

Introduction to Nanoscience and Nanotechnology: A Workbook



M. Kuno

August 22, 2005

Contents

1	Introduction	1
2	Structure	11
3	Length scales	37
4	Excitons	45
5	Quantum mechanics review	57
6	Confinement	75
7	Nondegenerate perturbation theory	107
8	Density of states	121
9	More density of states	129
10	Even more density of states	139
11	Joint density of states	151
12	Absorption	165
13	Interband transitions	173
14	Emission	195
15	Bands	213
16	K · P Approximation (Pronounced k dot p)	235

17 Tunneling	261
18 The WKB approximation	279
19 Synthesis	307
20 Tools	325
21 Applications	333
Acknowledgments	355

List of Figures

1.1	CdSe quantum dot	4
1.2	Quantum confinement	6
1.3	Dimensionality	7
1.4	Size dependent absorption and emission of CdSe	8
1.5	Artificial solid	9
2.1	14 3D Bravais lattices	12
2.2	Atoms per unit cell	14
2.3	Atom sharing	15
2.4	FCC unit cell	16
2.5	BCC unit cell	16
2.6	Hexagonal unit cell	17
2.7	Diamond structure unit cell	18
2.8	Zincblende or ZnS structure unit cell	19
2.9	NaCl structure unit cell	20
2.10	CsCl structure unit cell	21
2.11	Wurtzite unit cell	22
2.12	Miller index examples	23
2.13	More Miller index examples	24
2.14	Simple cubic	28
2.15	FCC	30
4.1	Exciton types	46
4.2	Exciton absorption	46
6.1	Particle in an infinite box	76
6.2	Half a harmonic oscillator	78
6.3	Various 1D potentials	79
6.4	Particle in a finite box	80
6.5	Particle in a finite box: solutions	81

6.6	Particle in a finite well: Mathcad solutions	85
6.7	Particle in an infinite circle	87
6.8	Particle in a sphere	92
8.1	3D density of states	123
8.2	2D density of states	125
8.3	1D density of states	126
8.4	0D density of states	127
9.1	3D density of CB and VB states	134
9.2	3D Fermi level	137
10.1	2D density of CB states	140
10.2	2D density of VB states	142
10.3	1D density of CB states	144
10.4	1D density of VB states	147
11.1	Vertical transitions	153
11.2	3D joint density of states	155
11.3	2D joint density of states	158
11.4	1D joint density of states	161
11.5	Summary, joint density of states	162
14.1	Einstein A and B coefficients	196
14.2	Derived emission spectrum: Einstein A and B coefficients . .	209
14.3	Pulsed experiment and lifetime	209
14.4	Radiative decay of excited state	211
14.5	Multiple pathway decay of excited state	211
15.1	Kronig-Penney rectangular potential	213
15.2	Kronig-Penney delta function potential	224
15.3	General Kronig Penney model: Mathcad solutions	227
15.4	General Kronig Penney model continued: Mathcad solutions	228
15.5	Kronig Penney model revisited: Mathcad solutions	229
15.6	Kronig Penney model, delta functions: Mathcad solutions . .	230
15.7	Kronig Penney model, delta functions continued: Mathcad solutions	231
15.8	From metals to insulators	232
17.1	Potential step	261
17.2	Potential step ($\varepsilon > V$)	262

17.3 Potential step ($\varepsilon < V$)	263
17.4 Potential barrier	269
17.5 Potential barrier ($\varepsilon > V$)	269
17.6 Potential barrier ($\varepsilon < V$)	275
17.7 Semiconductor junction	278
18.1 Arbitrary potential step	281
18.2 Arbitrary potential drop	286
18.3 Arbitrary potential barrier	290
18.4 Field emission	298
18.5 Shottky barrier	300
18.6 Parabolic barrier	301
18.7 Linear barrier	303
18.8 Parabolic barrier	304
19.1 Cartoon of a MBE apparatus	308
19.2 Cartoon of a MOCVD apparatus	309
19.3 Colloidal synthesis apparatus	311
19.4 LaMer model	317
19.5 LaMer model: size distribution	319
20.1 Transmission electron microscopy	326
20.2 Secondary electron microscopy	327
20.3 Atomic force microscopy	328
20.4 Scanning tunneling microscopy	329
20.5 Dip pen nanolithography	330
20.6 Microcontact printing	331
21.1 Nanowire device	334
21.2 Nanowire sensor	335
21.3 Quantum dot/dye photobleaching	337
21.4 Quantum dot/dye absorption/emission spectra	338
21.5 Density of states for lasing	340
21.6 Solar spectrum and QD absorption/emission spectra	342
21.7 Quantum dot LED schematic	344
21.8 Orthodox model of single electron tunneling	346
21.9 Coulomb Staircase	350

List of Tables

2.1	Common metals	25
2.2	Group IV semiconductors	25
2.3	Group III-V semiconductors	26
2.4	Group II-VI semiconductors	26
2.5	Group IV-VI semiconductors	27

Preface

This set of lecture notes about nanoscience and nanotechnology was initially written over the spring and summer of 2003. After my initial appointment as an assistant professor in chemistry, I agreed to teach an introductory class on nanoscience and nanotechnology for incoming graduate students and upper level undergraduates at the University of Notre Dame. However after accepting this task, it quickly became apparent to me that there were few resources available for teaching such a class, let alone any textbook. So while waiting for equipment to arrive, I undertook it upon myself to compile a series of lecture notes that would explain to the student some of the underlying concepts behind “nano”. The motivation for this was to describe to the student the physics behind each concept or assumption commonly encountered in the nano literature rather than just providing a qualitative overview of developments in the field. I have also tried to illustrate and motivate these concepts with instances in the current literature where such concepts are applied or have been assumed. In this manner, the goal is to provide the student with a foundation by which they can critically analyze, and possibly in the future, contribute to the emerging nano field. It is also my hope that one day, these lecture notes can be converted into an introductory text so that others may benefit as well.

Masaru Kenneth Kuno
Notre Dame, IN
August 24, 2003

Revision History

- Version 1.0 complete 8/03
- Corrected for typos and such as well as added the Kane chapter 6/04
- Added Absorption section 7/05
- Added Exciton section 7/05
- Added Quantum mechanics review section 7/05
- Added Non-degenerate perturbation theory section 7/05
- Added Time dependent perturbation theory section 7/05

Chapter 1

Introduction

Preliminaries

What is “nano”? Well, without providing a definite answer to this question, nano is a popular (emerging) area of science and technology today. It has attracted the attention of researchers from all walks of life, from physics to chemistry to biology and engineering. Further impetus for this movement comes from the tremendous increase in public and private funding for nano over the last ten years. A prime example of this is the new National Nanotechnology Initiative (NNI) created by former President Bill Clinton. The NNI increases funding in national nanoscience and nanotechnology research by hundreds of millions of dollars yearly. In addition, private sector contributions have jumped dramatically as evidenced by the plethora of small startup firms lining the tech corridors of the east and west.

Nano has even entered popular culture. It’s used as a buzzword in contemporary books, movies and television commercials. For example, in the recent blockbuster, *Spiderman*, the Willem Dafoe character (the Green Goblin) is a famous (and wildly wealthy) nanotechnologist whose papers the Tobey McGuire character (*Spiderman*) has read and followed (see the scene outside of Columbia university). Likewise, in the movie “*Minority Report*” Tom Cruise’s character undergoes eye surgery to avoid biometric fingerprinting. This scene involves a retinal eye transplant aided by so called “nano reconstructors”. A scene in the DC metro shows him reading a newspaper with the headline “nanotechnology breakthrough”. In television, a current GE commercial for washers and dryers features the storyline of: geeky nanotechnologist bumps into a supermodel at the laundromat resulting in love at first sight. We’re therefore, implicitly, told to believe that their mix of

brains and beauty embody GE's new washer/dryer line. In books, the New York Times bestseller "Prey" by Michael Crichton features nanotechnology run amok with spawns of tiny nano bots escaping from the laboratory and hunting down people for food. (Sounds like the "Andromeda Strain" except recast with nano as opposed to an alien virus.)

The mantle of nano has also been adopted by various scientific visionaries. Perhaps the most prominent is Eric Drexler who has founded an institute, called the Foresight Institute, devoted to exploring his ideas. Concepts being discussed include developing tiny nano robots that will "live" inside us and repair our blood vessels when damaged, preventing heart attacks. They will also kill cancer, cure us when we are sick, mend bones when broken, make us smarter and even give us immortality. These nano robots will also serve as tiny factories, manufacturing anything and everything from food to antibiotics to energy. In turn, nanotechnology will provide a solution to all of mankind's problems whether it be hunger in developing countries or pollution in developed ones. Drexler therefore envisions a vast industrial revolution of unprecedented size and scale. At the same time, concurrent with his visions of a utopian future is a darker side, involving themes where such nano robots escape from the laboratory and evolve into sentient beings completely out of mankind's control. Such beings could then sow the seeds to mankind's own destruction in the spirit of recent movies and books such as *The Terminator*, *The Matrix* and *Prey*. Now, whether such predictions and visions of the future will ever become reality remains to be seen. However, any such developments will ultimately rely on the scientific research of today, which is, on a daily basis, laying down the foundation for tomorrow's nanoscience and nanotechnology.

In today's scientific realm, the word nano describes physical lengthscales that are on the order of a billionth of a meter long. Nanoscale materials therefore lie in a physical size regime between bulk, macroscale, materials (the realm of condensed matter physics) and molecular compounds (the realm of traditional chemistry). This mesoscopic size regime has previously been unexplored and beckons the researcher with images of a scientific wild west with opportunities abound for those willing to pack their wagons and head into the scientific and technological hinterland. In this respect, nanoscale physics, chemistry, biology and engineering asks basic, yet unanswered, questions such as how the optical and electrical properties of a given material evolve from those of individual atoms or molecules to those of the parent bulk. Other questions that nanoscience asks include:

- How does one make a nanometer sized object?

- How do you make many (identical) nanometer sized objects?
- How do the optical and electrical properties of this nanoscale object change with size?
- How do its optical and electrical properties change with its “dimensionality”?
- How do charges behave in nanoscale objects?
- How does charge transport occur in these materials?
- Do these nanoscale materials possess new and previously undiscovered properties?
- Are they useful?

The transition to nanoscience begins at this last point when we ask how these nanoscale materials might be exploited to improve our lives. Venture capital firms and mainstream industry have therefore taken up this challenge with many small startups trying to apply nanoscale materials in products ranging from better sunscreen lotions to fluorescent labels for biological imaging applications to next generation transistors that will one day store the entire content of the Library of Congress on the head of a pin. More established companies, such as GE, HP, Lucent and IBM, have also started their own in house nano programs to revolutionize consumer lighting, personal computing, data storage and so forth. So whether it be for household lighting or consumer electronics, a nano solution exists and there is very likely a company or person pursuing this vision of a nano future.

So what is nano? This series of lecture notes tries to answer this question by explaining the physical concepts behind why such small, nanoscale, materials are so interesting and potentially useful.

Overview

The idea behind these lecture notes is as follows: First in Chapter 2, the composition of solids is discussed to introduce common crystal structures found in nanomaterials. Solids come in a number of forms, from amorphous (glass-like) to polycrystalline (multiple domains) to crystalline. Much of nanoscience and nanotechnology focuses on nanometer sized crystalline solids, hence the emphasis on crystal structure. In addition, the structure

section also illustrates the increase in surface to volume ratio for nanomaterials over bulk. This is because in nanometer sized systems up to 50% of the atoms lie at the surface of a nanostructure, in direct contrast to macroscopic solids where such numbers are typically much smaller. The surface is therefore potentially important in dictating a material's optical and electrical properties when nanometer sized. Furthermore, the increase in surface area is important to applications where the surface to volume ratio plays a critical role such as in catalysis as well as in photovoltaics. Developments in this area, using nanostructures, have led to increasingly efficient solar cells such as the Gratzel cell. Finally, the concept of crystal structure and the periodic potential due to the ordered arrangement of atoms is central to the concept of electronic bands, which we will discuss later on.

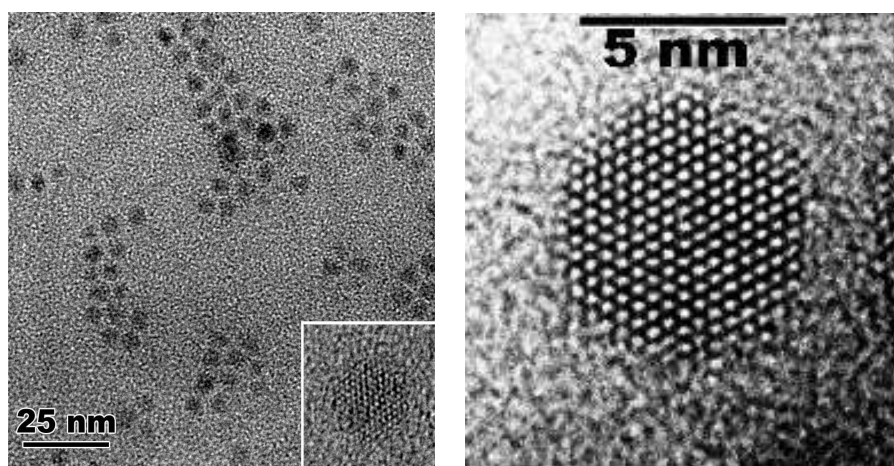


Figure 1.1: Transmission electron micrograph of individual CdSe quantum dots

Chapter 3 introduces the concept of length scales to put into perspective the actual physical lengths relevant to nano. Although being nanometer sized is often considered the essence of “nano”, the relevant physical length scales are actually relative to the natural electron or hole length scales in the parent bulk material. These natural length scales can either be referred to by their deBroglie wavelength or by the exciton Bohr radius. Thus, while a given nanometer sized object of one material may qualify for nano, a similar sized object of another material may not.

Next the concept of quantum confinement is introduced in Chapter 6 through the simple quantum mechanical analogy of a particle in a 1 di-

mensional, 2 dimensional and 3 dimensional box. Quantum confinement is most commonly associated with nano in the sense that bulk materials generally exhibit continuous absorption and electronic spectra. However, upon reaching a physical length scale equivalent to or less than either the exciton Bohr radius or deBroglie wavelength both the optical and electronic spectra become discrete and more atomic-like. In the extreme case of quantum dots, confinement occurs along all three physical dimensions, x,y,and z such that the optical and electrical spectra become truly atomic-like. This is one reason why quantum dots or nanocrystals are often called artificial atoms.

Analogies comparing the particle in a one dimensional box to a quantum well, the particle in a two dimensional box to a quantum wire and the particle in a three dimensional box to a quantum dot provide only half the solution. If one considers that in a quantum well only one dimension is confined and that two others are “free”, there are electronic states associated with these extra two degrees of freedom. Likewise in the case of a quantum wire, with two degrees of confinement, there exists one degree of freedom. So solving the particle in a two dimensional box problem models the electronic states along the two confined directions but does not address states associated with this remaining degree of freedom. To gain better insight into these additional states we introduce the concept of density of states (DOS) in Chapters 8,9,and 10. The density of states argument is subsequently applied to both the valence band and conduction band of a material. Putting together both valence and conduction band density of states introduces the concept of the joint density of states (JDOS) in Chapter 11 which, in turn, is related to the absorption coefficient of a material.

After describing the absorption spectra of 3D (bulk), 2D (quantum well), 1D (quantum wire), and 0D (quantum dot) systems we turn to the concept of photoluminescence. Generally speaking, in addition to absorbing light, systems will also emit light of certain frequencies. To describe this process, the Einstein A and B coefficients and their relationships are introduced and derived in Chapter 14. Finally, the emission spectrum of a bulk 3D material is calculated using the derived Einstein A and B coefficients. The concept of quantum yields and lifetimes, which describe the efficiency and timescale of the emission, completes this section.

Bands are introduced in Chapter 15. This topic is important because metals, semiconductors, and semi-metals all have bands due to the periodic potential experienced by the electron in a crystal. As mentioned earlier in the section on structure, this periodic potential occurs due to the ordered and repeated arrangement of atoms in a crystal. Furthermore, metals, semiconductors, insulators, and semi-metals can all be distinguished through

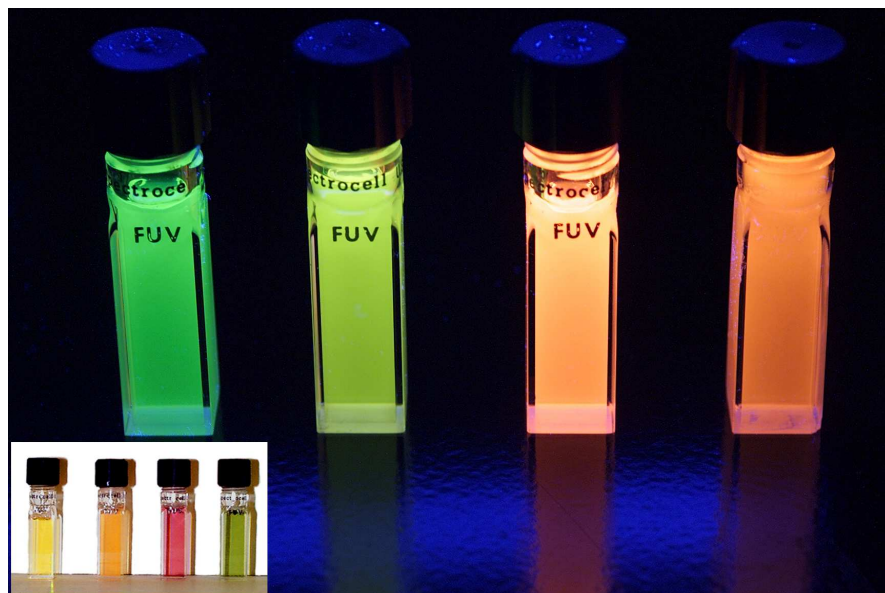


Figure 1.2: Photograph of the size dependent emission spectra of both HgS (top) and CdSe (bottom) quantum dots. Small quantum dots absorb and emit blue/green light, larger dots absorb and emit red light.

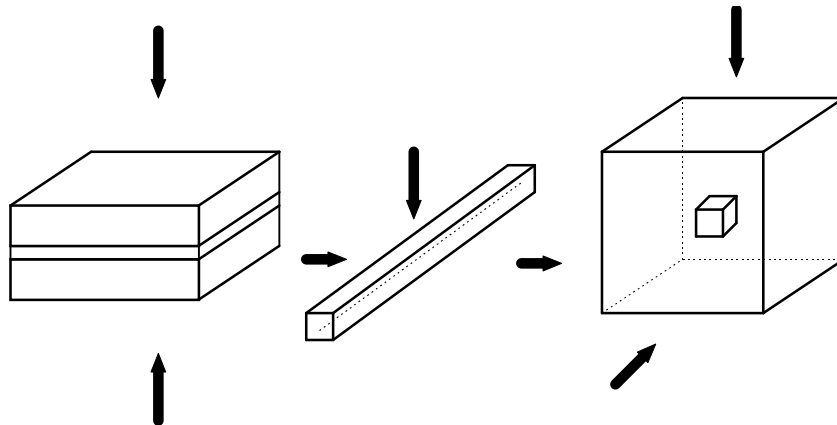


Figure 1.3: Cartoon of confinement along 1, 2 and 3 dimensions. Analogous to a quantum well, quantum wire and quantum dot.

the occupation of these bands by electrons. Metals have “full” conduction bands while semiconductors and insulators have “empty” conduction bands. At the same time, in the case of semiconductors and insulators there is a range of energies that cannot be populated by carriers separating the valence band from the conduction band. This forbidden range of energies (a no man’s land for electrons) is referred to as the band gap. The band gap is extremely important for optoelectronic applications of semiconductors. For example, the band gap will generally determine what colors of light a given semiconductor material will absorb or emit and, in turn, will determine their usefulness in applications such as solar energy conversion, photodetectors, or lasing applications. An exploration of the band gap concept ultimately touches on the effects quantum confinement has on the optical and electrical properties of a material, which leads to the realization of a size dependent band gap.

Introducing the concept of bands is also important for another reason since researchers have envisioned that ordered arrays of quantum wells, or wires, or dots, much like the arrangement of atoms in a crystal, can ultimately lead to new “artificial” solids with artificial bands and corresponding band gaps. These bands and associated gaps are formed from the delocalization of carriers in this new periodic potential defined by the ordered arrangement of quantum dots, quantum wires, or quantum wells. Talk about designer materials. Imagine artificial elements or even artificial metals, semimetals, and semiconductors. No wonder visionaries such as Drexler envision

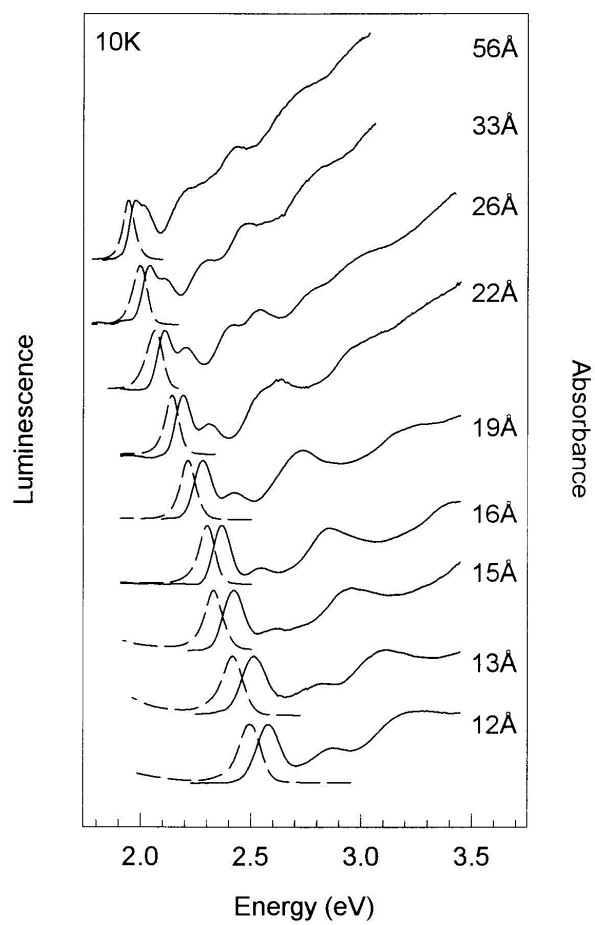


Figure 1.4: Size dependent absorption and emission spectra of colloidal CdSe quantum dots.

so much potential in nano. In the case of quantum wells, stacks of closely spaced wells have been grown leading to actual systems containing these minibands. In the case of quantum dots and wires, minibands have not been realized yet but this is not for lack of trying. The concept of creating artificial solids with tailor made bands from artificial atoms has been tantalizing to many.

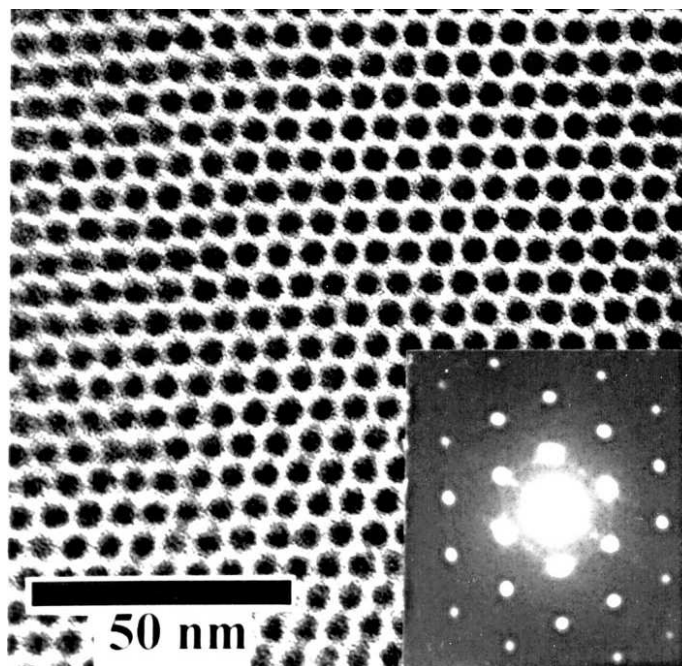


Figure 1.5: TEM micrograph of an array of colloidal CdSe quantum dots ordered into an artificial crystal. Each dark spot is an individual quantum dot. The white space in between dots is the organic ligands passivating the surface of the particle.

Not only do semiconductors, metals, insulators, and semi-metals absorb and emit light, but they also have electrical properties as well. These properties are greatly affected by quantum confinement and the discreteness of states just as with the aforementioned optical properties. Transport properties of these systems take center stage in the realm of devices where one desires to apply quantum dots, quantum wells, and quantum wires within opto-electronic devices such as single electron transistors. To this end, we introduce the concept of tunneling in Chapter 17 to motivate carrier trans-

port in nanometer-sized materials. Tunneling is a quantum mechanical effect where carriers can have non-zero probability of being located in energetically forbidden regions of a system. This becomes important when one considers that the discreteness of states in confined systems may mean that there are substantial “barriers” for carrier transport along certain physical directions of the material.

The WKB approximation is subsequently introduced in Chapter 18 to provide an approximate solution to Schrodinger’s equation in situations where the potential or potential barrier for the carrier varies slowly. In this fashion, one can repeat the same tunneling calculations as in the previous section in a faster, more general, fashion. A common result for the form of the tunneling probability through an arbitrary barrier is derived. This expression is commonly seen and assumed in much of the nano literature especially in scanning tunneling microscopy, as well as in another emerging field called molecular electronics.

After providing a gross overview of optical, electrical and transport properties of nanostructures, we turn to three topics that begin the transition from nanoscience to nanotechnology. Chapter 19 describes current methods for making nanoscale materials. Techniques such as molecular beam epitaxy (MBE), metal-organic chemical vapor deposition (MOCVD) and colloidal synthesis are described. A brief overview of each technique is given. Special emphasis is placed on better understanding colloidal growth models currently being used by chemists to make better, more uniform, quantum dots and nanorods. To this end, the classical LaMer and Dinegar growth model is derived and explained. Relations to the behavior of an ensemble size distribution are then discussed completing the section.

Once created, tools are needed to study as well as manipulate nanoscale objects. Chapter 20 describes some of the classical techniques used to characterize nanostructures such as transmission electron microscopy (TEM), secondary electron microscopy (SEM), atomic force microscopy (AFM) and scanning tunneling microscopy (STM). Newer techniques now coming into prominence are also discussed at the end of the section. In particular, the concepts behind dip-pen nanolithography and microcontact printing are illustrated.

Finally, Chapter 21 discusses applications of quantum dots, quantum wires, and quantum wells using examples from the current literature. Special emphasis is placed on the Coulomb blockade and Coulomb staircase problem, which is the basis of potential single electron transistors to be used in next generation electronics.

Chapter 2

Structure

Crystal structure of common materials

This section is not meant to be comprehensive. The interested reader may consult a number of excellent introductory references such as Kittel's introduction to solid state physics. The goal, however, is to illustrate the crystal structure of common materials often encountered in the nano literature. Solids generally appear in three forms, amorphous (no long range order, glass-like), polycrystalline (multiple domains) or crystalline (a single extended domain with long range order). Since nano typically concerns itself with crystalline metal nanoparticles and semiconductor nanocrystals, wires, and wells, having a basic picture of how the elements arrange themselves in these nanocrystalline systems is important. In this respect, crystal structure comes into play in many aspects of research, from a material's electronic spectra to its density and even to its powder x-ray diffraction pattern.

Atoms in a crystal are generally pictured as being arranged on an imaginary lattice. Individual atoms (or groups of atoms) are hung off of the lattice, much like Christmas ornaments. These individual (or groups of) atoms are referred to as the "basis" of the lattice. The endless repetition of basis atom(s) on a lattice makes up the crystal. In the simplest case, the basis consists of only a single atom and each atom is located directly over a lattice point. However, it is also very common to see a basis consisting of multiple atoms, which is the case when one deals with binary or even ternary semiconductors. Here the basis atoms do not necessarily sit at the same position as a lattice point, potentially causing some confusion when you first look at the crystal structures of these materials.

There are 14 three dimensional Bravais lattices shown in Figure 2.1.

These are also referred to as conventional unit cells (i.e. used in everyday life) as opposed to the primitive unit cell of which only the simple cubic (aka “primitive”) lattice qualifies. That is, most of these unit cells are not the simplest repeating units of an extended lattice; one can find even simpler repeating units by looking harder. Rather, these conventional cells happen to be easy to visualize and interpret and hence are the ones most commonly used.

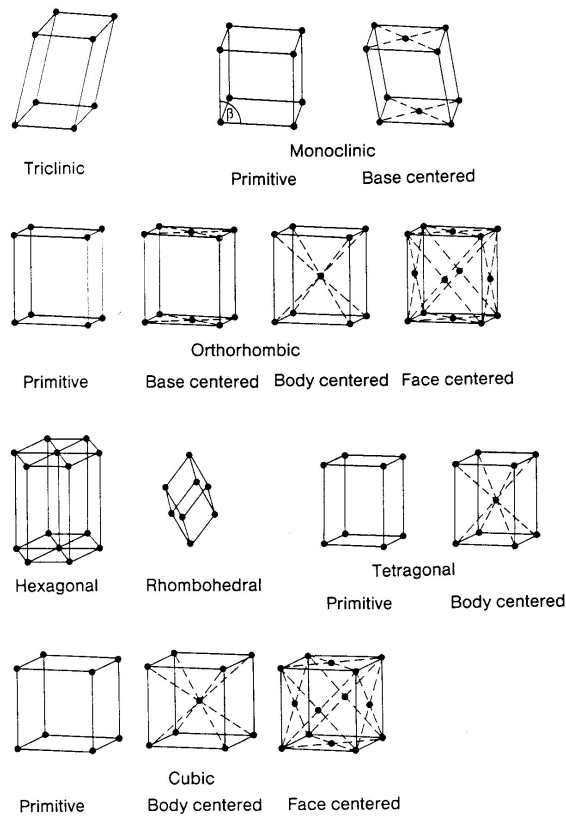


Figure 2.1: 14 3-dimensional Bravais lattices. From Ibach and Luth.

It is often useful to know the number of atoms in a unit cell. A general counting scheme for the number of atoms per unit cell follows.

- atoms entirely inside the unit cell: worth 1
- corner atoms: worth $\frac{1}{8}$

- face atoms: worth $\frac{1}{2}$
- edge atoms: worth $\frac{1}{4}$

Figure 2.2 illustrates the positions of an atom in the corner, face, edge and interior positions. The weighting scheme follows because each of these unit cells is surrounded by other unit cells, hence all atoms with the exception of interior ones are shared. The corner atoms are shared by 8 unit cells while the face atoms are shared by 2 cells and finally the edge atoms by four. Any atom in the interior of a unit cell is exclusive to that cell. This is illustrated in Figure 2.3. By counting atoms in this fashion, one can determine the number of atoms per unit cell.

Single element crystals

In the case of metals, the cubic lattices are important, with particular emphasis on the face centered cubic (FCC) and body centered cubic (BCC) structures. Both FCC and BCC structures have a single atom basis; thus, they strongly resemble the Bravais lattices or conventional unit cells seen in the previous diagram. The number of atoms per unit cell in the FCC case is 4 (8 corner atoms and 6 face atoms). Likewise, the number of atoms per BCC unit cell, using the above counting scheme, is 2 (1 interior atom and 8 corner atoms). Note that an alternative name exists for the FCC unit cell: cubic close packed (CCP), which should be remembered when reading the literature. Both unit cells are shown in Figures 2.15 and 2.5. Typical elements that crystallize in the FCC structure include: Cu, Ag, Au, Ni, Pd, Pt, and Al. Typical elements that crystallize in the BCC structure include: Fe, Cr, V, Nb, Ta, W and Mo. More complete tables can be found in Kittel.

Analogous to the FCC lattice is the hexagonal close packed (HCP) structure. A simple way to differentiate the two is the atomic packing order, which follows ABCABC in the case of FCC and ABABA in the case of HCP in the $\hat{j}001\hat{i}$ and $\hat{j}0001\hat{i}$ directions respectively. The letters A, B, and C etc... represent different atom planes. The HCP structure has a conventional unit cell shown in Figure 2.6. It contains 2 atoms per unit cell (8 on the corners and 1 inside).

Another conventional unit cell that is often encountered is called the “diamond” structure. The diamond structure differs from its FCC and BCC counterparts because it has a multi atom basis. Therefore, it does not immediately resemble any of the 14 Bravais lattices. In fact, the diamond structure has a FCC lattice with a two atom basis. The first atom is at $(0, 0, 0)$ and the second atom is at $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. It is adopted by elements that

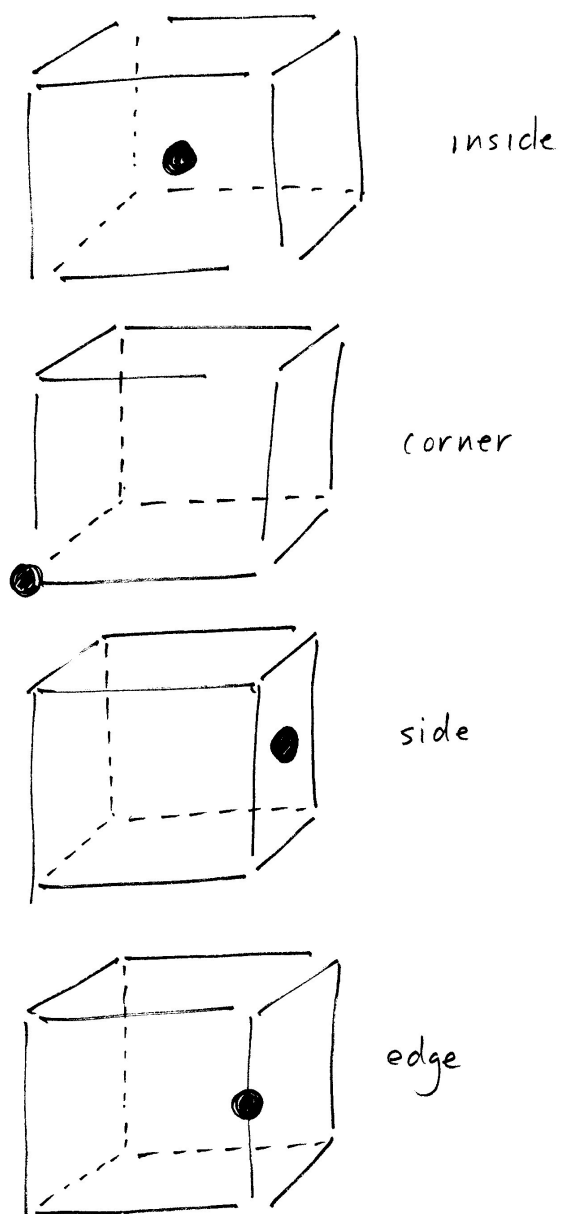


Figure 2.2: Number of atoms per unit cell. Counting scheme.

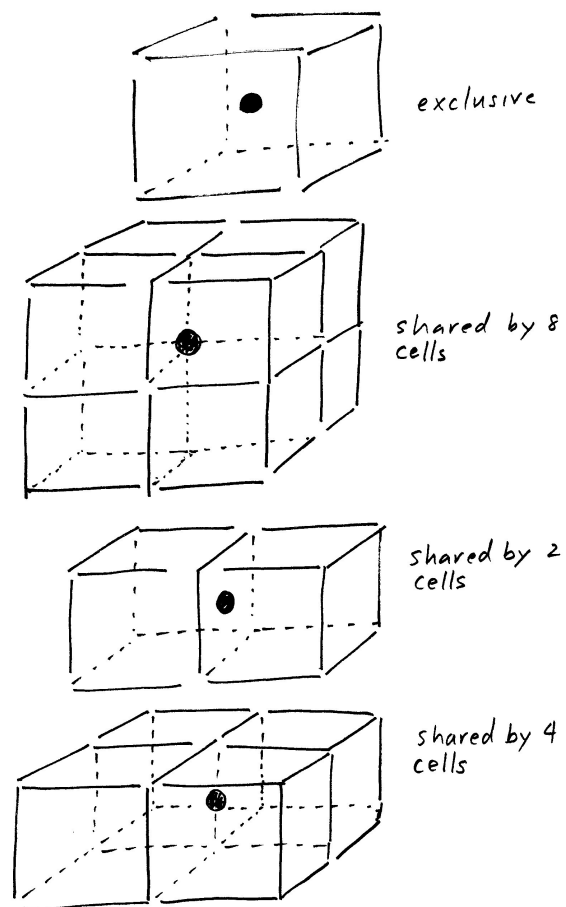


Figure 2.3: Cartoon showing sharing of atoms by multiple unit cells.

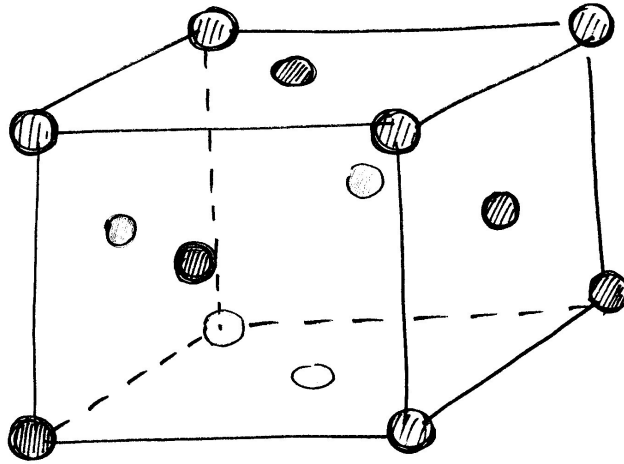


Figure 2.4: FCC unit cell

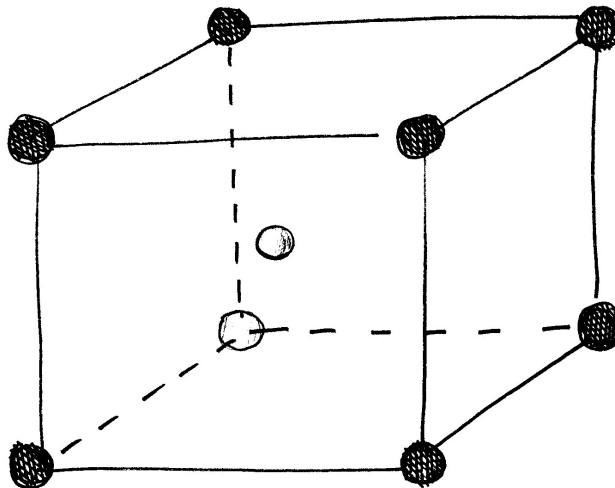


Figure 2.5: BCC unit cell

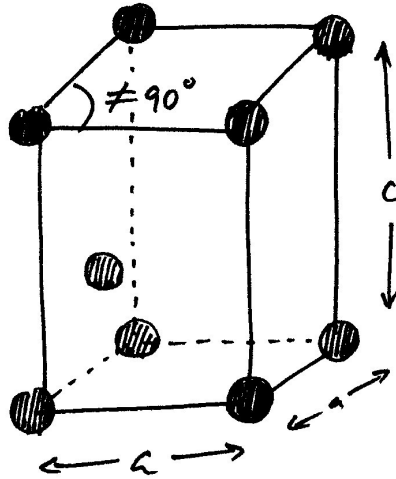


Figure 2.6: Hexagonal unit cell.

have a tendency to form strong covalent bonds, resulting in tetrahedral bonding arrangements (Figure 2.7). The number of atoms per unit cell in this case is 8 (8 corner atoms, 4 interior atoms, 6 face atoms). Some common elements that crystallize in the diamond structure include: C, Si, Ge and Sn. As a side note, one way to visualize the diamond unit cell is to picture two interpenetrating FCC lattices, offset from each other by a $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ displacement.

Summary, diamond

- lattice:FCC
- basis atom 1: $(0, 0, 0)$
- basis atom 2: $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$

Compound crystals

In the case of binary compounds, such as III-V and II-V semiconductors, things get a little more complicated. One doesn't have the benefit of conventional unit cells that resemble any of the 14 standard Bravais lattices.

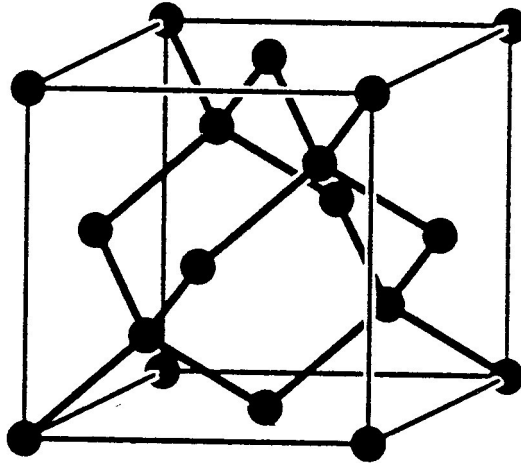


Figure 2.7: Diamond structure unit cell

Instead these conventional unit cells often have names such as the “NaCl” structure or the “ZnS” structure and so forth. This is because, unlike simple FCC or BCC metals, we no longer have a single atom basis, but rather a basis consisting of multiple atoms as well as a basis made up of different elements.

Common crystal lattices for semiconductors include the “ZnS”, “NaCl” and “CsCl” lattices.

The ZnS, also called zinc blende (ZB) or sphalerite, structure can be visualized as two interpenetrating FCC lattices offset by $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ in Figure 2.8. It is identical to the diamond structure we saw in the case of single element crystals. The only real difference is that now we have two elements making up the atom basis of the unit cell. The underlying lattice is FCC. The two atom, two element basis has positions of element1 $(0,0,0)$ and element2 $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. Using the above counting scheme we find that there are 8 atoms per unit cell. This is further subdivided into 4 atoms of element 1, and 4 atoms of element 2. You will notice in the figure that the 4 atoms of one element are completely inside the unit cell and that the atoms of the other element are arranged as 8 corner and 6 face atoms.

Summary, ZnS

- lattice: FCC
- basis, element 1 $(0, 0, 0)$
- basis, element 2 $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$

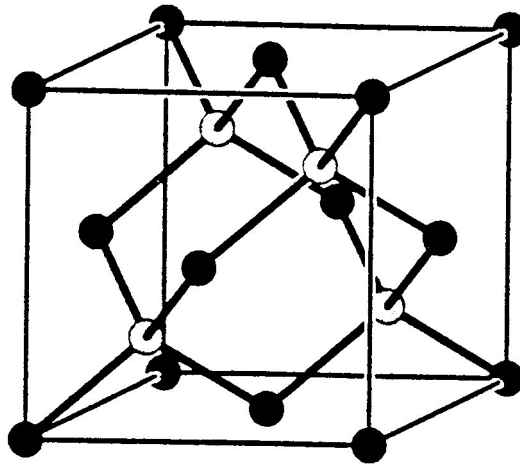


Figure 2.8: Zincblende or ZnS structure unit cell.

The NaCl structure (also called rocksalt structure) can be visualized as 2 interpenetrating FCC lattices offset by $(\frac{1}{2}, 0, 0)$ in Figure 2.9. The underlying lattice is FCC. The two element, two atom basis has locations of element1 $(0, 0, 0)$ and element2 $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$. It has 8 atoms per unit cell. This is broken up into 4 atoms from element 1 and 4 atoms from element 2. One can see in the figure that for element 1 there are 8 corner atoms and 6 face atoms. For element 2 there are 12 edge atoms and 1 interior atom. Examples of materials that crystallize in this structure include PbS, PbSe, and PbTe.

Summary, NaCl

- lattice: FCC
- basis, element 1 $(0, 0, 0)$

- basis, element 2 $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$

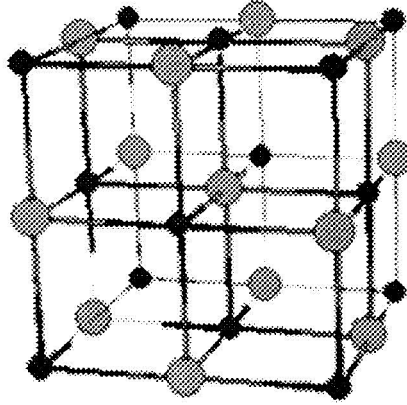


Figure 2.9: NaCl structure unit cell.

The CsCl structure is the compound material version of the single element BCC unit cell. It is shown in Figure 2.10 where one can see that there are two elements present with one of them being the center atom. The atoms from the other element take up corner positions in the unit cell. The underlying lattice is simple cubic and the two atom, two element basis is element 1 at $(0, 0, 0)$ and element 2 at $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$. The CsCl has two atoms per unit cell, 1 from each element. Examples of materials that crystallize in this structure include, CsCl and CuZn.

Summary, CsCl

- lattice: SC
- basis, element 1 $(0, 0, 0)$
- basis, element 2 $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$

The wurtzite crystal structure is the compound material version of the single element HCP structure. It has a multi atom basis. The underlying lattice is hexagonal. The two element, two atom basis is element 1 $(0, 0, 0)$ and element 2 $(\frac{2}{3}, \frac{1}{3}, \frac{1}{2})$. The unit cell is shown in Figure 2.11 and contains 4 atoms per unit cell, 2 atoms from element 1 and 2 atoms from element 2. Examples of materials that crystallize in this structure include CdSe and GaN.

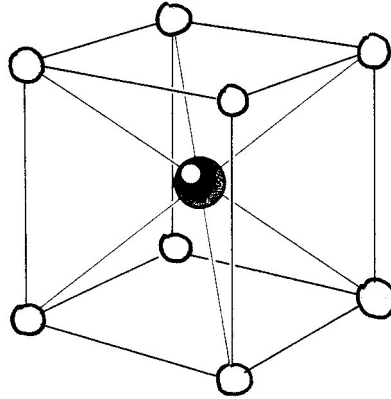


Figure 2.10: CsCl structure unit cell.

Summary, wurtzite

- lattice: hexagonal
- basis, element 1 $(0, 0, 0)$
- basis, element 2 $(\frac{2}{3}, \frac{1}{3}, \frac{1}{2})$

Miller indices

Sometimes you will see the orientation of a crystal plane described by (001) and so forth. These numbers are referred to as Miller indices. They are generated using some simple rules described below.

- Take the desired plane and see where it intersects each x, y, z axis in multiples of the lattice constant. For the case of cubic lattices the lattice constant, a , is the same in all x, y, and z directions.
- Next take the reciprocal of each intersection point and reduce the three values to their lowest integer values. (i.e. divide out any common integer)
- Express the plane through these integers in parentheses as (abc)

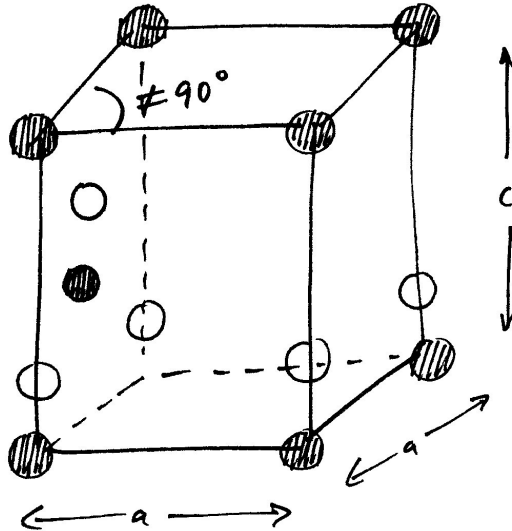


Figure 2.11: Wurtzite unit cell.

- Should the plane not intersect an axis, say the z axis, just write a 0. For example $(ab0)$
- If the intercept is in the negative side of an axis, say the y axis, just put a bar over the number, for example $(a\bar{b}c)$.

Miller notation

- (hkl) = crystal plane
- $\{hkl\}$ = equivalent planes
- $[hkl]$ = crystal direction
- $\langle hkl \rangle$ = equivalent directions

Examples are illustrated in Figures 2.12 and 2.13.

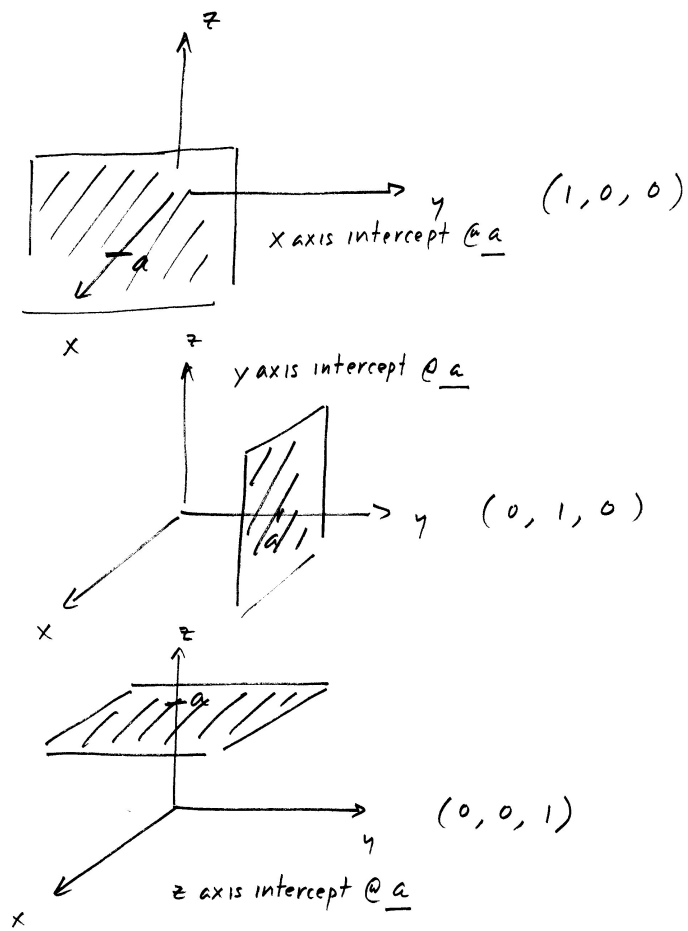


Figure 2.12: Examples of using Miller indices.

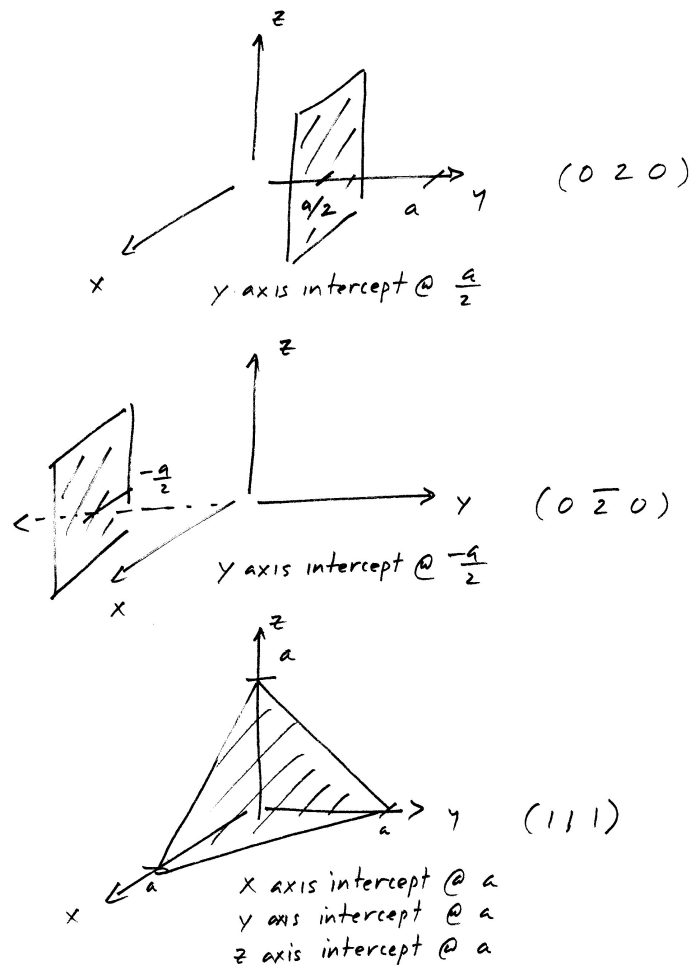


Figure 2.13: More examples of using Miller indices.

Quick tables

Short tables of common metals and semiconductors are provided below with their standard crystal structure.

Common Metals

Table 2.1: Common metals

I	II	III	IV	V	VI
		B	C	N	
		Al	Si	P	S
Cu	Zn	Ga	Ge	As	Se
Ag	Cd	In	Sn	Sb	Te
Au	Hg	Tl	Pb	Bi	Po

- Ag=FCC [cubic] (alternatively called cubic closest packed)
- Au=FCC [cubic] (alternatively called cubic closest packed)

Common Semiconductors

Group IV

Table 2.2: Group IV semiconductors

I	II	III	IV	V	VI
		B	C	N	
		Al	Si	P	S
Cu	Zn	Ga	Ge	As	Se
Ag	Cd	In	Sn	Sb	Te
Au	Hg	Tl	Pb	Bi	Po

- Si=diamond structure
- Ge=diamond structure

III-V

- GaN=ZB [cubic] (alternatively called ZnS structure)

Table 2.3: Group III-V semiconductors

I	II	III	IV	V	VI
		B	C	N	
		Al	Si	P	S
Cu	Zn	Ga	Ge	As	Se
Ag	Cd	In	Sn	Sb	Te
Au	Hg	Tl	Pb	Bi	Po

- GaAs=ZB [cubic] (alternatively called ZnS structure)
- InP=ZB [cubic] (alternatively called ZnS structure)
- InAs=ZB [cubic] (alternatively called ZnS structure)

II-VI

Table 2.4: Group II-VI semiconductors

I	II	III	IV	V	VI
		B	C	N	
		Al	Si	P	S
Cu	Zn	Ga	Ge	As	Se
Ag	Cd	In	Sn	Sb	Te
Au	Hg	Tl	Pb	Bi	Po

- ZnS=ZB [cubic]
- ZnSe=ZB [cubic]
- CdS=ZB [cubic]
- CdSe=wurtzite [hexagonal]
- CdTe=ZB [cubic]

IV-VI

- PbS=NaCl structure (alternatively called rocksalt)
- PbSe=NaCl structure (alternatively called rocksalt)
- PbTe=NaCl structure (alternatively called rocksalt)

Table 2.5: Group IV-VI semiconductors

I	II	III	IV	V	VI
		B	C	N	
		Al	Si	P	<i>S</i>
Cu	Zn	Ga	Ge	As	<i>Se</i>
Ag	Cd	In	Sn	Sb	<i>Te</i>
Au	Hg	Tl	<i>Pb</i>	Bi	Po

Packing fraction

Supposing atoms to be rigid spheres, what fraction of space is filled by atoms in the primitive cubic and fcc lattices. Assume that the spheres touch.

Simple cubic

Determine the packing fraction of the simple cubic lattice.

$$V_{sphere} = \frac{4}{3}\pi r^3$$

where $\frac{r=a}{2}$

$$\begin{aligned} V_{sphere} &= \frac{4}{3}\pi \left(\frac{a}{2}\right)^3 \\ &= \frac{\pi a^3}{6} \end{aligned}$$

Also there is only 1 sphere in the cube. Recall that each corner sphere is shared by 8 other neighboring cubes.

Next the volume of the cube is

$$V_{cube} = a^3$$

so that the ratio of volumes is

$$\begin{aligned} Ratio &= \frac{\frac{\pi a^3}{6}}{a^3} \\ &= \frac{\pi}{6} \end{aligned}$$

The desired simple cubic packing fraction is

$$\boxed{Ratio_{sc} = 0.524} \quad (2.1)$$

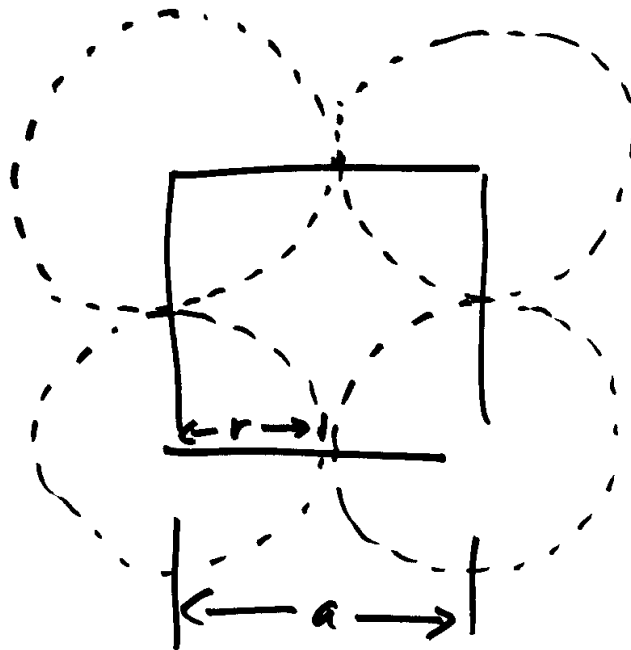


Figure 2.14: simple cubic.

FCC

Determine the packing fraction of the fcc lattice

$$\begin{aligned} c^2 &= a^2 + a^2 \\ &= 2a^2 \\ c &= \sqrt{2}a \\ &= 4r \end{aligned}$$

This results in

$$r = \frac{\sqrt{2}}{4}a$$

The volume of a single sphere is

$$\begin{aligned} V_{sphere} &= \frac{4}{3}\pi r^3 \\ &= \frac{4}{3}\pi \left(\frac{\sqrt{2}}{4}\right)^3 a^3 \\ &= \frac{\sqrt{2}}{24}\pi a^3 \end{aligned}$$

Next we calculate how many spheres are in the cube. In the fcc lattice there are 8 corner spheres and 6 face spheres. The calculation is

- Corner: $8 \left(\frac{1}{8}\right) = 1$
- Face: $6 \left(\frac{1}{2}\right) = 3$

The total number of spheres in the cube is therefore 4. The total volume of the spheres in the cube is therefore

$$\begin{aligned} V_{spheretotal} = 4V_{sphere} &= 4 \left(\frac{\sqrt{2}\pi}{24}\right) \\ &= \frac{\sqrt{2}}{6}\pi a^3 \end{aligned}$$

The volume of the cube is simply

$$V_{cube} = a^3$$

The desired ratio of volumes is then

$$\begin{aligned} \text{Ratio} &= \frac{\sqrt{2}\pi a^3}{6} \\ &= \frac{\sqrt{2}}{6}\pi \\ &= 0.741 \end{aligned}$$

The packing fraction of the fcc lattice is

$$\boxed{\text{Ratio}_{fcc} = 0.741} \quad (2.2)$$

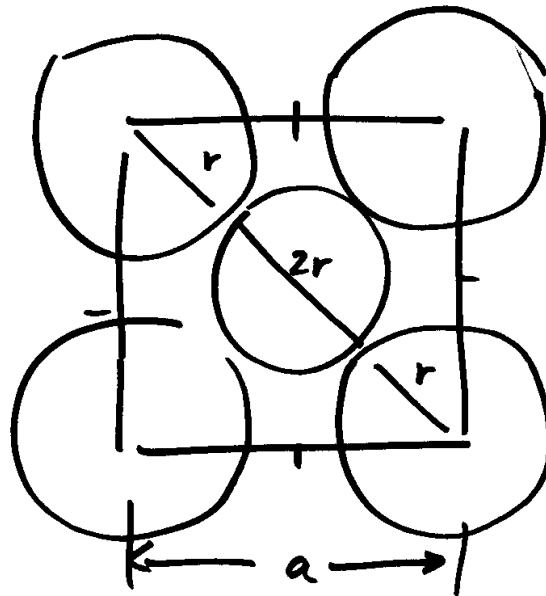


Figure 2.15: FCC.

Surface to volume ratio

Here we show that the surface to volume ratio of a number of systems scales as an inverse power law. Smaller sizes translate to larger surface to volume ratios.

Cube

Surface area

$$S = 6a^2$$

where a is the radius.

Volume

$$V = a^3$$

Surface to volume ratio

$$\begin{aligned} R = \frac{S}{V} &= \frac{6a^2}{a^3} \\ &= \frac{6}{a} \end{aligned}$$

Sphere

Surface area

$$S = 4\pi a^2$$

Volume of sphere

$$V = \frac{4}{3}\pi a^3$$

Surface to volume ratio

$$\begin{aligned} R = \frac{S}{V} &= \frac{4\pi a^2}{\frac{4}{3}\pi a^3} \\ &= \frac{3}{a} \end{aligned}$$

Cylinder

Surface area of cylinder

$$S = 2\pi al$$

Volume of cylinder

$$V = \pi a^2 l$$

Surface to volume ratio

$$\begin{aligned} R = \frac{S}{V} &= \frac{2\pi al}{\pi a^2 l} \\ &= \frac{2}{a} \end{aligned} \quad (2.3)$$

In all cases, the surface to volume ratio follows an inverse power law. But the bottom line is that as the size of the system decreases, the fraction of atoms on the surface will increase. This is both good and bad. In one case illustrated below it can lead to more efficient catalysts. However in other cases, such as in the optical microscopy of nanostructures it can be a bad thing because of defects at the surface which quench the emission of the sample.

Why, surface to volume ratio?

Since surfaces are useful for things such as catalysis, there is often an incentive to make nanostructured materials in the hopes that the increase in surface to volume ratio will lead to more catalytic activity. To illustrate consider the following example. Start with a 1 cm radius sphere. Calculate the surface area. Next, consider a collection of 1nm radius spheres that will have the same total volume as the original sphere. What is the total surface area of these little spheres. How does it compare?

The surface of the large sphere is

$$\begin{aligned} S_{large} &= 4\pi a^2 \\ &= 4\pi cm^2 \end{aligned}$$

The volume of the large sphere is

$$\begin{aligned} V_{large} &= \frac{4}{3}\pi a^3 \\ &= \frac{4}{3}\pi cm^3 \end{aligned}$$

Next, how many little spheres do I need to make up the same volume.

$$\begin{aligned} V_{little} &= \frac{4}{3}\pi(10^{-7})^3 \\ &= \frac{4}{3}\pi 10^{-21} \end{aligned}$$

The ratio of the large to small sphere volumes is

$$\begin{aligned}\frac{V_{large}}{V_{little}} &= \frac{\frac{4}{3}\pi}{\frac{4}{3}\pi 10^{-21}} \\ &= 10^{21}\end{aligned}$$

Next the surface area of an individual little sphere is

$$\begin{aligned}S_{little} &= 4\pi(10^{-7})^2 \\ &= 4\pi 10^{-14} \text{ cm}^2\end{aligned}$$

Multiply this by 10^{21} to get the total surface area,

$$\begin{aligned}V_{little,tot} &= 4\pi 10^{-14} 10^{21} \\ &= 4\pi 10^7 \text{ cm}^2\end{aligned}$$

So finally, before we had a total surface area of $4\pi \text{ cm}^2$. Now we have 10^7 times more surface area with the little spheres.

Exercises

1. The lattice constant of Si is 5.43 \AA . Calculate the number of silicon atoms in a cubic centimeter.
2. Calculate the number of Ga atoms per cubic centimeter in a GaAs crystal. The lattice constant of GaAs is 5.65 \AA . Do the same for As atoms.
3. Consider an actual silicon device on a wafer with physical dimensions of $5 \times 5 \times 1$ microns. Calculate the total number of atoms in the device.
4. Consider a slightly larger GaAs laser on the same wafer. Its dimensions are $100 \times 100 \times 20$ microns. How many total atoms exist on this device.
5. Calculate the number of atoms in a 1.4 nm diameter Pt nanoparticle using the total number of unit cells present. Consider a FCC unit cell with a lattice constant of $a = 0.391 \text{ nm}$.
6. Calculate the number of atoms in a 1.4 nm diameter Pt nanoparticle using the bulk density of Pt. Consider $\rho = 21.5 \text{ g/cm}^3$.

7. Estimate the number of surface atoms and percentage of surface atoms in a 1.4 nm diameter Pt nanoparticle. One can calculate this through a unit cell approach but use whatever approach you like.
8. Cobalt is usually found with a hexagonal crystal structure. It was recently found to crystallize with a simple cubic structure, now called ϵ -cobalt. The lattice constant of ϵ -cobalt is $a = 6.097\text{\AA}$. The density is $\rho = 8.635\text{g/cm}^3$. The unit cell contains 20 atoms. Calculate the number of atoms in a 2 nm diameter nanocrystal through a unit cell argument.
9. Calculate the number of atoms in a 2 nm diameter ϵ -cobalt nanocrystal through a density argument. Use $\rho = 8.635\text{g/cm}^3$.
10. Calculate the number of surface atoms and percentage of surface atoms in a 2 nm diameter ϵ -cobalt particle.
11. CdSe has a hexagonal unit cell (unit cell volume $\simeq 112\text{\AA}^3$). The lattice constants are $a = 4.3\text{\AA}$ and $c = 7\text{\AA}$. Calculate the total number of atoms in a CdSe quantum dot for a: 1, 2, 3, 4, 5, 6 nm diameter particle. How many atoms of each element are there?
12. For the same CdSe dots considered, calculate the fraction of surface atoms in each case and plot this on a graph.
13. Draw the surface of a Ag crystal cut along the (111) and (100) plane.

References

1. "Introduction to solid state physics"
C. Kittel
Wiley, New York, New York 1986.
2. "Solid state physics"
N. W. Ashcroft and N. D. Mermin
Holt, Rinehart and Winston, New York, New York 1976.
3. "Solid-state physics"
H. Ibach and H. Luth
Springer Verlag, Berlin 1990.

Relevant literature

1. "A solution-phase chemical approach to a new crystal structure of cobalt"
D. P. Dinega and M. G. Bawendi
Angew. Chem. Int. Ed. 38, 1788 (1999).

Chapter 3

Length scales

What are the relevant length scales for nano? Well, I guess it depends on who you talk to. On one hand some people call nano anything smaller than stuff on the micro level. This could mean delaing with stuff on the hundreds of nm scale. But as Uzi Landman once said, this is pretty high on the chutzpa scale. Now others will tell you that nano is for extremely small things like molecules and such. But isn't this just traditional chemistry or even biology? After all chemists, biologists and physicists have been dealing with small things for a pretty long time even before the development of instruments such as transmission electron micrographs. This is the opposite extreme-calling traditional chemistry, physics and biology, nano to put a new spin on things.

One useful perspective on a definition for the appropriate lengths scales for nano is a regime where the chemical, physical, optical and electrical properties of matter all become size and shape dependent. For semiconducting materials this is given by the bulk exciton Bohr radius. We will describe excitons (electron hole pairs) in more detail shortly, but for now just picture the classic Bohr model of electrons circling a heavy nucleus in stable orbits.

DeBroglie wavelength

So what is the deBroglie wavelength? Well this arose within the context of a classic debate about the true nature of light back in the old days. Was light a wave or was it a stream of particles? On one side you had Huygens who thought it was a wave. Then on the other side you had Newton who thought it was a stream of particles. Tastes great, less filling anyone?

So to make a long story short, deBroglie comes along later and says "both". This was his Ph.D thesis for which he got the Nobel prize in 1929.

But deBroglie’s claim actually goes beyond light because what it really says is that matter (whether it be you, me, a baseball, a car, a plane, an atom, an electron, a quantum dot or a nanowire—all of these things have both wave-like and particle-like properties). The wave-like properties of matter become important when you deal with small things so although macroscopic objects have wave-like properties they are better described by Newton’s laws. We will mostly deal with electrons from here on out.

Therefore according to deBroglie associated with each object is a wavelength.

$$\lambda = \frac{h}{p} \quad (3.1)$$

where λ is the deBroglie wavelength, h is Planck’s constant ($6.62 \times 10^{-34} J \cdot s$) and $p = mv$ is the momentum with m as the mass.

DeBroglie wavelength and exciton Bohr radius

Here we derive the relationship between the deBroglie wavelength and the exciton Bohr radius. The reason we do this is that often in the literature one sees a statement that a nanomaterial is in the “quantum confinement” regime because its size is smaller than the corresponding deBroglie wavelength of an electron or hole. At other times one sees the statement that a nanomaterial is quantum confined because its size is smaller than the corresponding exciton Bohr radius. We ask if these are the same statement.

In this section we show that the two are related and that, in fact, both statements essentially say the same thing.

Textbook Bohr radius

Here is the textbook equation for the Bohr radius of an electron

$$a_0 = \frac{4\pi\epsilon_0\hbar^2}{mq^2} \quad (3.2)$$

where $\epsilon_0 = 8.85 \times 10^{-12}$ F/m (permittivity), $\hbar = 1.054 \times 10^{-34}$ $J \cdot s$ (Planck’s constant over 2π), $m_e = 9.11 \times 10^{-31}$ kg (mass of a free electron) and $q = 1.602 \times 10^{-19}$ C (charge). If you plug all the numbers in and do the math you come up with the result

$$\begin{aligned} a_0 &= 5.28 \times 10^{-11} \text{meters} \\ &= 0.528 \text{ Angstroms} \end{aligned} \quad (3.3)$$

This is the standard Bohr radius one sees all the time.

Derivation

Basically we need to equate the centrifugal force of a carrier with the Coulomb attractive (inward) force.

$$\frac{mv^2}{r} = \frac{q^2}{4\pi\epsilon_0 r^2} \quad (3.4)$$

Here we make use of the relation

$$\boxed{2\pi r = n\lambda} \quad (3.5)$$

where n is an integer. Namely, that an integer number of wavelengths must “fit” into the circumference of the classic Bohr orbit. The deBroglie relation comes in by relating the wavelength $\lambda = \frac{h}{p}$ where h is Planck’s constant and p is the momentum of the particle ($p = mv$). Starting with the above equation we rearrange it to get

$$\lambda = \frac{2\pi r}{n} = \frac{h}{p} = \frac{h}{mv}$$

Solve for v to get (express v in terms of r)

$$v = \frac{nh}{2\pi mr} = \frac{n\hbar}{mr}$$

Replace this into the main equation (3.3)

$$\frac{n^2\hbar^2}{mr} = \frac{q^2}{4\pi\epsilon_0}$$

Rearrange this to get

$$r = \frac{4\pi\epsilon_0 n^2 \hbar^2}{mq^2}$$

If $n = 1$ (the lowest orbit) this gives us the Bohr radius

$$a_0 = \frac{4\pi\epsilon_0 \hbar^2}{mq^2}$$

which is the standard textbook equation we showed earlier.

At this point we note that if the electron or carrier is not in vacuum the equation should be modified to take into account the dielectric constant of the medium. (Instead of ϵ replace it with $\epsilon\epsilon_0$).

$$\boxed{a_0 = \frac{4\pi\epsilon_0\epsilon\hbar^2}{mq^2}} \quad (3.6)$$

This is because if you are not in vacuum which you aren't in a solid, liquid or gas, then one must account for possible extra charges or dipoles that can move to oppose the electric field that one establishes between the original two oppositely charged species. Any mobile charges or dipoles in the system will move or re-orient to oppose the field, trying to cancel it out. The net effect is that the original electric field as well as Coulomb potential are diminished by the response of the material. The factor by which it is diminished is the relative dielectric constant, denoted commonly as ϵ .

Furthermore, for the case of an exciton (electron hole pair) in a semiconductor just replace the mass of the electron with the effective mass of the exciton.

$$\boxed{\frac{1}{m_{eff}} = \frac{1}{m_e} + \frac{1}{m_h}} \quad (3.7)$$

where m_e and m_h are the effective masses of electron and hole in the material.

Note that equation 3.4 basically gives the relation between the deBroglie wavelength and the exciton Bohr radius. So, in effect, our initial statements about confinement dealing with either the exciton Bohr radius or deBroglie wavelength are essentially one and the same. The deBroglie wavelength or exciton Bohr radius are therefore natural length scales by which to compare the physical size of a nanomaterial. In general objects with dimensions smaller than these natural length scales will exhibit quantum confinement effects. This will be discussed in more detail in subsequent chapters.

Examples

Here are some values for some common systems where I've taken values of the relative dielectric constant, electron and hole effective masses from the literature. One can derive the exciton Bohr radius of these systems, using the values below, in a straightforward fashion. This list is not meant to be comprehensive and the interested reader should consult the Landolt Bornstein tables for more complete values.

- GaAs:
 $m_e = 0.067m_0$
 $m_h = 0.45m_0$
 $\epsilon = 12.4$
- InAs:
 $m_e = 0.02m_0$
 $m_h = 0.4m_0$
 $\epsilon = 14.5$
- InP:
 $m_e = 0.07m_0$
 $m_h = 0.4m_0$
 $\epsilon = 14$
- CdS:
 $m_e = 0.2m_0$
 $m_h = 0.7m_0$
 $\epsilon = 8.6$
- CdSe:
 $m_e = 0.13m_0$
 $m_h = 0.45m_0$
 $\epsilon = 9.4$

Worked example

Case: (GaAs)

$$\begin{aligned}\frac{1}{m_{eff}} &= \frac{1}{m_e} + \frac{1}{m_h} \\ &= \frac{1}{0.067m_0} + \frac{1}{0.45m_0}\end{aligned}$$

leading to the effective mass

$$m_{eff} = 0.058m_0$$

$$\begin{aligned}
 a_b &= \frac{4\pi\epsilon_0\epsilon\hbar^2}{0.058m_0q^2} \\
 &= \frac{4\pi(8.85 \times 10^{-12})(12.4)(1.054 \times 10^{-34})^2}{(0.058)(9.11 \times 10^{-31})(1.602 \times 10^{-19})^2} \\
 &= 11.3 \text{ nm}
 \end{aligned}$$

Exciton Bohr radius for GaAs.

Case: (CdSe)

$$\begin{aligned}
 \frac{1}{m_{eff}} &= \frac{1}{m_e} + \frac{1}{m_h} \\
 &= \frac{1}{0.13m_0} + \frac{1}{0.45m_0}
 \end{aligned}$$

leading to the effective mass

$$m_{eff} = 0.1m_0$$

$$\begin{aligned}
 a_b &= \frac{4\pi\epsilon_0\epsilon\hbar^2}{0.058m_0q^2} \\
 &= \frac{4\pi(8.85 \times 10^{-12})(9.4)(1.054 \times 10^{-34})^2}{(0.1)(9.11 \times 10^{-31})(1.602 \times 10^{-19})^2} \\
 &= 4.97 \text{ nm}
 \end{aligned}$$

Exciton Bohr radius for CdSe.

So in the end you can see that because of the variable relative dielectric constant and effective masses of carriers in a semiconductor there will be various exciton Bohr radii. Now finally let's put things into perspective. Consider the case of GaAs. We just found that the bulk exciton Bohr radius was $a_B = 11.3\text{nm}$. The lattice constant of GaAs which crystallizes in the zinc blende form is $a = 5.65\text{\AA}$. Thus the radius of the electron orbit is about 20 unit cells away from the counterpart hole. You can see that in semiconductors the electron can move very far away from its counterpart due to the screening effect of the material.

Exercises

1. What is the wavelength of a 1 eV photon?
2. What is the wavelength of a 2 eV photon?
3. What is the wavelength of a 3 eV photon?
4. What is your deBroglie wavelength (what's your weight in kg?) when moving at 10 m/s?
5. What is the deBroglie wavelength of a baseball (0.15 kg) moving at 50 m/s?
6. What is the deBroglie wavelength of C_{60} moving at 220 m/s? Read the corresponding article if you are interested. Arndt et. al. Nature 401, 680 (1999).
7. Calculate the exciton Bohr radius for the following semiconductors. If needed use values for what is called the heavy hole. Consult a good resource such as Landolt Bornstein.
 - II-VI compounds*
 - CdS
 - CdSe
 - CdTe
 - III-V compounds*
 - InP
 - InAs
 - IV-VI compounds*
 - PbS
 - PbSe
 - PbTe
8. Explain what the size of the exciton Bohr radius means for achieving quantum confinement. What systems are easiest for achieving confinement.

Chapter 4

Excitons

Excitons are bound electron-hole pairs caused by the absorption of a photon in a semiconductor. Specifically, we have an electron in the conduction band of the semiconductor and a hole in the valence band of the semiconductor. Excitons are potentially mobile and are charge neutral. There are two types of excitons, Mott-Wannier excitons and Frenkel excitons, distinguished by how close the electron and hole are. Mott-Wannier excitons have weak electron-hole interactions caused by a small Coulomb attraction and hence the carriers are relatively far apart. Corresponding binding energies are on the order of $10meV$. By contrast, Frenkel excitons have strong Coulomb interactions between the electron and hole. The carriers are close together. Corresponding binding energies are on the order of $100meV$.

The excitons most commonly dealt with in semiconductors is the Mott-Wannier exciton due to the large relative dielectric constant of most bulk semiconductors. Frenkel excitons are commonly seen in organic semiconductors.

Excitons can be seen in the absorption spectrum of bulk semiconductors. They generally appear just below the band edge of the semiconductor. This is because the energy of the exciton is lower than the band edge transition by its binding energy. Namely, the exciton energy $E_{exc} = E_g - E_{bind}$ where E_g is the bulk band gap of the semiconductor (analogous to the HOMO-LUMO gap in molecular systems) and E_{bind} is the exciton binding energy that we will derive shortly.

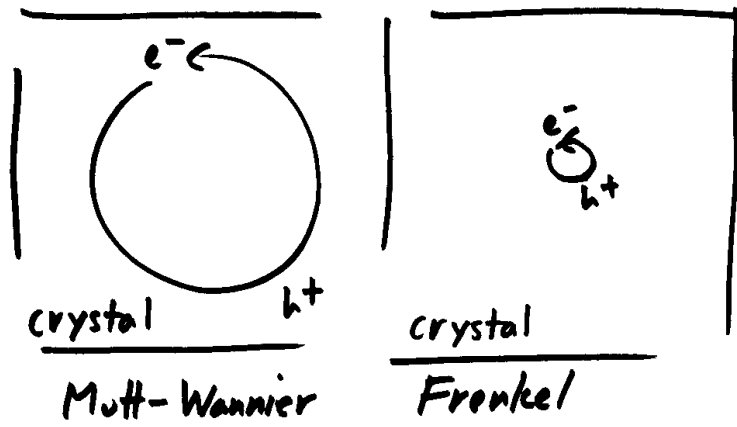


Figure 4.1: Exciton types

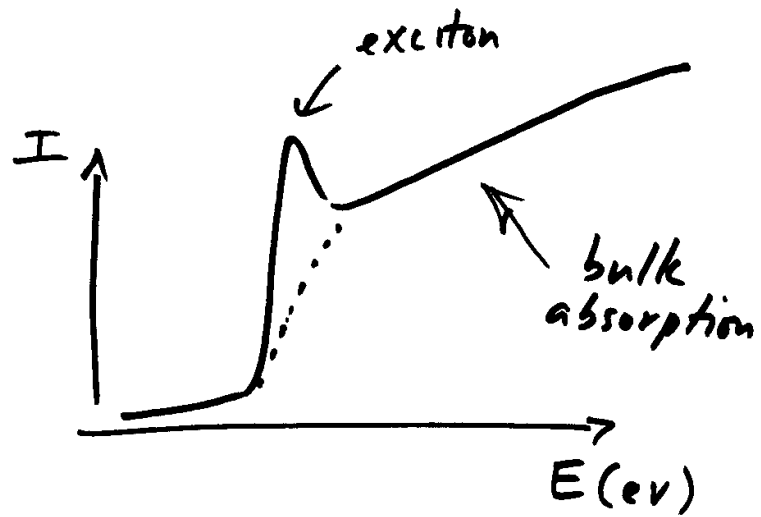


Figure 4.2: Exciton absorption

Exciton Bohr radius and binding energies

From the previous chapter we know that the Bohr radius of the electron-hole pair is

$$a_B = \frac{4\pi\epsilon\epsilon_o\hbar^2}{\mu q^2}$$

where μ is the exciton reduced mass (i used m last time, sorry for the notation change).

Usually folks like to write the exciton Bohr radius in terms of the standard textbook Bohr radius for the electron, a_o . Recall that

$$a_o = \frac{4\pi\epsilon_o\hbar^2}{m_o q^2}$$

Therefore the bulk exciton Bohr radius can be re-written as

$$a_B = \frac{4\pi\epsilon_o\hbar^2}{m_o q^2} \frac{\epsilon m_o}{\mu}$$

where ϵ is the relative dielectric constant of the semiconductor and μ is the reduced mass of the exciton. This gives

$$\boxed{a_B = \left(\frac{\epsilon m_o}{\mu}\right) a_o} \quad (4.1)$$

where a_o is everyone's favorite textbook Bohr radius ($a_o = 0.528\text{\AA}$). This is a more convenient way to go about calculating the exciton Bohr radius of something.

Note that we have omitted something pretty important. In real life the relative dielectric constant, ϵ , is frequency dependent. It isn't readily apparent which ϵ to use in the above expressions since there is the static dielectric constant, the high frequency dielectric constant ϵ_∞ and anything inbetween. Generally speaking, use the dielectric constant at the same frequency corresponding to the energy required to create the exciton.

Next we want the energy of the exciton. The total energy of the exciton relative to its ionization limit is the difference between its kinetic energy and its Coulomb (attractive) potential energy.

$$\epsilon_{tot} = \frac{1}{2}mv^2 - \frac{q^2}{4\pi\epsilon\epsilon_o r}$$

Since we previously saw in deriving the Bohr radius that

$$\begin{aligned}\frac{mv^2}{r} &= \frac{q^2}{4\pi\epsilon\epsilon_0 r^2} \\ mv^2 &= \frac{q^2}{4\pi\epsilon\epsilon_0 r}\end{aligned}$$

so that

$$\begin{aligned}\epsilon_{tot} &= \frac{1}{2} \left(\frac{q^2}{4\pi\epsilon\epsilon_0 r} \right) - \left(\frac{q^2}{4\pi\epsilon\epsilon_0 r} \right) \\ &= -\frac{1}{2} \left(\frac{q^2}{4\pi\epsilon\epsilon_0 r} \right)\end{aligned}$$

But recall that we just found that in a bound exciton

$$r = a_B = \frac{4\pi\epsilon\epsilon_0 \hbar^2 n^2}{\mu q^2}$$

Note that many times you will see the exciton Bohr radius written as a_X . Same thing.

Then we get our desired energy

$$\boxed{\epsilon_{tot} = -\frac{1}{2} \left(\frac{\mu q^4}{(4\pi\epsilon\epsilon_0)^2 \hbar^2} \right) \frac{1}{n^2}} \quad (4.2)$$

Again, you will often see this expressed in many texts in terms of the Rydberg and the exciton Rydberg (R_X). To express this like everyone else

$$\epsilon_{tot} = -\frac{1}{2} \left(\frac{m_o q^4}{(4\pi\epsilon_0)^2 \hbar^2} \right) \left(\frac{\mu}{\epsilon^2 m_o} \right) \frac{1}{n^2}$$

Now recognize that the first term in the expression is our textbook Rydberg $R = \frac{1}{2} \left(\frac{m_o q^4}{(4\pi\epsilon_0)^2 \hbar^2} \right)$ which refers to an energy of 13.6 eV. So we now have

$$\epsilon_{tot} = -R \left(\frac{\mu}{\epsilon^2 m_o} \right) \frac{1}{n^2}$$

Next define the exciton Rydberg as the product $R_X = R \left(\frac{\mu}{\epsilon^2 m_o} \right)$ to get the desired textbook expression

$$\boxed{\epsilon_{tot} = -\frac{R_X}{n^2}} \quad (4.3)$$

Check me

You can check that indeed the Rydberg is correctly defined. Start with our above definition and simplify it a little to get

$$R = \frac{m_o q^4}{8\epsilon_o^2 h^2}$$

Numerically we have

$$R = \frac{(9.11 \times 10^{-31})(1.602 \times 10^{-19})^4}{8(8.85 \times 10^{-12})^2(6.62 \times 10^{-34})^2}$$

all in SI units. Evaluate this to get

$$R = 2.185 \times 10^{-18}$$

The units are Joules. If so then divide this by 1.602×10^{-19} J/eV to get

$$R = 13.6eV$$

So we did things right.

In terms of units to prove that the last number was really in units of joules we have

$$\begin{aligned} & \frac{kgc^4}{\frac{F^2}{m^2}j^2s^2} \\ &= \left(\frac{kgm^2}{s^2}\right) \frac{c^4}{F^2j^2} \\ &= \frac{jc^4}{j^2F^2} \\ &= \frac{c^4}{jF^2} \\ &= \frac{c^4V^2}{jc^2} \\ & \text{since } F = \frac{c}{V} \\ &= \frac{c^2V^2}{j} \\ &= \frac{j^2}{j} \\ & \text{since } j = cv \\ &= j \end{aligned}$$

Ok, the units work out. Great.

Finally we want the binding energy of the exciton. This is the difference in energy between the electron-hole pair in a given orbit (say $n = 1$) and being at an infinite separation ($n = \infty$). If

$$\epsilon_{tot} = -R \left(\frac{\mu}{\epsilon^2 m_o} \right) \frac{1}{n^2}$$

then $\epsilon_{bind} = \epsilon_{\infty} - \epsilon_{n=1}$ giving

$$\epsilon_{bind} = -R \left(\frac{\mu}{\epsilon^2 m_o} \right) \left(\frac{1}{n_{\infty}^2} - \frac{1}{n_1^2} \right)$$

For the lowest orbit this gives

$$\boxed{\epsilon_{bind} = R \left(\frac{\mu}{\epsilon^2 m_o} \right)} \quad (4.4)$$

In general

$$\boxed{\epsilon_{bind} = \frac{mq^4}{2(4\pi\epsilon\epsilon_o)^2 \hbar^2} \left(\frac{1}{n^2} \right)} \quad (4.5)$$

Exciton Schrodinger equation

Excitons can be described as quasi particles and hence there is a quantum mechanical description for their energies and wavefunctions. We begin with the Hamiltonian for two independent particles (an electron and a hole) coupled by a common Coulomb term and then re-express the equation into one for the bound exciton.

In one dimension we have

$$\left[-\frac{\hbar^2}{2m_e} \frac{d^2}{dx_e^2} - \frac{\hbar^2}{2m_h} \frac{d^2}{dx_h^2} - \frac{e^2}{4\pi\epsilon\epsilon_o(r_e - r_h)} \right] \Psi_e \Psi_h = \epsilon \Psi_e \Psi_h$$

More generally however we have

$$\left[-\frac{\hbar^2}{2m_e} \nabla_e^2 - \frac{\hbar^2}{2m_h} \nabla_h^2 - \frac{e^2}{4\pi\epsilon\epsilon_o(r_e - r_h)} \right] \Psi_e \Psi_h = \epsilon \Psi_e \Psi_h$$

To begin converting this into an exciton picture let

$$\begin{aligned} R &= \frac{m_e r_e}{m_e + m_h} + \frac{m_h r_h}{m_e + m_h} \\ &= \frac{m_e r_e}{M} + \frac{m_h r_h}{M} \end{aligned}$$

$$\boxed{R = \frac{m_e r_e}{M} + \frac{m_h r_h}{M}} \quad (4.6)$$

where $M = m_e + m_h$ is the total mass of the exciton. This is the center of mass coordinate. Next let

$$\boxed{r = r_e - r_h} \quad (4.7)$$

be the relative electron hole coordinate.

It can be shown that

$$\boxed{r_e = R + \frac{\mu r}{m_e}} \quad (4.8)$$

$$\boxed{r_h = R - \frac{\mu r}{m_h}} \quad (4.9)$$

where $\mu = \frac{m_e m_h}{(m_e + m_h)}$ is the reduced mass.

“Proof”, First expression

$$\begin{aligned} r_e &= R + \frac{\mu r}{m_e} \\ &= \frac{m_e r_e}{M} + \frac{m_h r_h}{M} + \frac{m_e m_h}{M m_e} (r_e - r_h) \\ &= \frac{m_e r_e}{M} + \frac{m_h r_h}{M} + \frac{m_h}{M} (r_e - r_h) \\ &= \frac{m_e r_e}{M} + \frac{m_h r_h}{M} + \frac{m_h r_e}{M} - \frac{m_h r_h}{M} \\ &= \frac{m_e r_e}{M} + \frac{m_h r_e}{M} \\ &= \left(\frac{m_e + m_h}{M} \right) r_e \\ &= r_e \end{aligned}$$

“Proof”, Second expression

$$\begin{aligned} r_h &= R - \frac{\mu r}{m_h} \\ &= \frac{m_e r_e}{M} + \frac{m_h r_h}{M} - \frac{m_e m_h}{M m_h} (r_e - r_h) \\ &= \frac{m_e r_e}{M} + \frac{m_h r_h}{M} - \frac{m_e r_e}{M} + \frac{m_e r_h}{M} \\ &= \frac{m_h r_h}{M} + \frac{m_e r_h}{M} \\ &= \left(\frac{m_e + m_h}{M} \right) r_h \\ &= r_h \end{aligned}$$

By the chain rule we can now re-express ∇_e and ∇_h as follows since both r_e and r_h are seen to depend on both r and R .

First expression

$$\begin{aligned}\nabla_e = \frac{\partial}{\partial r_e} &= \frac{\partial}{\partial r} \frac{\partial r}{\partial r_e} + \frac{\partial}{\partial R} \frac{\partial R}{\partial r_e} \\ &= \frac{\partial}{\partial r}(1) + \frac{\partial}{\partial R} \left(\frac{m_e}{M} \right) \\ &= \frac{\partial}{\partial r} + \left(\frac{m_e}{M} \right) \frac{\partial}{\partial R}\end{aligned}$$

yielding

$$\boxed{\nabla_e = \nabla_r + \frac{m_e}{M} \nabla_R} \quad (4.10)$$

Second expression

$$\begin{aligned}\nabla_h = \frac{\partial}{\partial r_h} &= \frac{\partial}{\partial r} \frac{\partial r}{\partial r_h} + \frac{\partial}{\partial R} \frac{\partial R}{\partial r_h} \\ &= \frac{\partial}{\partial r}(-1) + \frac{\partial}{\partial R} \left(\frac{m_h}{M} \right) \\ &= -\frac{\partial}{\partial r} + \frac{m_h}{M} \frac{\partial}{\partial R}\end{aligned}$$

yielding

$$\boxed{\nabla_h = -\nabla_r + \frac{m_h}{M} \nabla_R} \quad (4.11)$$

Put everything together now

Back to our original expression

$$\left[-\frac{\hbar^2}{2m_e} \nabla_e^2 - \frac{\hbar^2}{2m_h} \nabla_h^2 - \frac{e^2}{4\pi\epsilon\epsilon_o(r_e - r_h)} \right] \Psi_{tot} = \epsilon \Psi_{tot}$$

where we will flip this Hamiltonian to something in terms of r and R only.

$$\left[-\frac{\hbar^2}{2m_e} \left(\nabla_r + \frac{m_e}{M} \nabla_R \right) \left(\nabla_r + \frac{m_e}{M} \nabla_R \right) - \frac{\hbar^2}{2m_h} \left(-\nabla_r + \frac{m_h}{M} \nabla_R \right) \left(-\nabla_r + \frac{m_h}{M} \nabla_R \right) - \frac{e^2}{4\pi\epsilon\epsilon_0 r} \right] \Psi_{tot} = \epsilon \Psi_{tot}$$

$$\left[-\frac{\hbar^2}{2m_e} \left(\nabla_r^2 + \frac{m_e}{M} \nabla_r \nabla_R + \frac{m_e}{M} \nabla_R \nabla_r + \left(\frac{m_e}{M} \right)^2 \nabla_R^2 \right) - \frac{\hbar^2}{2m_h} \left(\nabla_r^2 - \frac{m_h}{M} \nabla_r \nabla_R - \frac{m_h}{M} \nabla_R \nabla_r + \left(\frac{m_h}{M} \right)^2 \nabla_R^2 \right) - \frac{e^2}{4\pi\epsilon\epsilon_0 r} \right] \Psi_{tot} = \epsilon \Psi_{tot}$$

$$\left[\left(-\frac{\hbar^2}{2m_e} - \frac{\hbar^2}{2m_h} \right) \nabla_r^2 + \left(-\frac{\hbar^2}{2m_e} \left(\frac{m_e}{M} \right)^2 \nabla_R^2 - \frac{\hbar^2}{2m_h} \left(\frac{m_h}{M} \right)^2 \nabla_R^2 \right) - \frac{\hbar^2}{2m_e} \left(\frac{m_e}{M} \right) \nabla_r \nabla_R - \frac{\hbar^2}{2m_e} \left(\frac{m_e}{M} \right) \nabla_R \nabla_r + \frac{\hbar^2}{2m_h} \left(\frac{m_h}{M} \right) \nabla_r \nabla_R + \frac{\hbar^2}{2m_h} \left(\frac{m_h}{M} \right) \nabla_R \nabla_r - \frac{e^2}{4\pi\epsilon\epsilon_0 r} \right] \Psi_{tot} = \epsilon \Psi_{tot}$$

$$\left[-\frac{\hbar^2}{2} \left(\frac{1}{m_e} + \frac{1}{m_h} \right) \nabla_r^2 - \frac{\hbar^2}{2M} \left(\frac{m_e \nabla_R^2}{M} + \frac{m_h \nabla_R^2}{M} \right) - \frac{\hbar^2}{2M} \nabla_r \nabla_R - \frac{\hbar^2}{2M} \nabla_R \nabla_r + \frac{\hbar^2}{2M} \nabla_r \nabla_R + \frac{\hbar^2}{2M} \nabla_R \nabla_r - \frac{e^2}{4\pi\epsilon\epsilon_0 r} \right] \Psi_{tot} = \epsilon \Psi_{tot}$$

$$\left[-\frac{\hbar^2(m_e + m_h)}{2m_e m_h} \nabla_r^2 - \frac{\hbar^2}{2M} \nabla_R^2 - \frac{e^2}{4\pi\epsilon\epsilon_0 r} \right] \Psi_{tot} = \epsilon \Psi_{tot}$$

$$\left[-\frac{\hbar^2}{2\mu} \nabla_r^2 - \frac{\hbar^2}{2M} \nabla_R^2 - \frac{e^2}{4\pi\epsilon\epsilon_0 r} \right] \Psi(r)\Psi(R) = \epsilon \Psi(r)\Psi(R)$$

Since this is a separable problem, let $\epsilon = \epsilon_{cm} + \epsilon_{rel}$ giving

$$\begin{aligned} & \left[-\frac{\hbar^2}{2\mu} \nabla_r^2 - \frac{\hbar^2}{2M} \nabla_R^2 - \frac{e^2}{4\pi\epsilon\epsilon_0 r} - \epsilon_{cm} - \epsilon_{rel} \right] \Psi(r)\Psi(R) = 0 \\ & \left[\left(-\frac{\hbar^2}{2\mu} \nabla_r^2 - \frac{e^2}{4\pi\epsilon\epsilon_0 r} - \epsilon_{rel} \right) + \left(-\frac{\hbar^2}{2M} \nabla_R^2 - \epsilon_{cm} \right) \right] \Psi(r)\Psi(R) = 0 \\ & -\Psi(R) \frac{\hbar^2}{2\mu} \nabla_r^2 \Psi(r) - \epsilon' \Psi(r)\Psi(R) - \Psi(r) \frac{\hbar^2}{2M} \nabla_R^2 \Psi(R) - \epsilon_{cm} \Psi(r)\Psi(R) = 0 \end{aligned}$$

where $\epsilon' = \frac{e^2}{4\pi\epsilon\epsilon_0 r} - \epsilon_{rel}$. Next, divide by $\Psi(r)\Psi(R)$ to get

$$-\frac{1}{\Psi(r)} \frac{\hbar^2}{2\mu} \nabla_r^2 \Psi(r) - \epsilon' - \frac{1}{\Psi(R)} \frac{\hbar^2}{2M} \nabla_R^2 \Psi(R) - \epsilon_{cm} = 0$$

This gives you two separate equations. One which depends only on r (left one) and one which depends only on R (right one). So you get the two separate equations for the exciton relative motion and the center of mass motion. Both are solved independently. Solve for the eigenvalues of the relative expression to get the energies of the bound electron-hole pair.

$$\boxed{\left[-\frac{\hbar^2}{2\mu} \nabla_r^2 - \frac{e^2}{4\pi\epsilon\epsilon_0 r} \right] \Psi(r) = \epsilon_{rel} \Psi(r)} \quad (4.12)$$

$$\boxed{\left[-\frac{\hbar^2}{2M} \nabla_R^2 \right] \Psi(R) = \epsilon_{cm} \Psi(R)} \quad (4.13)$$

Exercises

1. Provide a value for the thermal energy, kT , in units of meV at these temperatures, 300K , 77K , 10K and 4K . Low temperature experiments are commonly conducted at 4K and 77K . Any suggestions as to why?
2. GaAs has a zinc blende crystal structure with a lattice constant of $a = 0.56\text{nm}$, a dielectric constant of $\epsilon = 12.8$, and electron/hole effective masses of $m_e = 0.067m_o$ and $m_h = 0.2m_o$. Provide the following: (a) the bulk exciton Bohr radius of this material. (b) the number of unit cells contained within the lowest exciton orbit $n = 1$. (c) The highest temperature at which stable excitons are observable.

Confinement regimes

At the heart of the binding energy is the Coulomb attraction between the oppositely charged electron and hole. Recall that this Coulomb term is proportional to $\frac{1}{r}$. Now it will be shown that the confinement energy associated with either the electron or hole is proportional to $\frac{1}{r^2}$. It's clear that the latter term will grow faster than the former when the size of a material becomes small. So in bulk materials, the Coulomb term and associated exciton is important. By contrast, in nanomaterials the exciton begins to experience competition from the “confinement” of individual electrons and holes.

There are three size (confinement) regimes that one must consider in real life. They are referred to as the strong, intermediate and weak confinement regimes. More about confinement will be illustrated in the following chapters. However, for now note that life isn't all that simple.

- Strong confinement: $a < a_{B,e}, a_{B,h}$
- Intermediate confinement: $a_{B,h} < a < a_{B,e}$
- Weak confinement: $a > a_{B,e}, a_{B,h}$

Weak confinement

In this case $a > a_{B,e}, a_{B,h}$. As a consequence the binding energy of the exciton is larger than the individual confinement energies of the electron and hole. This is definitely the case with bulk materials and with “large” nano scale materials. The energy of optical transitions here is basically the band gap energy minus the exciton binding energy.

Intermediate confinement

In this case we have the physical radius of the material smaller than one carrier's individual Bohr radius but larger than the individual Bohr radius of the other carrier. Because the effective mass of the electron is smaller than the effective mass of the hole, one often sees the criteria, $a_{B,h} < a < a_{B,e}$.

We have

$$a_{B,e} = \frac{4\pi\epsilon\epsilon_0\hbar^2}{m_e q^2}$$

$$a_{B,h} = \frac{4\pi\epsilon\epsilon_0\hbar^2}{m_h q^2}$$

Strong confinement

This is commonly seen in small nanomaterials. Here the physical size of the system is smaller than both individual Bohr radii of the electron and hole, $a < a_{B,e}, a_{B,h}$. In this regime the optical properties of the material are dominated by quantum confinement effects of the electron and hole. This will be seen in the following Chapters.

Exercises

1. For InP which has the following parameters, $m_e = 0.077m_o$, $m_h = 0.2m_o$ and $\epsilon = 12.4$, what are the individual Bohr radii of the electron and hole. Provide physical size ranges for this material to be in the strong, intermediate and weak confinement regimes.
2. For CdSe which has the following parameters, $m_e = 0.13m_o$, $m_h = 0.45m_o$ and $\epsilon = 9.4$, what are the individual Bohr radii of the electron and hole. Provide physical size ranges for this material to be in the strong, intermediate and weak confinement regimes.

Chapter 5

Quantum mechanics review

Wavefunctions and such

Given the deBroglie wave-particle duality it turns out that we can mathematically express a particle like a wave using a “wavefunction” (usually denoted Ψ). This wavefunction replaces the classical concept of a trajectory and contains all the dynamical information about a system that you can know. Usually much of the work we will do here will be to find out what this wavefunction looks like given certain constraints on the system (called boundary conditions).

There is a probabilistic interpretation of the wavefunction called the Born interpretation. In this respect

$$\begin{aligned} |\Psi|^2 &= \Psi^* \Psi \text{ is considered as a probability density} \\ |\Psi|^2 dx &= \Psi^* \Psi dx \text{ is considered as a probability} \end{aligned}$$

Through these quantities one can determine the probability that the particle is somewhere. Note that from a physical perspective only $|\Psi|^2$ has some physical significance. Ψ can be real or imaginary or negative but $|\Psi|^2$ will be real and positive.

One consequence of the probabilistic interpretation of the wavefunction is that the wavefunction must be normalized.

$$\int |\Psi|^2 dx = 1$$

This is because the probability of finding the particle somewhere must be unity. So generally you will see that in front of $\Psi(x, t)$ will be a constant N which ensures normalization is met. Physical particles therefore have

normalizable wavefunctions, non-physical ones have wavefunctions that are not normalizable.

There are some important mathematical properties (or mathematical constraints) of the wavefunction described as follows

- Ψ must be “well behaved” in general
- Ψ must be finite (it does not blow up or show singularities. This constraint of course is related to being able to normalize the wavefunction. Can't do it if it goes nuts.)
- Ψ must be single valued. Obviously it doesn't make sense to have multiple probabilities at a given position.
- Ψ must be continuous.
- Ψ must have a first derivative (related to previous statement)
- Ψ must have a second derivative (related to previous, previous statement)
- Ψ must be integrable and hence normalizable (basically a consequence of 1 and 2)

Other properties to remember include

- normalization $\int |\Psi|^2 dx = 1$
- orthogonality $\int \Psi_1 \Psi_2 dx = 0$. Thus these two wavefunctions Ψ_1 and Ψ_2 have no mutual spatial overlap.

Observables

All dynamical information about the particle or system is contained in the wavefunction. This includes observables such as

- position
- momentum
- angular momentum
- energy

So while the wavefunction itself is not a measurable quantity, these observables are in principle.

In quantum mechanics, just like the particle is represented by a mathematical wavefunction, observables are represented by mathematical "operators".

An operator will act on a wavefunction to transform it into another function. Some examples of operators include

- x (multiply by x)
- $\frac{\partial}{\partial x}$ (take the partial derivative with respect to x)
- $\frac{\partial}{\partial t}$ (take the partial derivative with respect to time)

You will sometimes see operators denoted by a little hat on top of it. For example

- $\hat{x} = x$ operator (multiply by x)
- $\hat{y} = y$ operator (multiply by y)
- and so forth

Correspondence principle

In quantum mechanics all physical observables have corresponding operators. Two fundamental operators are the position operator \hat{x} and the momentum operator \hat{p} .

- position operator, $\hat{x} = x$. read as multiply by x
- momentum operator, $\hat{p} = -i\hbar\nabla = -i\hbar\frac{\partial}{\partial x}$ (read as take the derivative with respect to x and multiply by $-i\hbar$)

All other operators can be constructed from combinations of these two fundamental operators. For example, for the kinetic energy operator (used to extract out the energy of the particle through its wavefunction)

$$\begin{aligned}
 T &= \frac{p^2}{2m} \\
 &= \frac{1}{2m} \left(-i\hbar \frac{\partial}{\partial x} \right) \left(-i\hbar \frac{\partial}{\partial x} \right) \\
 &= \frac{1}{2m} \left(-\hbar^2 \frac{\partial^2}{\partial x^2} \right) \\
 &= -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2}
 \end{aligned}$$

This leads to

$$\boxed{T = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2}} \quad (5.1)$$

For the potential energy $V(x)$ we have

$$\boxed{V(x) = V(x)} \quad (5.2)$$

Read this as multiply by $V(x)$.

For the total energy in 1D we have the operator $H = T + V$ where H is called the Hamiltonian operator.

$$\begin{aligned} H &= T + V \\ H &= -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x) \end{aligned}$$

Therefore

$$\boxed{H = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x)} \quad (5.3)$$

This is the total energy operator called the Hamiltonian operator or Hamiltonian for short.

For the total energy in 3D

$$\begin{aligned} H &= \frac{p_x^2}{2m} + \frac{p_y^2}{2m} + \frac{p_z^2}{2m} + V(x, y, z) \\ &= \frac{1}{2m} \left(-i\hbar \frac{\partial}{\partial x} \right) \left(-i\hbar \frac{\partial}{\partial x} \right) + \frac{p_y^2}{2m} + \frac{p_z^2}{2m} + V(x, y, z) \\ &= -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} - \frac{\hbar^2}{2m} \frac{\partial^2}{\partial y^2} - \frac{\hbar^2}{2m} \frac{\partial^2}{\partial z^2} + V(x, y, z) \end{aligned}$$

This results in

$$\boxed{H = -\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) + V(x, y, z)} \quad (5.4)$$

$$\boxed{H = -\frac{\hbar^2}{2m} \nabla^2} \quad (5.5)$$

where ∇^2 is called the Laplacian or “Del-squared” and

$$\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right)$$

In addition $\Delta = \nabla^2$ (French)

So to summarize you can see that other operators can be built up from the fundamental operators \hat{x} and \hat{p} .

Eigenvalues and eigenfunctions

Now a special class of function which when operated on yields itself again (may be multiplied by a constant or something) is called an eigenfunction (or eigenvector) of the operator. The constant that falls out is called an eigenvalue of the operator. For example, you could have

$$\frac{\partial}{\partial x}(e^{\alpha x}) = \alpha e^{\alpha x}$$

Here the function $e^{\alpha x}$ is an eigenfunction of the $\frac{\partial}{\partial x}$ operator and the constant α is an eigenvalue of the operator.

Expectation values

The average of mean value of an observable A is by definition

$$\langle A \rangle = \int \Psi^* \hat{A} \Psi dx \quad (5.6)$$

This is called “sandwiching”.

In the special case that Ψ is an eigenfunction of the operator A

$$\hat{A}\Psi = a\Psi$$

Therefore

$$\begin{aligned} \langle A \rangle &= \int \Psi^* \hat{A} \Psi dx \\ &= \int \Psi^* a \Psi dx \\ &= a \int \Psi^* \Psi dx \\ &= a \text{ since by normalization the integral is } 1 \end{aligned}$$

Therefore

$$\langle A \rangle = a$$

in the special case that Ψ is an eigenfunction of A .

Dirac Bra-Ket notation (a quick aside)

Dirac suggested a shorthand notation for writing and dealing with wavefunctions, operators and such. This new notation turns out to be easier and faster than the default calculus approach.

In this notation

- A wavefunction Ψ is denoted by $|\Psi\rangle$ and is called a ket
- The complex conjugate Ψ^* is denoted by $\langle\Psi|$ and is called a bra

Bras and kets follow the rules of linear algebra

- $\langle a|b\rangle = \langle b|a\rangle^*$
- $|aA\rangle = a|A\rangle$ if a is a constant (a scalar)
- $\langle aA| = a^* \langle A|$

Likewise

- $\langle A|bB + cC\rangle = b \langle A|B\rangle + c \langle A|C\rangle$

Integrals are represented in the following fashion

$$\begin{aligned} \int \Psi_1^* \Psi_1 dx &= \langle 1|1\rangle = 1 \text{ normalization} \\ \int \Psi_2^* \Psi_1 dx &= \langle 2|1\rangle = 0 \text{ orthogonality} \\ \int \Psi_1^* \hat{A} \Psi_1 dx &= \langle 1|\hat{A}|1\rangle \text{ Sandwiching} \\ \int \Psi_2^* \hat{A} \Psi_1 dx &= \langle 2|\hat{A}|1\rangle \text{ Sandwiching} \end{aligned}$$

and so forth.

Operator math

Just as wavefunctions have to obey certain constraints, they are important restrictions on operators that represent observable quantities. This is called “linearity” and “hermiticity”

Linearity:

An operator \hat{A} is linear for any constants α and β and wavefunctions $|f\rangle$ and $|g\rangle$ if

$$\hat{A}(\alpha|f\rangle + \beta|g\rangle) = \alpha\hat{A}|f\rangle + \beta\hat{A}|g\rangle$$

So the operator acting on the whole superposition wavefunction $\Psi_{tot} = \alpha|f\rangle + \beta|g\rangle$ is the same as the operator acting individually on the parts.

Hermiticity:

An operator \hat{A} is Hermitian if its integrals obey the following relationship for all valid $|f\rangle$ and $|g\rangle$

$$\begin{aligned} \int \Psi_f^* \hat{A} \Psi_g dx &= \int \Psi_g^* \hat{A} \Psi_f dx \\ \langle f | \hat{A} | g \rangle^* &= \langle g | \hat{A} | f \rangle \end{aligned}$$

This restriction ensures that eigenvalues of \hat{A} are all real-valued (as opposed to complex).

If its going to be observable in real life the number has to be real (not complex).

“Proof”:

$|k\rangle$ is an eigenfunction of \hat{A} with eigenvalue k . Because \hat{A} is Hermitian

$$\begin{aligned} \langle k | \hat{A} | k \rangle^* &= \langle k | \hat{A} | k \rangle \\ k^* &= k \end{aligned}$$

Therefore k is real valued.

Furthermore, eigenfunctions of a Hermitian operator associated with different eigenvalues are all orthogonal.

“Proof”:

$$\begin{aligned} \langle j | \hat{A} | k \rangle^* &= \langle k | \hat{A} | j \rangle \\ k \langle j | k \rangle^* &= j \langle k | j \rangle \quad \text{where } \langle k | j \rangle = \langle j | k \rangle^* \\ (k - j) \langle k | j \rangle &= 0 \end{aligned}$$

Since $k \neq j$ (they are different eigenvalues)

$$\langle k | j \rangle = 0$$

More on operators

$$\hat{A}|\Psi\rangle = |\hat{A}\Psi\rangle = |\Psi'\rangle$$

so that

$$\begin{aligned} |\Psi'\rangle &\longleftrightarrow \langle\Psi'| \\ |\hat{A}\Psi\rangle &\longleftrightarrow \langle\hat{A}\Psi| \\ \hat{A}|\Psi\rangle &\longleftrightarrow \langle\Psi|\hat{A}^\dagger \end{aligned}$$

where \hat{A}^\dagger is called the adjoint of \hat{A} operator.

More games with the adjoint

1)

$$\begin{aligned} \langle\phi|\hat{A}|\Psi\rangle &= \langle\phi|\hat{A}\Psi\rangle \\ &= \langle\hat{A}\Psi|\phi\rangle^* \\ &= \langle\Psi|\hat{A}^\dagger|\phi\rangle^* \end{aligned}$$

Therefore

$$\boxed{\langle\phi|\hat{A}|\Psi\rangle = \langle\Psi|\hat{A}^\dagger|\phi\rangle^*} \quad (5.7)$$

of course if $A = A^\dagger$ then A is Hermitian.

2)

$$\boxed{\langle\phi|\hat{A}|\Psi\rangle = \langle\phi|\hat{A}\Psi\rangle} \quad (5.8)$$

$$\boxed{\langle\phi|\hat{A}|\Psi\rangle = \langle\hat{A}^\dagger\phi|\Psi\rangle} \quad (5.9)$$

The first expression reflects operation on the right. The second expression reflects operating on the left.

3)

$$\boxed{(\hat{A}\hat{B})^\dagger = \hat{B}^\dagger\hat{A}^\dagger} \quad (5.10)$$

Proof: Let

$$\begin{aligned} |\phi\rangle &= (\hat{A}\hat{B})|\Psi\rangle \\ &= \hat{A}|\hat{B}\Psi\rangle \end{aligned}$$

Take adjoint now

$$\begin{aligned}
 \hat{A}\hat{B}|\Psi\rangle &\rightarrow \langle B\Psi|\hat{A}^\dagger \\
 \hat{A}\hat{B}|\Psi\rangle &\rightarrow \langle \Psi|\hat{B}^\dagger\hat{A}^\dagger \\
 &\text{Call } \hat{C} = \hat{A}\hat{B} \text{ or } \hat{C}^\dagger = \hat{B}^\dagger\hat{A}^\dagger \text{ resulting in} \\
 \hat{C}|\Psi\rangle &\rightarrow \langle \Psi|\hat{C}^\dagger
 \end{aligned}$$

This ends up with

$$\boxed{(\hat{A}\hat{B})^\dagger = \hat{B}^\dagger\hat{A}^\dagger} \quad (5.11)$$

4) If \hat{A} and \hat{B} are Hermitian ($\hat{A} = \hat{A}^\dagger$ and $\hat{B} = \hat{B}^\dagger$)

$$(\hat{A}\hat{B})^\dagger = \hat{B}^\dagger\hat{A}^\dagger = \hat{B}\hat{A} = \hat{A}\hat{B} + [\hat{B}, \hat{A}]$$

This results in the relationship where $[\hat{B}, \hat{A}]$ is called the “commutator” of the two operators. This will be introduced in the next section.

$$\boxed{(\hat{A}\hat{B})^\dagger = \hat{A}\hat{B} + [\hat{B}, \hat{A}]} \quad (5.12)$$

Commutators

A wavefunction Ψ with a characteristic well-defined value of some observable quantity is an eigenfunction of the corresponding operator. However, this Ψ does not have to have a characteristic value of any other observable.

For a Ψ to have a characteristic value of 2 observables simultaneously, the corresponding operators must “commute”

Definition of “commute”

The action of the two operators say \hat{A} and \hat{B} taken in succession on any Ψ is identical to the action of the operators taken in reverse order.

$$\hat{A}\hat{B} = \hat{B}\hat{A} \quad (5.13)$$

or equivalently that the “commutator” of the two operators equals zero.

Definition of “commutator”

$$[\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A}$$

A special case is

$$[\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A} = 0$$

which occurs when \hat{A} and \hat{B} commute.

Now if such a Ψ exists then it is an eigenfunction of both operators. For example

$$\hat{A}(\hat{B}|\beta\rangle) = \hat{A}\beta|\beta\rangle = \beta\hat{A}|\beta\rangle$$

But now if $\hat{A}\hat{B} = \hat{B}\hat{A}$ then

$$\begin{aligned}\hat{A}\hat{B}|\beta\rangle &= \hat{B}\hat{A}|\beta\rangle \\ \beta\hat{A}|\beta\rangle &= \hat{B}(\hat{A}|\beta\rangle)\end{aligned}$$

which means that $(\hat{A}|\beta\rangle)$ is an eigenfunction of \hat{B} , call it $|\beta'\rangle$

$$\begin{aligned}&= \hat{B}|\beta'\rangle \\ &= \beta|\beta'\rangle\end{aligned}$$

This implies that $|\beta\rangle$ is an eigenfunction of \hat{A} . Hence $|\beta\rangle$ is an eigenfunction of both \hat{A} and \hat{B} if they commute.

Alternative derivation

Claim that if $|\beta\rangle$ is an eigenfunction of \hat{B}

$$\hat{B}|\beta\rangle = \beta|\beta\rangle$$

Then if \hat{A} and \hat{B} commute (i.e. $[\hat{B}, \hat{A}] = 0$), $|\hat{A}\beta\rangle$ is also an eigenfunction of \hat{B} with eigenvalue β . To illustrate:

$$\begin{aligned}\hat{B}|\hat{A}\beta\rangle &= \hat{B}\hat{A}|\beta\rangle = (\hat{A}\hat{B} + [\hat{B}, \hat{A}]|\beta\rangle \\ \text{since } [\hat{B}, \hat{A}] &= 0 \\ &= \hat{A}\hat{B}|\beta\rangle \\ &= \beta(\hat{A}|\beta\rangle) \\ &= \beta|\hat{A}\beta\rangle\end{aligned}$$

One therefore sees that

$$\boxed{\hat{B}|\hat{A}\beta\rangle = \beta|\hat{A}\beta\rangle} \quad (5.14)$$

Again this implies that $|\beta\rangle$ is an eigenfunction of \hat{A} .

Conversely, if \hat{A} and \hat{B} do not commute ($[\hat{A}, \hat{B}] \neq 0$), you cannot specify the eigenvalues of both simultaneously (only one or the other). This is the origin of the uncertainty principle, which we will discuss later in more detail.

To illustrate this locally however, \hat{x} and \hat{p} do not commute.

Proof: (note that when working with operators its easier to have an imaginary test function to work on. otherwise it is easy to get confused.)

$$\begin{aligned} [\hat{x}, \hat{p}]|\Psi\rangle &= (\hat{x}\hat{p} - \hat{p}\hat{x})|\Psi\rangle \\ &= \hat{x}\hat{p}|\Psi\rangle - \hat{p}\hat{x}|\Psi\rangle \\ &= -i\hbar\hat{x}\frac{d|\Psi\rangle}{dx} + i\hbar\frac{d}{dx}(x|\Psi\rangle) \\ &= i\hbar\left[-x\frac{d|\Psi\rangle}{dx} + x\frac{d|\Psi\rangle}{dx} + |\Psi\rangle\right] \\ &= i\hbar|\Psi\rangle \end{aligned}$$

Resulting in

$$[\hat{x}, \hat{p}]|\Psi\rangle = i\hbar|\Psi\rangle \quad (5.15)$$

or alternatively that

$$\boxed{[x, p] = i\hbar} \quad (5.16)$$

This value is non-zero showing that \hat{x} and \hat{p} do not commute. And as advertised earlier this will lead to the Uncertainty principle

$$\boxed{\Delta x \Delta p \geq \frac{\hbar}{2}} \quad (5.17)$$

Furthermore, given 2 commuting observables \hat{A} and \hat{B} it is always possible to construct an orthonormal basis with eigenstates common to both \hat{A} and \hat{B} .

Additional commutator math

Before we go on we will want to summarize some important commutator relations

- $[\hat{A}, \hat{A}] = 0$
- $[\hat{A}, \hat{B}] = -[\hat{B}, \hat{A}]$
- $[\hat{A}, \hat{B}\hat{C}] = [\hat{A}, \hat{B}]\hat{C} + \hat{B}[\hat{A}, \hat{C}]$
- $[\hat{A}\hat{B}, \hat{C}] = \hat{A}[\hat{B}, \hat{C}] + [\hat{A}, \hat{C}]\hat{B}$
- $[\hat{A}, \hat{B} + \hat{C}] = [\hat{A}, \hat{B}] + [\hat{A}, \hat{C}]$
- $[\hat{A}, b\hat{B}] = b[\hat{A}, \hat{B}]$
- $[\hat{A}, [\hat{B}, \hat{C}]] = [\hat{A}, \hat{B}\hat{C}] - [\hat{A}, \hat{C}\hat{B}]$
- $[\hat{A} + \hat{B}, \hat{C} + \hat{D}] = [\hat{A}, \hat{C}] + [\hat{A}, \hat{D}] + [\hat{B}, \hat{C}] + [\hat{B}, \hat{D}]$

Examples

Evaluate $[\hat{p}, \hat{x}^2]$. Remember to add some dummy wavefunction Ψ to help out.

$$\begin{aligned}
 [\hat{p}, \hat{x}\hat{x}]\Psi &= \hat{p}\hat{x}\hat{x}\Psi - \hat{x}^2\hat{p}\Psi \\
 \text{where } \hat{p} &= -i\hbar\frac{d}{dx} \\
 &= -i\hbar\frac{d}{dx}(\hat{x}^2\Psi) - \hat{x}^2\frac{d\Psi}{dx}(-i\hbar) \\
 &= -i\hbar\left[\hat{x}^2\frac{d\Psi}{dx} + 2\hat{x}\Psi\right] + i\hbar\hat{x}^2\frac{d\Psi}{dx} \\
 &= i\hbar\left[-\hat{x}^2\frac{d\Psi}{dx} - 2\hat{x}\Psi + \hat{x}^2\frac{d\Psi}{dx}\right] \\
 &= -2i\hbar\hat{x}\Psi
 \end{aligned}$$

Therefore we conclude that

$$\boxed{[\hat{p}, \hat{x}^2] = -2i\hbar\hat{x}} \quad (5.18)$$

Evaluate $[\hat{p}^2, \hat{x}]$. Remember to add some dummy wavefunction Ψ to help

out.

$$\begin{aligned}
[\hat{p}^2, \hat{x}] \Psi &= (\hat{p}^2 \hat{x} - \hat{x} \hat{p}^2) \Psi \\
&= \hat{p}^2 \hat{x} \Psi - \hat{x} \hat{p}^2 \Psi \\
&= -i\hbar \frac{d}{dx} \left[-i\hbar \frac{d}{dx} (x\Psi) \right] - \hat{x} \frac{d^2}{dx^2} (-i\hbar)^2 \Psi \\
&= -\hbar^2 \frac{d}{dx} \left[x \frac{d\Psi}{dx} + \Psi \right] + \hbar^2 x \frac{d^2\Psi}{dx^2} \\
&= -\hbar^2 \left[x \frac{d^2\Psi}{dx^2} + \frac{d\Psi}{dx} + \frac{d\Psi}{dx} \right] + \hbar^2 x \frac{d^2\Psi}{dx^2} \\
&= -\hbar^2 \left[x \frac{d^2\Psi}{dx^2} + 2 \frac{d\Psi}{dx} \right] + \hbar^2 x \frac{d^2\Psi}{dx^2} \\
&= -\hbar^2 x \frac{d^2\Psi}{dx^2} - 2\hbar^2 \frac{d\Psi}{dx} + \hbar^2 x \frac{d^2\Psi}{dx^2} \\
&= -2\hbar^2 \frac{d\Psi}{dx} \\
&= \left[-2\hbar^2 \frac{d}{dx} \right] \Psi \\
&= \left[\frac{2\hbar^2}{i\hbar} (-i\hbar) \frac{d}{dx} \right] \Psi \\
&= -2i\hbar \hat{p} \Psi
\end{aligned}$$

Resulting in

$$\boxed{[\hat{p}^2, \hat{x}] = -2i\hbar \hat{p}} \quad (5.19)$$

The time independent Schrödinger equation

The general form of the Schrodinger equation which we will need to solve many times is

$$\boxed{i\hbar \frac{\partial}{\partial t} \Psi(r, t) = H \Psi(r, t)} \quad (5.20)$$

where H is the Hamiltonian operator. Alternatively

$$i\hbar \frac{\partial}{\partial t} \Psi(r, t) = \left[-\frac{\hbar^2}{2m} \nabla^2 + V \right] \Psi(r, t)$$

where V is some generic potential.

This equation derived by Schrodinger describes the wavefunction of the particle and hence is a pretty important. The general form of the Schrodinger equation has time dependence as you see above. However it can be shown that there is a time independent form of the equation also called the Schrodinger equation. To distinguish these two equations the former is often referred to as the time-dependent Schrodinger equation. Note that one typically deals with the time-independent version for the first semester of quantum and then get into the time-dependent version in the second semester.

To obtain the time-independent version of the Schrodinger equation assume that V does not depend on time. It is only dependent on position $V \rightarrow V(r)$. We have

$$i\hbar \frac{\partial}{\partial t} \Psi(r, t) = \left(-\frac{\hbar^2 \nabla^2}{2m} + V \right) \Psi(r, t)$$

Note that the left side has time dependence while the right side only has a spatial dependence. Look for solutions of the form

$$\begin{aligned} \Psi(r, t) &= \Psi(r) f(t) \\ &= \Psi f \text{ shorthand expression} \end{aligned}$$

We now have

$$\begin{aligned} i\hbar \frac{\partial}{\partial t} (\Psi f) &= -\frac{\hbar^2 \nabla^2}{2m} \Psi f + V \Psi f \\ i\hbar \Psi \frac{\partial f}{\partial t} &= -\frac{\hbar^2 f}{2m} \nabla^2 \Psi + V \Psi f \\ \frac{i\hbar \Psi}{f} \frac{\partial f}{\partial t} &= \frac{1}{\Psi} \left[-\frac{\hbar^2 \nabla^2}{2m} \Psi + V \Psi \right] \end{aligned}$$

Note that the left hand side is independent of position and the right hand side is independent of time. Both sides must equal a constant (call it ϵ) in order to be consistent. It turns out that this constant is the system eigenenergy which is why we are calling it ϵ already.

Left hand side

$$\begin{aligned} \frac{i\hbar}{f} \frac{\partial f}{\partial t} &= \epsilon \\ \frac{1}{f} \frac{\partial f}{\partial t} &= \frac{\epsilon}{i\hbar} \\ \frac{1}{f} \frac{\partial f}{\partial t} &= \frac{-i\epsilon}{\hbar} \\ \ln f &= -\frac{i\epsilon t}{\hbar} \end{aligned}$$

note that there is a constant of integration
but we ignore it here

This results in

$$\boxed{f = e^{-\frac{i\epsilon t}{\hbar}}} \quad (5.21)$$

The form of the function is exponential but it has an imaginary exponent. So what does this mean? Well, it means that one has plane waves. Note that $e^{ikx} = \cos kx + i \sin kx$.

Right hand side

$$\begin{aligned} \frac{1}{\Psi} \left[-\frac{\hbar^2 \nabla^2 \Psi}{2m} + V\Psi \right] &= \epsilon \\ H\Psi &= \epsilon\Psi \end{aligned}$$

This expression is something that we are familiar with. Its our favorite eigenvalue and eigenvector problem.

Summary

So putting everything together we have the full form of the wavefunction (both the time-dependent and time-independent parts).

$$\boxed{\Psi = \Psi(r) e^{-\frac{i\epsilon t}{\hbar}}} \quad (5.22)$$

Recall that additional constraints on $\Psi(r)$ include

- $\Psi(r)$ is finite

- $\Psi(r)$ is continuous
- $\Psi'(r)$ is continuous

Furthermore, boundary conditions on the wavefunction will induce quantization on the energies ϵ as well as the mathematical form of the wavefunction. Here you will note that typically the first semester of quantum mechanics is spent solving the time independent Schrodinger equation for a number of different situation and boundary conditions. These problems are referred to as model problems and will be described in the next sections.

What do we mean

What exactly did we mean above. Well let's go back to the full time-dependent Schrodinger equation

$$i\hbar \frac{\partial}{\partial t} \Psi(r, t) = H\Psi(r, t)$$

where we just found that

$$\Psi(r, t) = \Psi(r) e^{-\frac{i\epsilon t}{\hbar}}$$

Replace this into the Schrodinger equation and evaluate.

$$\begin{aligned} i\hbar \frac{\partial}{\partial t} \left[\Psi(r) e^{-\frac{i\epsilon t}{\hbar}} \right] &= H\Psi(r) e^{-\frac{i\epsilon t}{\hbar}} \\ i\hbar \Psi(r) \frac{\partial}{\partial t} \left[e^{-\frac{i\epsilon t}{\hbar}} \right] &= H\Psi(r) e^{-\frac{i\epsilon t}{\hbar}} \\ i\hbar \left(-\frac{i\epsilon}{\hbar} \right) \Psi(r) e^{-\frac{i\epsilon t}{\hbar}} &= H\Psi(r) e^{-\frac{i\epsilon t}{\hbar}} \\ i\hbar \left(-\frac{i\epsilon}{\hbar} \right) \Psi(r) &= H\Psi(r) \end{aligned}$$

leaving

$$\boxed{H\Psi(r) = \epsilon\Psi(r)} \quad (5.23)$$

which is the time independent Schrodinger equation or our familiar eigenvalue and eigenvector problem. We will solve this equation a bunch of times for a number of model systems. Later on we will go back and deal more directly with the time dependent Schrodinger equation.

Some relevant applications of the time independent Schrodinger equation include

- particle in a box (absorption)
- harmonic oscillator (vibrations)
- rigid rotor (rotational spectra)

Chapter 6

Confinement

Quantum wells, wires and dots are often described using the analogy to a particle in a 1D box, a 2D box and a 3D box. This is because when the actual physical length scale of the system is smaller than the exciton Bohr radius or corresponding deBroglie wavelength (as we saw in the previous section), either or both the electron and hole experience confinement. In turn, the energies of the carrier along that dimension of the material are no longer continuous as in the case where there is no confinement. The appearance of discrete states is one of the fundamental signatures of nanomaterials. Since solving the Schrodinger equation of a carrier to find its eigenvalues and eigenfunctions involves using boundary conditions one can also immediately predict that the actual shape of a quantum well, wire or dot will also play a role in dictating the ordering and spacing of states. A nanowire will have a similar but different progression of states than a quantum dot (or nanocrystal). The same applies to quantum wells as well as more exotic shapes of nanostructures.

In this chapter we solve the simple analytical problems of a particle in a 1 dimensional rectangular box, a cylindrical wire and a sphere (particle in a spherical box) to illustrate the discreteness as well as progression of states in wells, wires and dots.

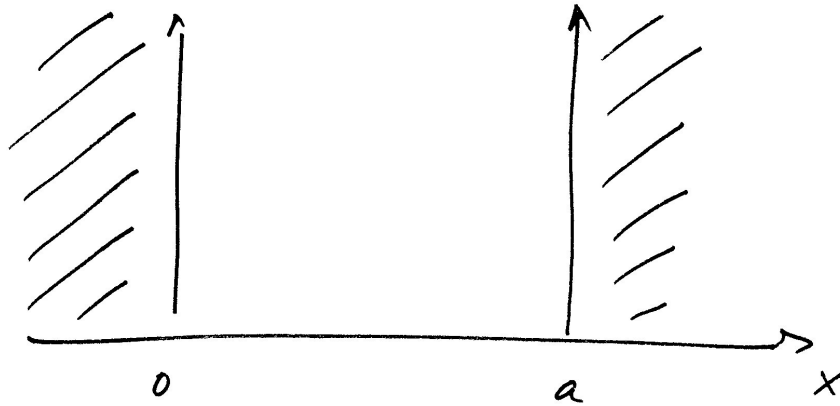


Figure 6.1: Cartoon of a one dimensional infinite barrier potential

1 Dimension of confinement

Particle in a 1D infinite box

The potential is

$$V(x) = \begin{cases} \infty & \text{if } x \leq 0 \\ 0 & \text{if } 0 < x < a \\ \infty & \text{if } x \geq a \end{cases}$$

The boundary conditions are

$$\Psi(0) = 0$$

$$\Psi(a) = 0$$

The Shrodinger equation to solve is

$$\boxed{-\frac{\hbar^2}{2m} \frac{d^2\Psi}{dx^2} + V\Psi = \varepsilon\Psi} \quad (6.1)$$

Rearrange to yield in the box region where $V = 0$

$$\frac{d^2\Psi}{dx^2} + k^2\Psi = 0 \quad (6.2)$$

where $k = \sqrt{\frac{2m\varepsilon}{\hbar^2}}$ General solutions are of the form

$$\Psi = Ae^{ikx} + Be^{-ikx} \quad (6.3)$$

Note that the first term describes a right going wave and the second describes a left going wave. Also if you need it $e^{ikx} = \cos(kx) + i\sin(kx)$ and $e^{-ikx} = \cos(kx) - i\sin(kx)$. Apply the boundary conditions now to simplify

$$\begin{aligned} \Psi(0) &= A + B \rightarrow B = -A \\ \Psi(a) &= Ae^{ika} - Ae^{-ika} = 0 \end{aligned}$$

This latter equation reduces to

$$2iA \frac{e^{ika} - e^{-ika}}{2i} = 0$$

$$2iA \sin(ka) = 0$$

where we employ the identities $\sin(kx) = \frac{e^{ikx} - e^{-ikx}}{2i}$ and $\cos(kx) = \frac{e^{ikx} + e^{-ikx}}{2}$. For this to be true and non-trivial ($A \neq 0$), $ka = n\pi$. This leads to

$$\sqrt{\frac{2m\varepsilon}{\hbar^2}} a = n\pi$$

The energy can be solved for, to give

$$\boxed{\varepsilon = \frac{n^2 \hbar^2}{8ma^2}} \quad (6.4)$$

where n is an integer. Now by normalizing the wavefunction

$$\Psi^* \Psi = 1$$

one basically gets the equation

$$N^2 \int_0^a \sin^2(kx) dx = 1$$

The integral can be evaluated by recalling that $\sin^2(kx) = \frac{1}{2}(1 - \cos(2kx))$. This is readily integrated to give $N = \sqrt{\frac{2}{a}}$ and the complete wavefunction as

$$\boxed{\Psi(x) = \sqrt{\frac{2}{a}} \sin(kx)} \quad (6.5)$$

Exercises

1. Estimate the first few energies ($n = 1, 2, 3$) for an electron in GaAs quantum wells of width 10 nm and 4 nm. Assume the mass is $0.067m_0$. Repeat the same calculation for a “heavy” hole ($m_{hh} = .5m_0$) and “light” hole ($m_{lh} = 0.082m_0$). In all cases assume an infinite box model.
2. Use Mathcad, Matlab, Mathematica or your favorite mathematical modeling program and numerically determine the first 10 energies of the particle in an infinite box (previous problem). Draw the corresponding wavefunctions as well.
3. Consider the following potential in Figure 6.2. Find the energy levels and eigenfunctions. Assume knowledge of the harmonic oscillator wavefunctions and energies. (look them up if you need to) This problem does not require any extensive work.
4. Using your knowledge of the energy level spacing for a particle in an infinite box as well as the energy level spacing in a harmonic oscillator predict whether the energy levels will converge or diverge for the following two potentials. That is as the energies get larger, do the spacings between states increase or decrease.

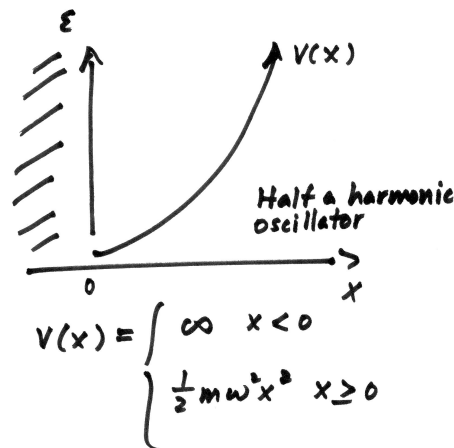


Figure 6.2: Half a harmonic oscillator

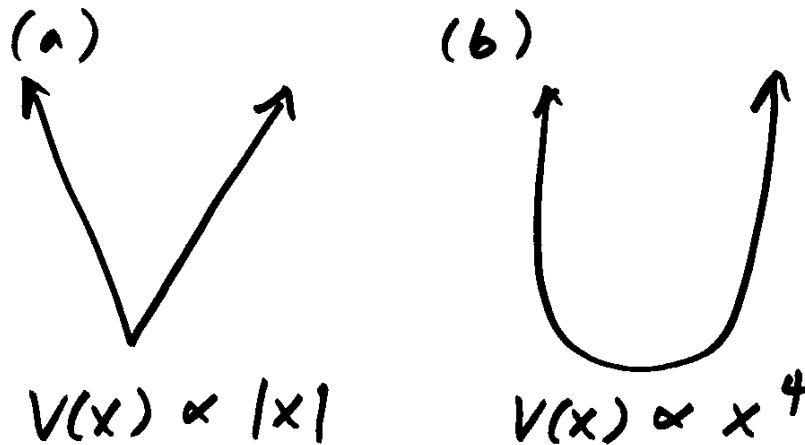


Figure 6.3: Various 1D potentials

Particle in a 1D finite box

This problem is a little more complicated. From quantum mechanics we know that the solution in the box region where the potential is zero will be wavelike. In the barrier region we also know that the solutions will be exponentially decaying.

The potential is

$$V(x) = \begin{cases} V & \text{if } x \leq 0 \\ 0 & \text{if } 0 < x < a \\ V & \text{if } x \geq a \end{cases}$$

The solutions are

$$\begin{aligned} \Psi_1(x) &= Ae^{\beta x} + Be^{-\beta x} \\ \Psi_2(x) &= Ce^{ikx} + De^{-ikx} \\ \Psi_3(x) &= Fe^{\beta x} + Ge^{-\beta x} \end{aligned}$$

where $\beta = \sqrt{\frac{2m(V-\epsilon)}{\hbar^2}}$ and $k = \sqrt{\frac{2m\epsilon}{\hbar^2}}$. By finiteness of the wavefunction $B = 0$. In addition $F = 0$. The wavefunction must not blow up in the barrier region. Must be well behaved.

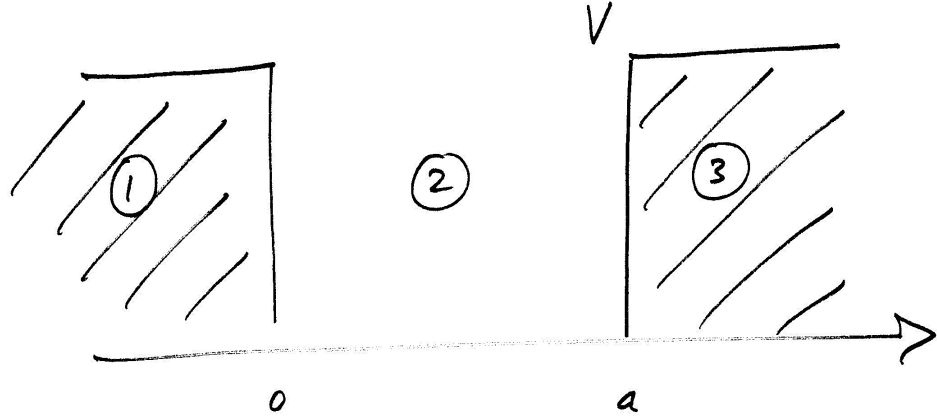


Figure 6.4: Cartoon of a one dimensional finite barrier potential

This leaves us with

$$\Psi_1(x) = Ae^{\beta x} \quad (6.6)$$

$$\Psi_2(x) = Ce^{ikx} + De^{-ikx} \quad (6.7)$$

$$\Psi_3(x) = Ge^{-\beta x} \quad (6.8)$$

Apply the boundary conditions and matching conditions as follows

$$\Psi_1(0) = \Psi_2(0) \rightarrow A = C + D$$

$$\Psi_1'(0) = \Psi_2'(0) \rightarrow A\beta = ikC - ikD$$

$$\Psi_2(a) = \Psi_3(a) \rightarrow Ce^{ika} + De^{-ika} = Ge^{-\beta a}$$

$$\Psi_2'(a) = \Psi_3'(a) \rightarrow ikCe^{ika} - ikDe^{-ika} = -\beta Ge^{-\beta a}$$

This leads to a system of four equations and four unknowns (A,C,D,G). Arranged in matrix form it looks like

$$\begin{pmatrix} 1 & -1 & -1 & 0 \\ \beta & -ik & ik & 0 \\ 0 & e^{ika} & e^{-ika} & -e^{-\beta a} \\ 0 & ike^{ika} & -ike^{-ika} & \beta e^{-\beta a} \end{pmatrix} \begin{pmatrix} A \\ C \\ D \\ G \end{pmatrix} = 0$$

Here either we get the trivial solution where $A=C=D=G=0$ or that the

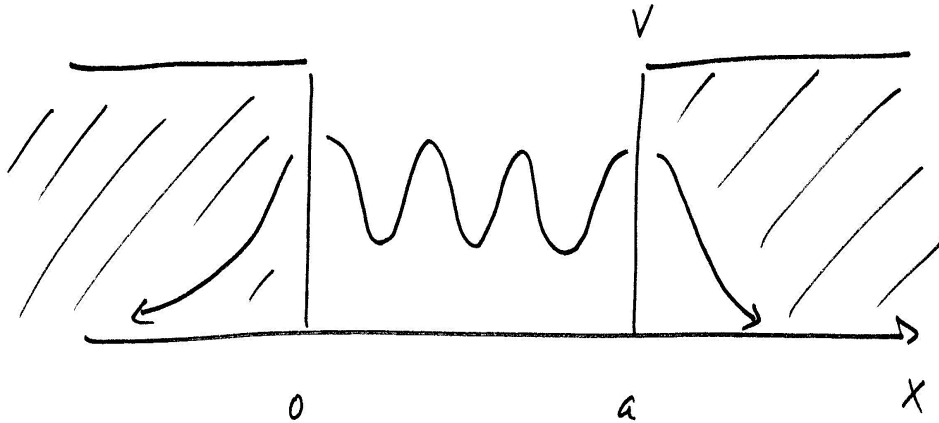


Figure 6.5: Sketch showing the behavior of the wavefunctions in the finite box model

determinant of the large matrix is zero.

$$\begin{vmatrix} 1 & -1 & -1 & 0 \\ \beta & -ik & ik & 0 \\ 0 & e^{ika} & e^{-ika} & -e^{-\beta a} \\ 0 & ike^{ika} & -ike^{-ika} & \beta e^{-\beta a} \end{vmatrix} = 0 \quad (6.9)$$

Simplify this determinant. This can be done a number of ways. One possible path is shown.

$$-ik(\text{row } 3) + (\text{row } 4) \rightarrow (\text{row } 4)$$

The determinant becomes

$$\begin{vmatrix} 1 & -1 & -1 & 0 \\ 0 & (\beta - ik) & (\beta + ik) & 0 \\ 0 & e^{ika} & e^{-ika} & -e^{-\beta a} \\ 0 & 0 & -2ike^{-ika} & (\beta + ik)e^{-\beta a} \end{vmatrix} = 0$$

Followed by

$$\begin{aligned} -\frac{1}{\beta - ik}(\text{row } 2) &\rightarrow (\text{row } 2) \\ e^{-ika}(\text{row } 3) &\rightarrow (\text{row } 3) \end{aligned}$$

yielding

$$\begin{vmatrix} 1 & -1 & -1 & 0 \\ 0 & -1 & -\frac{(\beta+ik)}{(\beta-ik)} & 0 \\ 0 & 1 & e^{-2ika} & -e^{-a(\beta+ik)} \\ 0 & 0 & -2ike^{-ika} & (\beta+ik)e^{-\beta a} \end{vmatrix} = 0$$

Additional steps

$$\begin{aligned} (\text{row } 2) + (\text{row } 3) &\rightarrow (\text{row } 3) \\ -(\text{row } 2) &\rightarrow (\text{row } 2) \end{aligned}$$

giving

$$\begin{vmatrix} 1 & -1 & -1 & 0 \\ 0 & 1 & \frac{(\beta+ik)}{(\beta-ik)} & 0 \\ 0 & 0 & e^{-2ika} - \frac{(\beta+ik)}{(\beta-ik)} & -e^{-a(\beta+ik)} \\ 0 & 0 & -2ike^{-ika} & (\beta+ik)e^{-\beta a} \end{vmatrix} = 0$$

Finding this determinant basically means finding the sub 2x2 determinant

$$\begin{vmatrix} e^{-2ika} - \frac{(\beta+ik)}{(\beta-ik)} & -e^{-a(\beta+ik)} \\ -2ike^{-ika} & (\beta+ik)e^{-\beta a} \end{vmatrix} = 0$$

This is the same as

$$\left(e^{-2ika} - \frac{(\beta+ik)}{(\beta-ik)} \right) (\beta+ik)e^{-\beta a} - 2ike^{-ika-a(\beta+ik)} = 0$$

$$(\beta+ik)e^{-2ika-\beta a} - \frac{(\beta+ik)^2}{(\beta-ik)}e^{-\beta a} - 2ike^{-2ika-\beta a} = 0$$

Drop the $e^{-\beta a}$ term to give

$$(\beta+ik)e^{-2ika} - \frac{(\beta+ik)^2}{(\beta-ik)} - 2ike^{-2ika} = 0$$

$$(\beta+ik - 2ik)e^{-2ika} = \frac{(\beta+ik)^2}{(\beta-ik)}$$

$$(\beta-ik)^2 e^{-2ika} = (\beta+ik)^2$$

Multiply both sides by e^{ika}

$$(\beta - ik)^2 e^{-ika} = (\beta + ik)^2 e^{ika}$$

$$(\beta + ik)^2 e^{ika} - (\beta - ik)^2 e^{-ika} = 0$$

Continue simplifying

$$(\beta^2 + 2i\beta k - k^2)e^{ika} - (\beta^2 - 2i\beta k - k^2)e^{-ika} = 0$$

$$(\beta^2 - k^2)(e^{ika} - e^{-ika}) + 2i\beta k(e^{ika} + e^{-ika}) = 0$$

$$2i(\beta^2 - k^2)\sin(ka) + 4i\beta k\cos(ka) = 0$$

$$(\beta^2 - k^2)\sin(ka) + 2\beta k\cos(ka) = 0$$

$$(\beta^2 - k^2)\tan(ka) + 2\beta k = 0$$

Giving our final expression

$$\boxed{\tan(ka) = \frac{2\beta k}{k^2 - \beta^2}} \quad (6.10)$$

where $\beta = \sqrt{\frac{2m(V-\varepsilon)}{\hbar^2}}$ and $k = \sqrt{\frac{2m\varepsilon}{\hbar^2}}$. If we replace this into the above equation we get the equation

$$\boxed{\tan\left(\sqrt{\frac{2m\varepsilon}{\hbar^2}}a\right) = \frac{2\sqrt{\varepsilon(V-\varepsilon)}}{2\varepsilon - V}} \quad (6.11)$$

Solve this numerically to get all allowed values of the energy (i.e. find the roots).

Exercises

1. Consider the same GaAs quantum wells in the previous exercise (well widths of 4 nm and 10 nm). Assume the effective mass of the electron is $0.067m_o$. Also assume, now rather than an infinite box, we have a finite box with a 0.3 eV barrier. Calculate the energy of electron states trapped in the box. Increase the barrier to 1 eV. Calculate the energy

of electron states trapped in the box (also known as bound states). This can be done using your favorite mathematical modeling software such as Mathcad, Matlab, Mathematica etc. . . .

Particle in a 1D triangular well

Here the potential is

$$V(z) = eFz$$

where e is the fundamental unit of charge and F is the magnitude of the electric field giving rise to such a potential. The left side of the potential is infinite, beginning at the origin.

The Schrodinger equation is

$$\left[-\frac{\hbar^2}{2m} \frac{d^2}{dz^2} + V(z) \right] \Psi(z) = \epsilon \Psi(z)$$

The relevant boundary condition for the wavefunction is that it must go to zero at the origin $\Psi(0) = 0$.

To simplify things and converge on a name equation that has already been solved we do the following. First multiply through by $\frac{\hbar^2}{2m}$

$$\begin{aligned} \frac{d^2\Psi}{dz^2} - \left(\frac{2meFz}{\hbar^2} \right) \Psi &= -\frac{2m\epsilon}{\hbar^2} \Psi \\ \frac{d^2\Psi}{dz^2} - \left(\frac{2meF}{\hbar^2} \right) z \Psi &= -\frac{2m\epsilon}{\hbar^2} \Psi \end{aligned}$$

Now let $\bar{z} = \frac{z}{z_o}$. We get

$$\begin{aligned} \frac{1}{z_o^2} \frac{d^2\Psi}{d\bar{z}^2} - \left(\frac{2meF}{\hbar^2} \right) z_o \bar{z} \Psi &= -\frac{2m\epsilon}{\hbar^2} \Psi \\ \frac{d^2\Psi}{d\bar{z}^2} - \left(\frac{2meF}{\hbar^2} \right) z_o^3 \bar{z} \Psi &= -\frac{2m\epsilon}{\hbar^2} z_o^2 \Psi \end{aligned}$$

Next, let $z_o = \frac{1}{\left(\frac{2meF}{\hbar^2} \right)^{\frac{1}{3}}}$. We then get

$$\frac{d^2\Psi}{\bar{z}^2} - \bar{z} \Psi = -\frac{2m\epsilon}{\hbar^2} \left(\frac{\hbar^2}{2meF} \right)^{\frac{2}{3}} \Psi$$

Particle in a Finite Potential Well

definitions we need

$\hbar := 1.054 \cdot 10^{-34}$ Self explanatory, units of joule seconds

$m_0 := 9.11 \cdot 10^{-31}$ Free electron mass, units of kg

$eV := 1.602 \cdot 10^{-19}$ Conversion factor from eV to joules, units of coulombs

User defined parameters

$\beta := 0.067$ Effective mass prefactor, unitless

$a := 1 \cdot 10^{-8}$ Width of the potential well, units of meters

$V := 0.3$ Barrier height in units of eV

Something to simplify the expression for the computer

$\varepsilon(\beta, a) := \sqrt{\frac{2 \cdot \beta \cdot m_0}{\hbar^2}} \cdot a \cdot \sqrt{eV}$ $\varepsilon(\beta, a) = 13.268$ Just a check to make sure not unreasonable number

Now this is the equation we derived and whose roots we want

$$F(E) := \tan(\varepsilon(\beta, a) \cdot \sqrt{E}) - \frac{2 \cdot \sqrt{E \cdot (V - E)}}{(2E - V)}$$

Need some guess to help the computer, the closer the guess the better

$g1 := 0.03$ Energy guesses, units of eV

$g2 := 0.16$

$E1 := \text{root}(F(g1), g1)$ The heart of this mathcad sheet, computer doing its job here

$E2 := \text{root}(F(g2), g2)$

$E1 = 0.034$ These are the energies of the bound levels

$E2 = 0.271$ Looks like there are only 2 bound states, since E2 close to 0.3 eV barrier height

Figure 6.6: Mathcad sheet showing numerical solutions to the particle in a finite box problem

Let $\bar{\epsilon} = \frac{\epsilon}{\epsilon_o}$ leading to

$$\begin{aligned} \frac{d^2\Psi}{\bar{z}^2} - \bar{z}\Psi &= -\frac{2m\bar{\epsilon}\epsilon_o}{\hbar^2} \left(\frac{\hbar^2}{2m\epsilon F} \right)^{\frac{2}{3}} \Psi \\ &= -\left(\frac{2m}{\hbar^2} \right)^{\frac{1}{3}} \left(\frac{1}{\epsilon F} \right)^{\frac{2}{3}} \epsilon_o \bar{\epsilon} \Psi \\ &= -\left(\frac{2m}{(\epsilon F \hbar)^2} \right)^{\frac{1}{3}} \epsilon_o \bar{\epsilon} \Psi \end{aligned}$$

Let $\epsilon_o = \frac{1}{\left(\frac{2m}{(\epsilon F \hbar)^2} \right)^{\frac{1}{3}}} = \left(\frac{(\epsilon F \hbar)^2}{2m} \right)^{\frac{1}{3}}$ giving

$$\begin{aligned} \frac{d^2\Psi}{\bar{z}^2} - \bar{z}\Psi &= -\bar{\epsilon}\Psi \\ \frac{d^2\Psi}{\bar{z}^2} &= (\bar{z} - \bar{\epsilon})\Psi \end{aligned}$$

Let $s = \bar{z} - \bar{\epsilon}$. Once this is done we are left with

$$\frac{d^2\Psi}{d\bar{z}^2} = s\Psi \quad (6.12)$$

This is Airy's equation and has general solution that are linear combinations of Airy functions (A_i and B_i). In general B_i will diverge as $z \rightarrow \infty$. Since we can't have the wavefunction diverge we drop the latter Airy function. We are then left with the generic solution for the wavefunctions

$$\Psi = N A_i(s) \quad (6.13)$$

where N is a normalization constant. To find the energies of the particle, we apply the one boundary condition that we have $\Psi(z = 0) = 0$. Since $s = \bar{z} - \bar{\epsilon}$ and $\bar{z} = 0$ if $z = 0$ we get $s = -\bar{\epsilon}$. The boundary condition is therefore $\Psi(-\bar{\epsilon}) = 0$. Let α be the (negative) zeros of the Airy function A_i . Therefore $\bar{\epsilon} = |\alpha|$ and recall $\bar{\epsilon} = \frac{\epsilon}{\epsilon_o}$. Solving for ϵ we get $\epsilon = \alpha \epsilon_o$ or

$$\epsilon = \alpha \left[\frac{(\epsilon F \hbar)^2}{2m} \right]^{\frac{1}{3}} \quad (6.14)$$

which are the desired energies of the particle in a 1D triangular potential.

2 Dimensions of confinement

Particle in an infinite circular box

The Schrodinger equation here is

$$\boxed{-\frac{\hbar^2}{2m}\nabla^2\Psi + V\Psi = \varepsilon\Psi}$$

The potential is

$$V(x) = \begin{cases} 0 & \text{if } r < a \\ \infty & \text{if } r \geq a \end{cases}$$

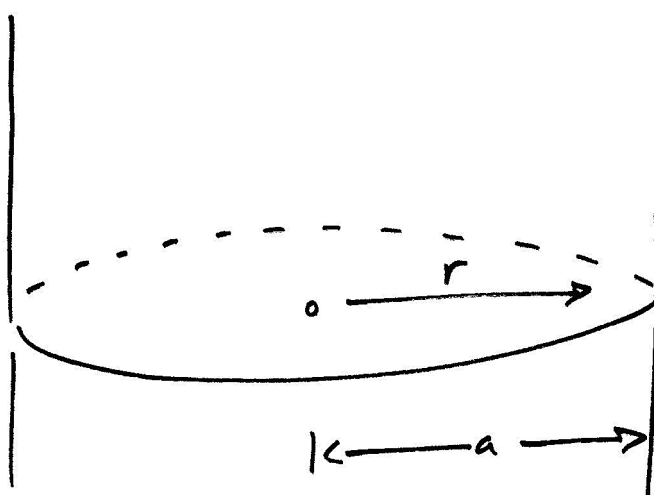


Figure 6.7: Cartoon of the infinite circular potential

In the region where the potential is zero

$$-\frac{\hbar^2}{2m}\nabla^2\Psi = \varepsilon\Psi$$

where ∇^2 , the Laplacian in r, θ coordinates, is

$$\nabla^2 = \frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2} \quad (6.15)$$

Note that the Laplacian in cartesian coordinates is $\nabla^2 = \frac{d^2}{dx^2} + \frac{d^2}{dy^2} + \frac{d^2}{dz^2}$. Insert this into the main equation to get

$$\begin{aligned} -\frac{\hbar^2}{2m} \left(\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial \Psi}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2 \Psi}{\partial \theta^2} \right) - \varepsilon \Psi &= 0 \\ \frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial \Psi}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2 \Psi}{\partial \theta^2} + \frac{2m\varepsilon}{\hbar^2} \Psi &= 0 \\ \frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial \Psi}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2 \Psi}{\partial \theta^2} + k^2 \Psi &= 0 \end{aligned}$$

Multiply through by r^2 to get

$$r \frac{\partial}{\partial r} \left(r \frac{\partial \Psi}{\partial r} \right) + (kr)^2 \Psi + \frac{\partial^2 \Psi}{\partial \theta^2} = 0$$

Note that you have part of this expression depends only on r and the other exclusively on θ . Assume a form of the wavefunction that is $\Psi = x(r)y(\theta)$. As shorthand just denote Ψ by xy . Let's evaluate the first two terms in the main equation above.

- $\frac{\partial \Psi}{\partial r} = yx'$
- $\left(r \frac{\partial \Psi}{\partial r} \right) = yrx'$
- $\frac{\partial}{\partial r} \left(r \frac{\partial \Psi}{\partial r} \right) = yrx'' + yx'$
- $r \frac{\partial}{\partial r} \left(r \frac{\partial \Psi}{\partial r} \right) = yr^2x'' + yrx'$

Replace the last item into the above main equation to get

$$yr^2x'' + yrx' + (kr)^2xy + xy'' = 0$$

Divide through by xy to get

$$\frac{r^2x''}{x} + \frac{rx'}{x} + (kr)^2 + \frac{y''}{y} = 0 \quad (6.16)$$

Solve for y first.

$$\frac{y''}{y} = - \left(\frac{r^2x''}{x} + \frac{rx'}{x} + (kr)^2 \right)$$

let

$$m^2 = \left(\frac{r^2 x''}{x} + \frac{rx'}{x} + (kr)^2 \right)$$

which you notice is independent of θ and so to y appears as a constant. Replacing this into the above expression gives

$$y'' = -m^2 y \quad (6.17)$$

or

$$y'' + m^2 y = 0$$

leading to a general solution of the form

$$y(\theta) = Ae^{im\theta} + Be^{-im\theta} \quad (6.18)$$

Now given an explicit form for $\frac{y''}{y}$ (equation 15) replace this back into equation 14. This is because ultimately what we want is $x(r)$. Equation 13 becomes

$$\begin{aligned} \frac{r^2 x''}{x} + \frac{rx'}{x} + (kr)^2 - m^2 &= 0 \\ \frac{x''}{x} + \frac{x'}{rx} + \left(k^2 - \frac{m^2}{r^2} \right) &= 0 \end{aligned}$$

leading to

$$\boxed{x'' + \frac{x'}{r} + x \left(k^2 - \frac{m^2}{r^2} \right) = 0} \quad (6.19)$$

which is the normal Bessel equation. Alternatively, to make this look like something you look up in a book, let $z = kr$. This results in Bessel's equation

$$\boxed{z^2 \frac{d^2 x}{dz^2} + z \frac{dx}{dz} + x(z^2 - m^2) = 0} \quad (6.20)$$

Solutions to this equation (for integer m) take the general form

$$\boxed{x(r) = AJ_m(z) + BY_m(z)} \quad (6.21)$$

which is a linear combination of $J_m(z)$ called Bessel functions of the first kind and $Y_m(z)$ called Bessel functions of the second kind. Note that Bessel functions of the first kind are well behaved at the origin but Bessel functions

of the second kind will diverge. Therefore, to obtain physically relevant solutions drop the Bessel functions of the second kind and only consider $J_m(z)$.

$$\boxed{x(r) = AJ_m(z)} \quad (6.22)$$

Finally, from our other boundary conditions, $J_m(ka) = 0$ Therefore $ka = \alpha$ which is the root of the Bessel function. Since $k = \sqrt{\frac{2m\varepsilon}{\hbar^2}}$ and $z = kr$, the eigenenergies that we desire are

$$\boxed{\varepsilon = \frac{\hbar^2 \alpha^2}{2ma^2}} \quad (6.23)$$

where a is the radius of the circle, m is the mass of the particle and α are the roots (first, second, third etc...) of the Bessel function. (Please note that the m in the denominator is the mass of the particle, not the order of the Bessel function. Sorry for the notation glitch)

Exercises

1. Consider an InP nanowire of diameter 10, 15 and 20 nm (see the Lieber paper below). Assume $m_e = 0.078m_o$ and $m_h = 0.4m_o$. Calculate the energy of the first 3 optical transitions by adding these particle in a circular box energies to the bulk room temperature band gap of InP. Ignore the length of the wire for simplicity. What colors do you expect these wires to emit, UV, visible, IR? How does it compare to the paper?
2. Consider a quantum corral as described below in the Eigler paper. Assume the diameter of the corral is 71.3\AA . Assume the effective mass of the electron is $m_e = 0.38m_o$. Calculate the first 3 electron energy levels of this corral. How does it compare to the paper?
3. For the same quantum corral above, draw the first three wavefunctions of the system. Use your favorite mathematical modeling program such as Mathcad, Matlab, Mathematica etc. . . .

Relevant reading

- “Size dependent photoluminescence from single indium phosphide nanowires” M. S. Gudiksen, J. Wang, C. M. Lieber J. Phys. Chem. B 106, 4036 (2002).

- “Confinement of electrons to quantum corrals on a metal surface” M. F. Crommie, C. P. Lutz, D. M. Eigler *Science*, 262, 218 (1993).
- “Quantum corrals” M. F. Crommie, C. P. Lutz, D. M. Eigler, E. J. Heller *Physica D*, 83, 98 (1995).
- “Energy levels in embedded semiconductor nanoparticles and nanowires” K. K. Nanda, F. E. Kruis, H. Fissan *Nano Lett.* 1, 605 (2001).

3 Dimensions of confinement

Particle in an infinite spherical box

This is a more complicated problem. Two approaches to a solution are illustrated with one leading to what are known as spherical Bessel functions and the other to a solution involving regular Bessel functions of half integer order. The Schrodinger equation is

$$\boxed{-\frac{\hbar^2}{2m}\nabla^2\Psi + V\Psi = \varepsilon\Psi}$$

The potential is

$$V(x) = \begin{cases} 0 & \text{if } r < a \\ \infty & \text{if } r \geq a \end{cases}$$

In the region inside the sphere where $V = 0$, this reduces to

$$-\frac{\hbar^2}{2m}\nabla^2\Psi = \varepsilon\Psi \quad (6.24)$$

where

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2\sin(\theta)}\frac{\partial}{\partial\theta}\left(\sin(\theta)\frac{\partial}{\partial\theta}\right) + \frac{1}{r^2\sin^2(\theta)}\frac{\partial^2}{\partial\phi^2} \quad (6.25)$$

If replaced into the above equation, multiply by $2mr^2$ on both sides to simplify giving

$$-\hbar^2r^2\nabla^2\Psi = 2mr^2\varepsilon\Psi$$

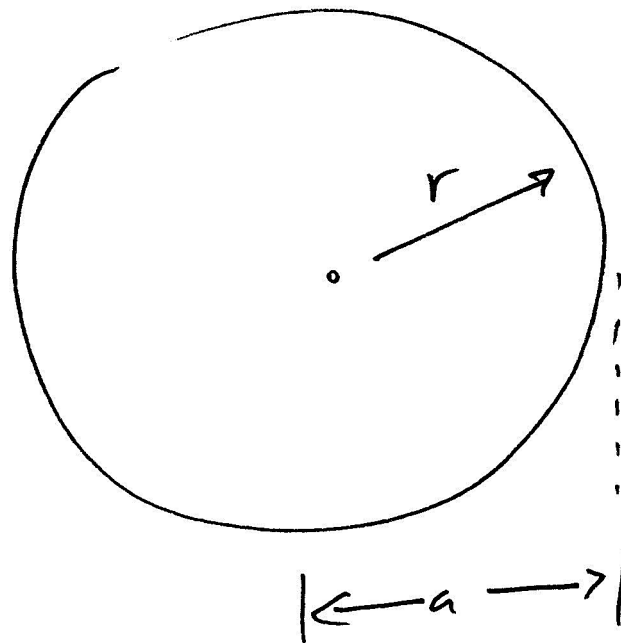


Figure 6.8: Cartoon of the infinite spherical potential

Expanded out this looks like

$$\begin{aligned}
& -\hbar^2 r^2 \left(\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \Psi}{\partial r} \right) + \frac{1}{r^2 \sin(\theta)} \frac{\partial}{\partial \theta} \left(\sin(\theta) \frac{\partial \Psi}{\partial \theta} \right) + \frac{1}{r^2 \sin^2(\theta)} \frac{\partial^2 \Psi}{\partial \phi^2} \right) \\
& \hspace{15em} = 2mr^2 \varepsilon \Psi \\
& -\hbar^2 \left(\frac{\partial}{\partial r} \left(r^2 \frac{\partial \Psi}{\partial r} \right) + \frac{1}{\sin(\theta)} \frac{\partial}{\partial \theta} \left(\sin(\theta) \frac{\partial \Psi}{\partial \theta} \right) + \frac{1}{\sin^2(\theta)} \frac{\partial^2 \Psi}{\partial \phi^2} \right) \\
& \hspace{15em} = 2mr^2 \varepsilon \Psi
\end{aligned}$$

Rearrange to give

$$\begin{aligned}
& - \hbar^2 \frac{\partial}{\partial r} \left(r^2 \frac{\partial \Psi}{\partial r} \right) - 2mr^2 \varepsilon \Psi \\
& - \hbar^2 \left(\frac{1}{\sin(\theta)} \frac{\partial}{\partial \theta} \left(\sin(\theta) \frac{\partial \Psi}{\partial \theta} \right) - \frac{1}{\sin^2(\theta)} \frac{\partial^2 \Psi}{\partial \phi^2} \right) \Psi = 0
\end{aligned}$$

where recall that the third term looks familiar. Basically

$$\hat{L}^2 = -\hbar^2 \left(\frac{1}{\sin(\theta)} \frac{\partial}{\partial \theta} \left(\sin(\theta) \frac{\partial}{\partial \theta} \right) - \frac{1}{\sin^2(\theta)} \frac{\partial^2}{\partial \phi^2} \right) \quad (6.26)$$

(Angular momentum operator). Our main equation then becomes

$$-\hbar^2 \frac{\partial}{\partial r} \left(r^2 \frac{\partial \Psi}{\partial r} \right) - 2mr^2 \varepsilon \Psi + \hat{L}^2 \Psi = 0$$

Furthermore, recall that $\hat{L}^2 \Psi = \hbar^2 l(l+1)$ (angular momentum eigenvalues) leading to

$$-\hbar^2 \frac{\partial}{\partial r} \left(r^2 \frac{\partial \Psi}{\partial r} \right) - 2mr^2 \varepsilon \Psi + \hbar^2 l(l+1) \Psi = 0$$

Simplify this

$$\begin{aligned}
& -\hbar^2 \frac{\partial}{\partial r} \left(r^2 \frac{\partial \Psi}{\partial r} \right) - \Psi (2mr^2 \varepsilon - \hbar^2 l(l+1)) = 0 \\
& \frac{\partial}{\partial r} \left(r^2 \frac{\partial \Psi}{\partial r} \right) + \Psi \left(\frac{2m\varepsilon}{\hbar^2} r^2 - l(l+1) \right) = 0
\end{aligned}$$

Let $k^2 = \frac{2m\varepsilon}{\hbar^2}$ giving

$$\frac{\partial}{\partial r} \left(r^2 \frac{\partial \Psi}{\partial r} \right) + \Psi (k^2 r^2 - l(l+1)) = 0 \quad (6.27)$$

At this point there are two ways to proceed from this equation. The first will lead to solutions involving so called “spherical” Bessel functions. The other will lead to solutions involving regular or normal Bessel functions of half integer order. Spherical and half integer order, normal, Bessel functions are proportional to each other so ultimately the solutions are the same.

Solution with spherical Bessel functions

Let $\Psi = x(r)y(\theta, \phi)$ The we can evaluate the following items using x and y as shorthand for $x(r)$ and $y(\theta, \phi)$

- $\frac{\partial \Psi}{\partial r} = yx'$
- $r^2 \frac{\partial \Psi}{\partial r} = yr^2x'$
- $\frac{\partial}{\partial r} (r^2 \frac{\partial \Psi}{\partial r}) = y(r^2x'' + 2rx')$

Replace the last item into equation 23 giving

$$y(r^2x'' + 2rx') + xy(k^2r^2 - l(l+1)) = 0$$

Divide out y and continue simplifying.

$$r^2x'' + 2rx' + x(k^2r^2 - l(l+1)) = 0$$

Let $z = kr$ (therefore $dr = \frac{dz}{k}$ and $dr^2 = \frac{dz^2}{k^2}$. these will be useful in a moment) and replace into the above expression

$$\begin{aligned} r^2x'' + 2rx' + x(z^2 - l(l+1)) &= 0 \\ r^2 \frac{d^2x}{dr^2} + 2r \frac{dx}{dr} + x(z^2 - l(l+1)) &= 0 \\ k^2r^2 \frac{d^2x}{dz^2} + 2rk \frac{dx}{dz} + x(z^2 - l(l+1)) &= 0 \end{aligned}$$

or

$$\boxed{z^2 \frac{d^2x}{dz^2} + 2z \frac{dx}{dz} + x(z^2 - l(l+1)) = 0} \quad (6.28)$$

This is the general spherical Bessel equation whose solutions take the form:

$$\boxed{x(r) = Aj_l(z) + By_l(z)} \quad (6.29)$$

where $j_l(z)$ are spherical Bessel functions of the first kind and $y_l(z)$ are spherical Bessel functions of the second kind (also known as spherical Neumann functions written as $n_l(z)$). Note that $j_l(z)$ is finite and well behaved at the origin while $y_l(z)$ diverges. So to get a physical solution, one must drop the spherical Bessel functions of the second kind leaving

$$\boxed{x(r) = A j_l(z)} \quad (6.30)$$

The equation and its solutions can be looked up in a text like “Handbook of Mathematical Functions” Abramowitz and Stegun, pg 437. Note that these spherical Bessel functions are denoted by little $j_n(z)$ as opposed to big $J_n(z)$ which are normal Bessel functions (same with spherical Bessel functions of the second kind). Furthermore, spherical bessel functions are related to half integer Bessel functions, a subclass of normal Bessel functions as shown at the end of this chapter. Examples of the spherical Bessel function solutions are

- $j_0(z) = \frac{\sin(z)}{z}$
- $j_1(z) = \frac{\sin(z)}{z^2} - \frac{\cos(z)}{z}$
- $j_2(z) = 3 \frac{\sin(z)}{z^3} - 3 \frac{\cos(z)}{z^2} - \frac{\sin(z)}{z}$

and so forth.

Solution with half integer (normal) Bessel functions

Alternatively assume that the wavefunction has the form $\psi = \frac{x(r)}{r} y(\theta, \phi) = R(r)y(\theta, \phi)$ Replace this in equation 24 and start simplifying.

$$\frac{\partial}{\partial r} \left(r^2 \frac{\partial \Psi}{\partial r} \right) + \Psi(k^2 r^2 - l(l+1)) = 0$$

We need the following bits of information

- $\frac{\partial \Psi}{\partial r} = yx \left(-\frac{1}{r^2} \right) + \frac{y}{r} x'$
- $r^2 \frac{\partial \Psi}{\partial r} = -xy + r y x'$
- $\frac{\partial}{\partial r} \left(r^2 \frac{\partial \Psi}{\partial r} \right) = -x' y + x'' r y + x' y = x'' r y$

Replace the last item into our main equation

$$x'' ry + \frac{xy}{r}(k^2 r^2 - l(l+1)) = 0$$

Get rid of y

$$rx'' + \frac{x}{r}(k^2 r^2 - l(l+1)) = 0$$

$$x'' + x \left(k^2 - \frac{l(l+1)}{r^2} \right) = 0$$

Now let $z = kr$ and $x = \sqrt{z}\phi(z)$ ($dr = \frac{dz}{k}$) giving

$$\frac{d^2 x}{dr^2} + x \left(k^2 - \frac{l(l+1)}{r^2} \right) = 0$$

$$k^2 \frac{d^2 x}{dz^2} + \sqrt{z}\phi \left(k^2 - \frac{k^2 l(l+1)}{z^2} \right) = 0$$

Use the following bits of information

- $\frac{dx}{dz} = \sqrt{z}\phi' + \phi \frac{1}{2}z^{-\frac{1}{2}}$
- $\frac{d^2 x}{dz^2} = \sqrt{z}\phi'' + \phi' \frac{1}{2}z^{-\frac{1}{2}} + \frac{1}{2} \left(\phi \left(-\frac{1}{2}\right) z^{-\frac{3}{2}} + z^{-\frac{1}{2}}\phi' \right)$
- $\frac{d^2 x}{dz^2} = \sqrt{z}\phi'' + \phi' z^{-\frac{1}{2}} - \frac{1}{4}\phi z^{-\frac{3}{2}}$

especially the last term to get

$$k^2 \left(\sqrt{z}\phi'' + \phi' z^{-\frac{1}{2}} - \frac{\phi}{4} z^{-\frac{3}{2}} \right) + \sqrt{z}\phi \left(k^2 - \frac{k^2 l(l+1)}{z^2} \right) = 0$$

Drop k^2 and continue simplifying

$$\sqrt{z}\phi'' + \phi' z^{-\frac{1}{2}} - \frac{\phi}{4} z^{-\frac{3}{2}} + \sqrt{z}\phi \left(1 - \frac{l(l+1)}{z^2} \right) = 0$$

$$z\phi'' + \phi' - \frac{\phi}{4z} + z\phi \left(1 - \frac{l(l+1)}{z^2} \right) = 0$$

$$z^2\phi'' + z\phi' - \frac{\phi}{4} + z^2\phi \left(1 - \frac{l(l+1)}{z^2} \right) = 0$$

$$z^2\phi'' + z\phi' + \phi \left(-\frac{1}{4} + z^2 - l(l+1) \right) = 0$$

$$z^2\phi'' + z\phi' + \phi \left(z^2 - \left(\frac{1}{4} + l(l+1) \right) \right) = 0$$

Note that the term in the inner parenthesis is equal to $(l + \frac{1}{2})^2$. This simplifies the full expression to

$$z^2\phi'' + z\phi' + \phi\left(z^2 - \left(l + \frac{1}{2}\right)^2\right) = 0$$

Let $n = (l + \frac{1}{2})$

$$\boxed{z^2\phi'' + z\phi' + \phi(z^2 - n^2) = 0} \quad (6.31)$$

This is the normal Bessel function equation. In this case where we have half-integer orders n then general solutions for ϕ are linear combinations of half integer Bessel functions. Furthermore, since $x(r) = \sqrt{z}\phi(z)$ and $R(r) = \frac{x(r)}{r}$ the general radial solution, $R(r)$, is

$$\begin{aligned} \phi(z) &= AJ_n(z) + BJ_n(z) \\ x(r) &= \sqrt{z}\left(A'J_n(z) + B'Y_n(z)\right) \end{aligned}$$

$$\boxed{R(r) = \frac{1}{\sqrt{z}}\left(A'J_n(z) + B'Y_n(z)\right)} \quad (6.32)$$

or more explicitly

$$\boxed{x(r) = \sqrt{z}\left(AJ_{(l+\frac{1}{2})}(z) + BY_{(l+\frac{1}{2})}(z)\right)} \quad (6.33)$$

We can look up the equation and its solution in a book like Abramowitz and Stegun but since Y_n diverges near the origin so it should be dropped immediately. This leaves

$$\boxed{R(r) = \frac{A'}{\sqrt{z}}J_{(l+\frac{1}{2})}(z)} \quad (6.34)$$

Summary

So in the end, regardless of how you came to the solution, to find the eigenvalues we need to find the roots, α , of either

$$\boxed{j_l(kr) = 0} \text{ spherical Bessel} \quad (6.35)$$

$$\boxed{J_{(l+\frac{1}{2})}(kr) = 0} \text{ half integer Bessel} \quad (6.36)$$

So on finding the root one also has $\alpha = kr$ and from this the desired value of k . Note that for each value of l there are an infinite number of roots. The eigenvalues are then

$$\begin{aligned} ka &= \alpha \\ \sqrt{\frac{2m\varepsilon}{\hbar^2}} a &= \alpha \\ \varepsilon &= \frac{\hbar^2 \alpha^2}{2ma^2} \end{aligned}$$

leading to

$$\begin{aligned} \varepsilon &= \frac{\hbar^2 k^2}{2m} \\ \boxed{\varepsilon = \frac{\hbar^2 \alpha^2}{2ma^2}} & \end{aligned} \quad (6.37)$$

Finally, the relation between spherical Bessel functions and half integer Bessel functions are:

$$j_l(z) = \sqrt{\frac{\pi}{2z}} J_{(l+\frac{1}{2})}(z) \quad (6.38)$$

$$y_l = n_l(z) = \sqrt{\frac{\pi}{2z}} Y_{(l+\frac{1}{2})}(z) \quad (6.39)$$

So either approach to solving the problem is valid. Its your choice on what's most convenient.

Exercises

1. Calculate the eigenenergies of a free electron (mass m_o) in a 5 nm diameter sphere for $l = 0, 1, 2, 3$ using the lowest root of the Bessel function.
2. Calculate the eigenenergies of a free electron (mass m_o) in a 5 nm diameter sphere for $l = 0, 1, 2, 3$ using the second lowest root of the Bessel function.
3. Calculate the eigenenergies of a free electron (mass m_o) in a 5 nm diameter sphere for $l = 0, 1, 2, 3$ using the third lowest root of the Bessel function.
4. Summarize all these energies in ascending order of energy in one table.

5. Use Mathcad, Matlab, Mathematica or your favorite mathematical modeling program to draw the radial wavefunctions corresponding to the lowest three energies in your table. Don't forget to normalize the wavefunction.
 6. Calculate the first 7 energy levels of a free electron in a 5 nm length three dimensional box. (All sides of the cube have length of 5 nm). Compare these results to the particle in a sphere case.
 7. Qualitatively compare the size dependence of the energies for a particle in a 1D box (well), an infinite circular potential (wire) and a particle in an infinite spherical box (dot). Basically ask yourself how all these energies scale with r or a . Any similarities? Now take a look at Yu et al., Nature Materials 2, 517 - 520 (01 Aug 2003) Letters and comment on it.
 8. Often enough there is a Coulomb attraction between the electron and hole in a confined system. As a consequence the energies one calculates via particle in a box type expression are often corrected for this Coulomb attraction. Think back to the basic expression for the Coulomb energy and qualitatively explain why this term is important or not important. Hint: think of how these energies all scale with r or a .
 9. Core shell semiconductor particles are pretty popular these days since they were found to improve the quantum yield of many semiconductor systems. Read the following papers with particular emphasis on the first one. **(1)** (CdSe)ZnS core-shell quantum dots: Synthesis and characterization of a size series of highly luminescent nanocrystallites" B. O. Dabbousi, J. Rodriguez-Viejo, F. V. Mikulec, J. R. Heine, H. Mattoussi, R. Ober, K. F. Jensen, M. G. Bawendi J. Phys. Chem. B 1997, 101, 9463. **(2)** "Quantum-dot quantum well CdS/HgS/CdS: Theory and experiment" D. Schooss, A. Mews, A. Eychmuller, H. Weller Phys. Rev. B 1994, 49, 17072. **(3)** "Type-II quantum dots: CdTe/CdSe (core/shell) and CdSe/ZnTe (core/shell) heterostructures" S. Kim, B. Fisher, H. J. Eisler, M. Bawendi J. Am. Chem. Soc. 2003, 125, 11466.
- Now Bashir (Dabbousi) has been good enough to plot the radial wavefunction of both the electron and hole in this system. See figure 14. You too can do this. Reproduce the first (CdSe/matrix) and second (CdSe/ZnS) graphs.

Particle in a cylinder

Here in cylindrical coordinates

$$\nabla^2 = \frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2} + \frac{\partial^2}{\partial z^2}$$

Solve the Schrodinger equation

$$H\Psi = \epsilon\Psi$$

where

$$H = -\frac{\hbar^2}{2m} \nabla^2 + V$$

where $V = 0$ inside the cylinder. Elsewhere the potential is infinite. The effective mass of the electron and the hole are used. Replacing this into the Schrodinger equation we get

$$-\frac{\hbar^2}{2m} \left[\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2} + \frac{\partial^2}{\partial z^2} \right] \Psi = \epsilon\Psi$$

Multiply both sides by $-\frac{2m}{\hbar^2}$ giving

$$\left[\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2} + \frac{\partial^2}{\partial z^2} \right] \Psi = -\frac{2m\epsilon}{\hbar^2} \Psi$$

Let $k^2 = \frac{2m\epsilon}{\hbar^2}$ giving

$$\left[\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2} + \frac{\partial^2}{\partial z^2} \right] \Psi = -\kappa^2 \Psi$$

Multiply by r^2 to get rid of the middle $\frac{1}{r^2}$ term. We get

$$\begin{aligned} \left[r \frac{\partial}{\partial r} \left(r \frac{\partial}{\partial r} \right) + \frac{\partial^2}{\partial \theta^2} + r^2 \frac{\partial^2}{\partial z^2} \right] \Psi &= -\kappa^2 r^2 \Psi \\ r \frac{\partial}{\partial r} \left(r \frac{\partial \Psi}{\partial r} \right) + r^2 \frac{\partial^2 \Psi}{\partial z^2} + \kappa^2 r^2 \Psi + \frac{\partial^2 \Psi}{\partial \theta^2} & \end{aligned}$$

Now assume a solution of the form

$$\Psi(r, \theta, z) = A(r, \theta)B(z) \equiv AB \tag{6.40}$$

through the method of separation of variables. We then have

$$\begin{aligned} r \frac{\partial}{\partial r} \left(r \frac{\partial AB}{\partial r} \right) + r^2 \frac{\partial^2 AB}{\partial z^2} + (\kappa r)^2 AB + \frac{\partial^2 AB}{\partial \theta^2} &= 0 \\ Br \frac{\partial}{\partial r} \left(r \frac{\partial A}{\partial r} \right) + Ar^2 \frac{\partial^2 B}{\partial z^2} + (kr)^2 AB + B \frac{\partial^2 A}{\partial \theta^2} &= 0 \end{aligned}$$

Divide through by AB to get

$$\frac{r}{A} \frac{\partial}{\partial r} \left(r \frac{\partial A}{\partial r} \right) + \frac{r^2}{B} \frac{\partial^2 B}{\partial z^2} + (kr)^2 + \frac{1}{A} \frac{\partial^2 A}{\partial \theta^2} = 0$$

Divide through by r^2 to get

$$\frac{1}{rA} \frac{\partial}{\partial r} \left(r \frac{\partial A}{\partial r} \right) + \frac{1}{B} \frac{\partial^2 B}{\partial z^2} + \kappa^2 + \frac{1}{r^2 A} \frac{\partial^2 A}{\partial \theta^2} = 0$$

Consolidate terms to get

$$\frac{1}{rA} \frac{\partial A}{\partial r} + \frac{1}{r^2 A} \frac{\partial^2 A}{\partial \theta^2} + \kappa^2 = -\frac{1}{B} \frac{\partial^2 B}{\partial z^2}$$

The LHS is a function of (r, θ) (although offset by κ^2). The RHS is a function of z . If these expressions are equivalent then they must both equal a constant with no dependence of r, θ or z . Let's call it m^2 . Therefore we get

$$\frac{1}{rA} \frac{\partial}{\partial r} \left(r \frac{\partial A}{\partial r} \right) + \frac{1}{r^2 A} \frac{\partial^2 A}{\partial \theta^2} + \kappa^2 = m^2 \quad (6.41)$$

$$-\frac{1}{B} \frac{\partial^2 B}{\partial z^2} = m^2 \quad (6.42)$$

Let's solve the easy equation first. Begin with

$$\begin{aligned} -\frac{1}{B} \frac{\partial^2 B}{\partial z^2} &= m^2 \\ \frac{d^2 B}{dz^2} &= -m^2 B \end{aligned}$$

leading to

$$\frac{d^2 B}{dz^2} + m^2 B = 0 \quad (6.43)$$

A general solution to this differential equation is

$$B = A_o e^{imz} + B_o e^{-imz} \quad (6.44)$$

where A_o and B_o are constants. The boundary conditions here require that

$$B(0) = 0 \quad (6.45)$$

$$B(L) = 0 \quad (6.46)$$

where L is the length of the rod or the wire.

Applying the first boundary condition we get

$$B(0) = A_o + B_o = 0$$

therefore

$$B_o = -A_o$$

Applying the second boundary condition we get

$$B(L) = A_o e^{imL} - A_o e^{-imL} = 0$$

or

$$\begin{aligned} 2iA_o \left(\frac{e^{imL} - e^{-imL}}{2i} \right) &= 0 \\ 2iA_o \sin mL &= 0 \end{aligned} \quad (6.47)$$

therefore we either get the trivial solution ($A_o = 0$) or we get $\sin mL = 0$. The latter solution requires that

$$\begin{aligned} mL &= n\pi \\ m &= \frac{n\pi}{L} \end{aligned}$$

where $n = 0, 1, 2, 3, \dots$

Putting everything together we get

$$B(z) = N_o \sin\left(\frac{n\pi}{L}z\right) \quad (6.48)$$

where N_o is a normalization constant. This is then the longitudinal part of the wavefunction.

Now we go and find the harder solutions. Begin with

$$\frac{1}{rA} \frac{d}{dr} \left(r \frac{dA}{dr} \right) + \frac{1}{r^2 A} \frac{d^2 A}{d\theta^2} + \kappa^2 = m^2$$

Multiply through by r^2 to get

$$\frac{r}{A} \frac{d}{dr} \left(r \frac{dA}{dr} \right) + \frac{1}{A} \frac{d^2 A}{d\theta^2} + (\kappa r)^2 = (mr)^2$$

Now multiply through by A

$$r \frac{d}{dr} \left(r \frac{dA}{dr} \right) + \frac{d^2 A}{d\theta^2} + ((\kappa r)^2 - (mr)^2) A = 0$$

Assume $A(r, \theta) \equiv C(r)D(\theta)$ again by the separation of variables approach.

$$A(r, \theta) = C(r)D(\theta) \equiv CD$$

giving

$$\begin{aligned} r \frac{\partial}{\partial r} \left(r \frac{\partial CD}{\partial r} \right) + \frac{\partial^2 CD}{\partial \theta^2} + [(\kappa r)^2 - (mr)^2] CD &= 0 \\ Dr \frac{\partial}{\partial r} \left(r \frac{\partial C}{\partial r} \right) + C \frac{\partial^2 D}{\partial \theta^2} + [(\kappa r)^2 - (mr)^2] CD &= 0 \end{aligned}$$

Divide through by CD

$$\frac{r}{C} \frac{\partial}{\partial r} \left(r \frac{\partial C}{\partial r} \right) + \frac{1}{D} \frac{\partial^2 D}{\partial \theta^2} + [(\kappa r)^2 - (mr)^2] = 0$$

Consolidate common terms

$$\frac{r}{C} \frac{\partial}{\partial r} \left(r \frac{\partial C}{\partial r} \right) + [(\kappa r)^2 - (mr)^2] = -\frac{1}{D} \frac{\partial^2 D}{\partial \theta^2} \quad (6.49)$$

Note that the left side depends only on r and the right side depends only on θ . Both sides therefore equal a constant. We will call it p^2 . This gives us the following two equations

$$\frac{r}{C} \frac{\partial}{\partial r} \left(r \frac{\partial C}{\partial r} \right) + [(\kappa r)^2 - (mr)^2] = p^2 \quad (6.50)$$

$$-\frac{1}{D} \frac{\partial^2 D}{\partial \theta^2} = p^2 \quad (6.51)$$

Let's start by solving the easy equation. Begin with

$$\begin{aligned} -\frac{1}{D} \frac{d^2 D}{d\theta^2} &= p^2 \\ \frac{d^2 D}{d\theta^2} &= -Dp^2 \end{aligned}$$

giving

$$\frac{d^2 D}{d\theta^2} + p^2 D = 0 \quad (6.52)$$

The general solutions to this differential equation are of the form

$$D = C_o e^{ip\theta} + D_o e^{-ip\theta} \quad (6.53)$$

where C_o and D_o are constants. By continuity we also know that

$$D(\theta) = D(\theta + 2\pi)$$

giving

$$\begin{aligned} D(0) &= C_o + D_o \\ D(2\pi) &= C_o e^{ip(2\pi)} + D_o e^{-ip(2\pi)} \end{aligned}$$

by inspection p is an integer and $p = 0, 1, 2, 3, \dots$

Now we go and solve the harder equation. Begin with

$$\frac{r}{C} \frac{d}{dr} \left(r \frac{dC}{dr} \right) + [(\kappa r)^2 - (mr)^2] = p^2$$

Multiply through by C

$$\begin{aligned} r \frac{d}{dr} \left(r \frac{dC}{dr} \right) + [(\kappa r)^2 - (mr)^2] C &= p^2 C \\ r \left[r \frac{d^2 C}{dr^2} + \frac{dC}{dr} \right] + [(\kappa r)^2 - (mr)^2] C &= p^2 C \\ r^2 \frac{d^2 C}{dr^2} + r \frac{dC}{dr} + [(\kappa r)^2 - (mr)^2] C &= p^2 C \\ r^2 \frac{d^2 C}{dr^2} + r \frac{dC}{dr} + [(\kappa r)^2 - (mr) - p^2] C &= 0 \end{aligned}$$

Divide through by r^2

$$\frac{d^2 C}{dr^2} + \frac{1}{r} \frac{dC}{dr} + \left[(\kappa^2 - m^2) - \frac{p^2}{r^2} \right] C = 0$$

Let $a^2 = (\kappa^2 - m^2)$

$$\frac{d^2 C}{dr^2} + \frac{1}{r} \frac{dC}{dr} + \left[a^2 - \frac{p^2}{r^2} \right] C = 0$$

Multiply through by r^2 again

$$r^2 \frac{dC}{dr} + r \frac{dC}{dr} + [(ar)^2 - p^2] C = 0$$

Let $\xi = ar$ or $d\xi = adr$ or $dr = \frac{d\xi}{a}$. This gives

$$a^2 r^2 \frac{dC}{d\xi} + ar \frac{dC}{d\xi} + [\xi^2 - p^2] C = 0$$

or

$$\xi^2 \frac{dC}{d\xi} + \xi \frac{dC}{d\xi} + [\xi^2 - p^2] C = 0 \quad (6.54)$$

This is Bessel's equation whose solutions are Bessel functions.

$$C = AJ_p(\xi) + BY_p(\xi)$$

where A, B are constants and $\xi = ar$ where $a = \sqrt{\kappa^2 - m^2}$ and $\kappa^2 = \frac{2m\epsilon}{\hbar^2}$. Because $Y_n(\xi) \rightarrow \infty$ as $r \rightarrow \infty$ we have $B = 0$. This leaves us with

$$C(\xi) = AJ_p(\xi) \quad (6.55)$$

This is the wavefunction of this part of the problem.

Now we summarize everything here

$$\begin{aligned} C(r) &= AJ_p(ar) \\ D(\theta) &= C_o e^{ip\theta} + D_o e^{-ip\theta} \\ B(z) &= N_o \sin\left(\frac{n\pi}{L} z\right) \end{aligned}$$

where $p = 0, 1, 2, 3, \dots$ and $n = 0, 1, 2, 3, \dots$. Putting it all together we have the total wavefunction as the product of the individual terms $\Psi = C(r)D(\theta)B(z)$ giving

$$\Psi(r, \theta, z) = n \left[J_p(ar) \left(C_o e^{ip\theta} + D_o e^{-ip\theta} \right) \right] \sin\left(\frac{n\pi}{L} z\right) \quad (6.56)$$

This is our desired wavefunction.

The desired energies come from the roots of the Bessel function

$$J_p(ar) = 0 \quad (6.57)$$

where $r = a_o$ is the radius of the cylinder. α will be the root of the Bessel function and $\alpha = aa_o$ where $a = \sqrt{\kappa^2 - m^2}$. Then we have

$$\begin{aligned}\alpha &= \sqrt{\kappa^2 - m^2}a_o \\ &= \sqrt{\frac{2m\epsilon}{\hbar^2} - m^2}a_o \\ \sqrt{\frac{2m\epsilon}{\hbar^2}} &= \frac{\alpha}{a_o} \\ \frac{2m\epsilon}{\hbar^2} - m^2 &= \frac{\alpha^2}{a_o^2} \\ \frac{2m\epsilon}{\hbar^2} &= \frac{\alpha^2}{a_o^2} + m^2\end{aligned}$$

giving

$$\epsilon = \frac{\hbar^2}{2m} \left(\frac{\alpha^2}{a_o^2} + m^2 \right)$$

where $m = \frac{n\pi}{L}$, a_o is the radius of the cylinder, and α are the roots of the Bessel function. If $a_o = R$ the the final desired energies of the particle in a cylinder are

$$\epsilon = \frac{\hbar^2}{2m} \left[\left(\frac{\alpha}{R} \right)^2 + \left(\frac{n\pi}{L} \right)^2 \right] \quad (6.58)$$

Chapter 7

Nondegenerate perturbation theory

The basic idea here is that the true Hamiltonian \hat{H} is close to an approximate Hamiltonian $\hat{H}^{(0)}$ whose eigenfunctions and eigenvalues are known. The difference $\hat{H}^{(1)}$ is called a perturbation. Higher order corrections such as $\hat{H}^{(2)}, \hat{H}^{(3)} \dots$ can be included if desired. Usually though you will stop at the first or second correction to either the energies or wavefunctions.

- We introduce a parameter λ to keep track of the order or alternatively, the order of refinement of the solutions.
- So 1st order corrections from $\hat{H}^{(1)}$ have a λ in front of it.
- 2nd order corrections from $\hat{H}^{(2)}$ will have λ^2 in front of it
- and so on in this fashion
- Also λ does not represent any real physical quantity. It is simply a mathematical technique for keeping track of the order of approximation. In the end we will usually set $\lambda = 1$ but the method is valid for any fraction of λ .

Our corrected Hamiltonian is therefore

$$\hat{H} = \hat{H}^{(0)} + \lambda\hat{H}^{(1)} + \lambda^2\hat{H}^{(2)} + \dots$$

where $\hat{H}^{(0)}$ is our unperturbed Hamiltonian whose eigenvalues and eigenvectors we know. Our wavefunctions and energies are

$$\begin{aligned}\Psi &= \Psi^{(0)} + \lambda\Psi^{(1)} + \lambda^2\Psi^{(2)} + \dots \\ \epsilon &= \epsilon^{(0)} + \lambda\epsilon^{(1)} + \lambda^2\epsilon^{(2)} + \dots\end{aligned}$$

The results of perturbation theory are derived by taking the exact Schrodinger equation $H\Psi = \epsilon\Psi$ and substituting in our above expressions for Ψ and ϵ . Next we group all terms with the same power of λ .

$$\begin{aligned} & [H^{(0)} + \lambda H^{(1)} + \lambda^2 H^{(2)}][\Psi^{(0)} + \lambda\Psi^{(1)} + \lambda^2\Psi^{(2)}] \\ = & [\epsilon^{(0)} + \lambda\epsilon^{(1)} + \lambda^2\epsilon^{(2)}][\Psi^{(0)} + \lambda\Psi^{(1)} + \lambda^2\Psi^{(2)}] \end{aligned}$$

Ok things get complicated here. We will separately evaluate the left and right hand side of these expressions. In addition, we will consider terms only up to second order (i.e. drop λ^3 and higher terms).

Left hand side

$$\begin{aligned} H^{(0)}\Psi^{(0)} & + \lambda H^{(0)}\Psi^{(1)} + \lambda^2 H^{(0)}\Psi^{(2)} + \lambda^3 H^{(1)}\Psi^{(2)} + \lambda^4 H^{(2)}\Psi^{(2)} \\ & + \lambda H^{(1)}\Psi^{(0)} + \lambda^2 H^{(1)}\Psi^{(1)} + \lambda^3 H^{(2)}\Psi^{(1)} \\ & + \lambda^2 H^{(2)}\Psi^{(0)} \end{aligned}$$

When we consider terms only up to second order we get

$$H^{(0)}\Psi^{(0)} + \lambda[H^{(0)}\Psi^{(1)} + H^{(1)}\Psi^{(0)}] + \lambda^2[H^{(0)}\Psi^{(2)} + H^{(1)}\Psi^{(1)} + H^{(2)}\Psi^{(0)}]$$

Right hand side

$$\begin{aligned} \epsilon^{(0)}\Psi^{(0)} & + \lambda\epsilon^{(0)}\Psi^{(1)} + \lambda^2\epsilon^{(0)}\Psi^{(2)} + \lambda^3\epsilon^{(1)}\Psi^{(2)} + \lambda^4\epsilon^{(2)}\Psi^{(2)} \\ & + \lambda\epsilon^{(1)}\Psi^{(0)} + \lambda^2\epsilon^{(1)}\Psi^{(1)} + \lambda^3\epsilon^{(2)}\Psi^{(1)} \\ & + \lambda^2\epsilon^{(2)}\Psi^{(0)} \end{aligned}$$

When we consider terms only up to second order we get

$$\epsilon^{(0)}\Psi^{(0)} + \lambda[\epsilon^{(0)}\Psi^{(1)} + \epsilon^{(1)}\Psi^{(0)}] + \lambda^2[\epsilon^{(0)}\Psi^{(2)} + \epsilon^{(1)}\Psi^{(1)} + \epsilon^{(2)}\Psi^{(0)}]$$

Now because the solution must be true for all values of λ one can equate the coefficients of power of λ giving

$$\boxed{H^{(0)}\Psi^{(0)} = \epsilon^{(0)}\Psi^{(0)}}$$

$$\boxed{H^{(0)}\Psi^{(1)} + H^{(1)}\Psi^{(0)} = \epsilon^{(0)}\Psi^{(1)} + \epsilon^{(1)}\Psi^{(0)}}$$

$$\boxed{H^{(0)}\Psi^{(2)} + H^{(1)}\Psi^{(1)} + H^{(2)}\Psi^{(0)} = \epsilon^{(0)}\Psi^{(2)} + \epsilon^{(1)}\Psi^{(1)} + \epsilon^{(2)}\Psi^{(0)}}$$

Let's rearrange these expressions to get

$$\boxed{(H^{(0)} - \epsilon^{(0)})\Psi^{(0)} = 0} \quad (7.1)$$

$$\boxed{(H^{(0)} - \epsilon^{(0)})\Psi^{(1)} + (H^{(1)} - \epsilon^{(1)})\Psi^{(0)} = 0} \quad (7.2)$$

$$\boxed{(H^{(0)} - \epsilon^{(0)})\Psi^{(2)} + (H^{(1)} - \epsilon^{(1)})\Psi^{(1)} + (H^{(2)} - \epsilon^{(2)})\Psi^{(0)} = 0} \quad (7.3)$$

The first of the three equations is our unperturbed case. It provides our zero-order solutions. Nothing special here.

Now multiply all expression by $\Psi^{*(0)}$ and integrate. Also let's switch to bra-ket notation so I don't have to write out the integrals.

$$\begin{aligned} \langle \Psi^{(0)} | H^{(0)} - \epsilon^{(0)} | \Psi^{(0)} \rangle &= 0 \\ \langle \Psi^{(0)} | H^{(0)} - \epsilon^{(0)} | \Psi^{(1)} \rangle + \langle \Psi^{(0)} | H^{(1)} - \epsilon^{(1)} | \Psi^{(0)} \rangle &= 0 \\ \langle \Psi^{(0)} | H^{(0)} - \epsilon^{(0)} | \Psi^{(2)} \rangle + \langle \Psi^{(0)} | H^{(1)} - \epsilon^{(1)} | \Psi^{(1)} \rangle + \langle \Psi^{(0)} | H^{(2)} - \epsilon^{(2)} | \Psi^{(0)} \rangle &= 0 \end{aligned}$$

Now in the case of the first order and second order expressions the first term in each case is zero.

$$\begin{aligned} \langle \Psi^{(0)} | H^{(0)} - \epsilon^{(0)} | \Psi^{(1)} \rangle &= 0 \\ \langle \Psi^{(0)} | H^{(0)} - \epsilon^{(0)} | \Psi^{(2)} \rangle &= 0 \end{aligned}$$

this is because $H^{(0)}$ is Hermitian $H^{(0)} = H^{\dagger(0)}$. So you can operate with $H^{(0)}$ on the bra and see that each expression will die. Alternatively, you could also invoke the orthogonality of $\langle \Psi^{(0)} |$ and $|\Psi^{(1)} \rangle$ after operating on the bra.

This leaves

$$\begin{aligned} \langle \Psi^{(0)} | H^{(1)} - \epsilon^{(1)} | \Psi^{(0)} \rangle &= 0 \text{ 1st order} \\ \langle \Psi^{(0)} | H^{(1)} - \epsilon^{(1)} | \Psi^{(1)} \rangle + \langle \Psi^{(0)} | H^{(2)} - \epsilon^{(2)} | \Psi^{(0)} \rangle &= 0 \text{ 2nd order} \end{aligned}$$

Next let's break each expression up and drop any zero terms. We get

$$\begin{aligned} \langle \Psi^{(0)} | H^{(1)} | \Psi^{(0)} \rangle - \epsilon^{(1)} \langle \Psi^{(0)} | \Psi^{(0)} \rangle &= 0 \\ \langle \Psi^{(0)} | H^{(1)} | \Psi^{(1)} \rangle - \epsilon^{(1)} \langle \Psi^{(0)} | \Psi^{(1)} \rangle + \langle \Psi^{(0)} | H^{(2)} | \Psi^{(0)} \rangle - \epsilon^{(2)} \langle \Psi^{(0)} | \Psi^{(0)} \rangle &= 0 \end{aligned}$$

reducing to

$$\begin{aligned} \langle \Psi^{(0)} | H^{(1)} | \Psi^{(0)} \rangle - \epsilon^{(1)} &= 0 \\ \langle \Psi^{(0)} | H^{(1)} | \Psi^{(1)} \rangle + \langle \Psi^{(0)} | H^{(2)} | \Psi^{(0)} \rangle - \epsilon^{(2)} &= 0 \end{aligned}$$

finally simplifying to

$$\begin{aligned} \langle \Psi^{(0)} | H^{(1)} | \Psi^{(0)} \rangle &= \epsilon^{(1)} \\ \langle \Psi^{(0)} | H^{(1)} | \Psi^{(1)} \rangle + \langle \Psi^{(0)} | H^{(2)} | \Psi^{(0)} \rangle &= \epsilon^{(2)} \end{aligned}$$

Therefore for the energies we have the following first and second order corrections to the zero order solutions.

$$\boxed{\epsilon^{(1)} = \langle \Psi^{(0)} | H^{(1)} | \Psi^{(0)} \rangle} \quad (7.4)$$

$$\boxed{\epsilon^{(2)} = \langle \Psi^{(0)} | H^{(1)} | \Psi^{(1)} \rangle + \langle \Psi^{(0)} | H^{(2)} | \Psi^{(2)} \rangle} \quad (7.5)$$

Note that to get the second order energy correction you already need to know what $|\Psi^{(1)}\rangle$ is. The same applies to higher order corrections and means that this can get pretty tedious. Usually you stop at first order corrections. So right now our second order expression is not necessarily in a useful form. We will revisit this later after finding out what $|\Psi^{(1)}\rangle$ looks like.

Example 1, expression for ϵ^3

Please calculate the expression for the third order correction ϵ^3

$$\begin{aligned} H &= H^{(0)} + \lambda H^{(1)} + \lambda^2 H^{(2)} + \lambda^3 H^{(3)} + \dots \\ \Psi &= \Psi^{(0)} + \lambda \Psi^{(1)} + \lambda^2 \Psi^{(2)} + \lambda^3 \Psi^{(3)} + \dots \\ \epsilon &= \epsilon^{(0)} + \lambda \epsilon^{(1)} + \lambda^2 \epsilon^{(2)} + \lambda^3 \epsilon^{(3)} + \dots \end{aligned}$$

Plugging this into the Schrodinger equation we get

$$\begin{aligned} & [H^{(0)} + \lambda H^{(1)} + \lambda^2 H^{(2)} + \lambda^3 H^{(3)}] [\Psi^{(0)} + \lambda \Psi^{(1)} + \lambda^2 \Psi^{(2)} + \lambda^3 \Psi^{(3)}] \\ &= [\epsilon^{(0)} + \lambda \epsilon^{(1)} + \lambda^2 \epsilon^{(2)} + \lambda^3 \epsilon^{(3)}] [\Psi^{(0)} + \lambda \Psi^{(1)} + \lambda^2 \Psi^{(2)} + \lambda^3 \Psi^{(3)}] \end{aligned}$$

Again, since this gets a little involved we will consider the left and right hand side of this equation separately.

Left hand side

$$\begin{aligned} H^{(0)}\Psi^{(0)} &+ \lambda H^{(0)}\Psi^{(1)} + \lambda^2 H^{(0)}\Psi^{(2)} + \lambda^3 H^{(0)}\Psi^{(3)} + \lambda^4 H^{(1)}\Psi^{(3)} \\ &+ \lambda^5 H^{(2)}\Psi^{(3)} + \lambda^6 H^{(3)}\Psi^{(3)} \\ &+ \lambda H^{(1)}\Psi^{(0)} + \lambda^2 H^{(1)}\Psi^{(1)} + \lambda^3 H^{(1)}\Psi^{(2)} + \lambda^4 H^{(2)}\Psi^{(2)} \\ &+ \lambda^5 H^{(3)}\Psi^{(2)} \\ &+ \lambda^2 H^{(2)}\Psi^{(0)} + \lambda^3 H^{(2)}\Psi^{(1)} + \lambda^4 H^{(3)}\Psi^{(1)} \\ &+ \lambda^3 H^{(3)}\Psi^{(0)} \end{aligned}$$

We want the third order corrections here.

Right hand side

$$\begin{aligned}
\epsilon^{(0)}\Psi^{(0)} &+ \lambda\epsilon^{(0)}\Psi^{(1)} + \lambda^2\epsilon^{(0)}\Psi^{(2)} + \lambda^3\epsilon^{(0)}\Psi^{(3)} + \lambda^4\epsilon^{(1)}\Psi^{(3)} + \lambda^5\epsilon^{(2)}\Psi^{(3)} + \lambda^6\epsilon^{(3)}\Psi^{(3)} \\
&+ \lambda\epsilon^{(1)}\Psi^{(0)} + \lambda^2\epsilon^{(1)}\Psi^{(1)} + \lambda^3\epsilon^{(1)}\Psi^{(2)} + \lambda^4\epsilon^{(2)}\Psi^{(2)} + \lambda^5\epsilon^{(3)}\Psi^{(2)} \\
&+ \lambda^2\epsilon^{(2)}\Psi^{(0)} + \lambda^3\epsilon^{(2)}\Psi^{(1)} + \lambda^4\epsilon^{(3)}\Psi^{(1)} \\
&+ \lambda^3\epsilon^{(3)}\Psi^{(0)}
\end{aligned}$$

We only want the 3rd order corrections here.

This leaves

$$\boxed{H^{(0)}\Psi^{(3)} + H^{(1)}\Psi^{(2)} + H^{(2)}\Psi^{(1)} + H^{(3)}\Psi^{(0)} = \epsilon^{(0)}\Psi^{(3)} + \epsilon^{(1)}\Psi^{(2)} + \epsilon^{(2)}\Psi^{(1)} + \epsilon^{(3)}\Psi^{(0)}}$$

Multiply both sides by $\Psi^{*(0)}$ and integrate. Switching to the Dirac bra-ket notation we get

$$(H^{(0)} - \epsilon^{(0)})|\Psi^{(3)}\rangle + (H^{(1)} - \epsilon^{(1)})|\Psi^{(2)}\rangle + (H^{(2)} - \epsilon^{(2)})|\Psi^{(1)}\rangle + (H^{(3)} - \epsilon^{(3)})|\Psi^{(0)}\rangle = 0$$

$$\begin{aligned}
&\langle \Psi^{(0)} | H^{(0)} - \epsilon^{(0)} | \Psi^{(3)} \rangle + \langle \Psi^{(0)} | H^{(1)} - \epsilon^{(1)} | \Psi^{(2)} \rangle \\
+ &\langle \Psi^{(0)} | H^{(2)} - \epsilon^{(2)} | \Psi^{(1)} \rangle + \langle \Psi^{(0)} | H^{(3)} - \epsilon^{(3)} | \Psi^{(0)} \rangle = 0
\end{aligned}$$

the first term drops out leaving

$$\begin{aligned}
&\langle \Psi^{(0)} | H^{(1)} - \epsilon^{(1)} | \Psi^{(2)} \rangle + \langle \Psi^{(0)} | H^{(2)} - \epsilon^{(2)} | \Psi^{(1)} \rangle \\
+ &\langle \Psi^{(0)} | H^{(3)} - \epsilon^{(3)} | \Psi^{(0)} \rangle = 0
\end{aligned}$$

becoming

$$\begin{aligned}
&\langle \Psi^{(0)} | H^{(1)} | \Psi^{(2)} \rangle - \epsilon^{(1)} \langle \Psi^{(0)} | \Psi^{(2)} \rangle + \langle \Psi^{(0)} | H^{(2)} | \Psi^{(1)} \rangle - \epsilon^{(2)} \langle \Psi^{(0)} | \Psi^{(1)} \rangle \\
+ &\langle \Psi^{(0)} | H^{(3)} | \Psi^{(0)} \rangle - \epsilon^{(3)} \langle \Psi^{(0)} | \Psi^{(0)} \rangle = 0
\end{aligned}$$

by orthonormality a bunch of terms drop out giving

$$\begin{aligned}
&\langle \Psi^{(0)} | H^{(1)} | \Psi^{(2)} \rangle + \langle \Psi^{(0)} | H^{(2)} | \Psi^{(1)} \rangle \\
+ &\langle \Psi^{(0)} | H^{(3)} | \Psi^{(0)} \rangle = \epsilon^{(3)}
\end{aligned}$$

Therefore our desired 3rd order correction is

$$\boxed{\epsilon^{(3)} = \langle \Psi^{(0)} | H^{(1)} | \Psi^{(2)} \rangle + \langle \Psi^{(0)} | H^{(2)} | \Psi^{(1)} \rangle + \langle \Psi^{(0)} | H^{(3)} | \Psi^{(0)} \rangle} \tag{7.6}$$

Example 2, Particle in a slanted box

We have a particle in a 1D box with a slanted bottom. The perturbation term is $H^{(1)} = \frac{V}{a}x$ where V is some constant. The zeroth order wavefunction is $\Psi^{(0)} = \sqrt{\frac{2}{a}}\sin\frac{n\pi x}{a}$ and the zeroth order energies are $\epsilon^{(0)} = \frac{n^2h^2}{8ma^2}$.

What is the 1st order correction to $\epsilon^{(0)}$?

From nondegenerate perturbation theory the first order correction is

$$\begin{aligned}\epsilon^{(1)} &= \langle \psi^{(0)} | H^{(1)} | \Psi^{(0)} \rangle \\ &= \int_0^a \left(\sqrt{\frac{2}{a}} \sin \frac{n\pi x}{a} \right) \left(\frac{V}{a} x \right) \left(\sqrt{\frac{2}{a}} \sin \frac{n\pi x}{a} \right) dx \\ &= \frac{2}{a} \left(\frac{V}{a} \right) \int_0^a x \sin^2 \frac{n\pi x}{a} dx \\ &\quad \text{let } b = \frac{n\pi}{a} \\ &= \frac{2V}{a^2} \int_0^a x \sin^2 bx dx \\ &\quad \text{where } \sin^2 bx = \frac{1}{2}(1 - \cos 2bx) \\ &= \frac{2V}{a^2} \left[\frac{1}{2} \int_0^a x(1 - \cos 2bx) dx \right] \\ &= \frac{V}{a^2} \left[\int_0^a x dx - \int_0^a x \cos 2bx dx \right] \\ &= \frac{V}{a^2} \left[\frac{x^2}{2} \Big|_0^a - \int_0^a x \cos 2bx dx \right] \\ &= \frac{V}{a^2} \left[\frac{a^2}{2} - \int_0^a x \cos 2bx dx \right]\end{aligned}$$

Solve the integral by parts. Let $u = x, du = dx, dv = \cos 2bx dx, v = \frac{1}{2b} \sin 2bx$. We get

$$\begin{aligned}&= \frac{V}{2} - \frac{V}{a^2} \left[\frac{x}{2b} \sin 2bx \Big|_0^a - \int_0^a \frac{1}{2b} \sin 2bx dx \right] \\ &= \frac{V}{2} - \frac{V}{a^2} \left[\frac{x}{2b} \sin 2bx \Big|_0^a + \frac{1}{2b} \frac{1}{2b} \cos 2bx \Big|_0^a \right] \\ &\quad \text{the first term is zero} \\ &= \frac{V}{2} - \frac{V}{a^2} \left[\frac{1}{4b^2} \cos 2bx \Big|_0^a \right] \\ &\quad \text{the term in brackets is also zero}\end{aligned}$$

therefore

$$\boxed{\epsilon^{(1)} = \frac{V}{2}} \quad (7.7)$$

This is our desired 1st order correction.

The final corrected energies are then

$$\boxed{\epsilon = \frac{n^2 h^2}{8ma^2} + \frac{V}{2}} \quad (7.8)$$

where $n = 1, 2, 3, \dots$

Example 3, An anharmonic oscillator

Consider an anharmonic oscillator whose potential is $U(x) = \frac{1}{2}kx^2 + \frac{1}{6}\gamma x^3$. Note that this problem does not require a lot of math.

Calculate the 1st order correction to the ground state energy. We have

$$\begin{aligned} \epsilon &= \left(n + \frac{1}{2}\right)h\nu \\ \Psi &= N_n H_n(\alpha^{1/2}x) e^{-\frac{\alpha x^2}{2}} \end{aligned}$$

where $\alpha = \sqrt{\frac{k\mu}{\hbar^2}}$ and the normalization constant is $N_n = \frac{1}{\sqrt{2^n n!}} \left(\frac{\alpha}{\pi}\right)^{\frac{1}{4}}$

The first order energy correction is

$$\epsilon^{(1)} = \langle \Psi^{(0)} | H^{(1)} | \Psi^{(0)} \rangle$$

where $H^{(1)} = \frac{1}{6}\gamma x^3$. We have

$$\begin{aligned} &= \frac{\gamma}{6} \int_{-\infty}^{\infty} \left(N_n H_n(\alpha^{1/2}x) e^{-\frac{\alpha x^2}{2}} \right) (x^3) \left(N_n H_n(\alpha^{1/2}x) e^{-\frac{\alpha x^2}{2}} \right) dx \\ &= \frac{\gamma}{6} N_n^2 \int_{-\infty}^{\infty} [H_n(\alpha^{1/2}x)] [x^3] [H_n(\alpha^{1/2}x)] e^{-\alpha x^2} dx \end{aligned}$$

Now if $n = 0$ $H_0(\alpha^{1/2}x) = 1$ leaving

$$= \frac{\gamma}{6} N_n^2 \int_{-\infty}^{\infty} x^3 e^{-\alpha x^2} dx$$

The first term in the integral is odd. The second is even. Now by symmetry their product is overall odd and over the range of the integral it evaluates to zero. Therefore

$$\boxed{\epsilon^{(1)} = 0} \quad (7.9)$$

This is our desired first order correction. So to first order the overall energies are

$$\boxed{\epsilon = (n + \frac{1}{2})h\nu} \quad (7.10)$$

Example 4, Another anharmonic oscillator

Calculate the 1st order correction to the ground state energy of an anharmonic oscillator whose potential energy is

$$\begin{aligned} U(x) &= \frac{1}{2}kx^2 + \frac{1}{6}\gamma x^3 + \frac{1}{24}bx^4 \\ H^{(1)} &= \frac{1}{6}\gamma x^3 + \frac{1}{24}bx^4 \\ \epsilon &= (n + \frac{1}{2})h\nu \\ \Psi &= N_n H_n(\alpha^{\frac{1}{2}}x) e^{-\frac{\alpha x^2}{2}} \end{aligned}$$

where $\alpha = \sqrt{\frac{k\mu}{\hbar^2}}$ and the normalization constant is $N = \frac{1}{\sqrt{2^n n!}} \left(\frac{\alpha}{\pi}\right)^{\frac{1}{4}}$ and $N_0 = \left(\frac{\alpha}{\pi}\right)^{\frac{1}{4}}$ if $n = 0$. The first order energy correction is

$$\begin{aligned} \epsilon^{(1)} &= \langle \Psi^{(0)} | H^{(1)} | \Psi^{(0)} \rangle \\ &= \int_{-\infty}^{\infty} \left(N_0 H_0(\alpha^{\frac{1}{2}}x) e^{-\frac{\alpha x^2}{2}} \right) \left(\frac{1}{6}\gamma x^3 + \frac{1}{24}bx^4 \right) \left(N_0 H_0(\alpha^{\frac{1}{2}}x) e^{-\frac{\alpha x^2}{2}} \right) dx \end{aligned}$$

We have already done the 1st term. By symmetry it drops out leaving.

$$= \frac{b}{24} N_0^2 \int_{-\infty}^{\infty} (H_0(\alpha^{\frac{1}{2}}x))(x^4)(H_0(\alpha^{\frac{1}{2}}x))e^{-\alpha x^2} dx$$

where $H_0(\alpha^{\frac{1}{2}}x) = 1$

$$= \frac{b}{24} N_0^2 \int_{-\infty}^{\infty} x^4 e^{-\alpha x^2} dx$$

this integral does not disappear by symmetry

and must be evaluated

$$= \frac{bN_0^2}{12} \int_{-\infty}^{\infty} x^4 e^{-\alpha x^2} dx$$

looking this up in a table of integrals

where

$$\boxed{\int_0^{\infty} x^{2n} e^{-\alpha x^2} dx = \frac{1 \cdot 3 \cdot 5 \cdot (2n-1)}{2^{n+1} \alpha^n} \sqrt{\frac{\pi}{\alpha}}}$$

Our integral evaluates to $\frac{3}{8\alpha^2}\sqrt{\frac{\pi}{\alpha}}$. The expression becomes

$$\begin{aligned}
 &= \frac{bN_0^2}{12} \left(\frac{3}{8\alpha^2} \right) \sqrt{\frac{\pi}{\alpha}} \\
 &= \frac{bN_0^2}{32\alpha^2} \sqrt{\frac{\pi}{\alpha}} \\
 &\quad \text{where } N_0^2 = \sqrt{\frac{\alpha}{\pi}} \text{ giving} \\
 &= \frac{b}{32\alpha^2} \\
 &\quad \text{where } \alpha^2 = \frac{k\mu}{\hbar^2} \text{ giving} \\
 &= \frac{b\hbar^2}{32k\mu}
 \end{aligned}$$

The overall energy with the first order correction is therefore

$$\begin{aligned}
 \epsilon &= \left(n + \frac{1}{2}\right)h\nu + \frac{b\hbar^2}{32k\mu} \\
 &\boxed{\epsilon = \frac{h\nu}{2} + \frac{b\hbar^2}{32k\mu}} \qquad (7.11)
 \end{aligned}$$

Wavefunctions

Now that we have the energy corrections, what about the wavefunctions? Note that you will usually go to second order in the energies but only 1st order in the wavefunctions. Note that this is also called the Raleigh Ritz method.

We have $\Psi_i^{(0)}$ which is our unperturbed wavefunction. We know that we have corrections like

$$\Psi_{tot} = \Psi_i^{(0)} + \lambda\Psi_i^{(1)} + \lambda^2\Psi_i^{(2)} + \dots$$

A convenient way to express $\Psi^{(1)}$ is as a linear expansion in unperturbed wavefunctions which we will denote as $\Psi_j^{(0)}$ or $|j\rangle$. We then have

$$\Psi_i^{(1)} = \sum_j c_{ij}|j\rangle$$

and where c_{ij} are weights representing how much of each unperturbed wavefunction is present.

If we go back to our expression where we derived the 1st order corrections in energy we have

$$\begin{aligned}(H^{(0)} - \epsilon_i^{(0)})\Psi_i^{(1)} + (H^{(1)} - \epsilon_i^{(1)})\Psi_i^{(0)} &= 0 \\ H^{(0)}\Psi_i^{(1)} + H^{(1)}\Psi_i^{(0)} &= \epsilon_i^{(0)}\Psi_i^{(1)} + \epsilon_i^{(1)}\Psi_i^{(0)}\end{aligned}$$

Now recall that $\Psi_i^{(1)} = \sum_j c_{ij}|j\rangle$ and replace into the above expression

$$\begin{aligned}H^{(0)} \sum_j c_{ij}|j\rangle + H^{(1)}|i\rangle &= \epsilon_i^{(0)} \sum_j c_{ij}|j\rangle + \epsilon_i^{(1)}|i\rangle \\ \sum_j c_{ij}H^{(0)}|j\rangle + H^{(1)}|i\rangle &= \epsilon_i^{(0)} \sum_j c_{ij}|j\rangle + \epsilon_i^{(1)}|i\rangle\end{aligned}$$

Recall that $|j\rangle$ are eigenfunctions of $H^{(0)}$ since we expanded in the unperturbed basis.

$$\begin{aligned}\sum_j c_{ij}\epsilon_j^{(0)}|j\rangle + H^{(1)}|i\rangle &= \epsilon_i^{(0)} \sum_j c_{ij}|j\rangle + \epsilon_i^{(1)}|i\rangle \\ \epsilon_j^{(0)} \sum_j c_{ij}|j\rangle + H^{(1)}|i\rangle &= \epsilon_i^{(0)} \sum_j c_{ij}|j\rangle + \epsilon_i^{(1)}|i\rangle\end{aligned}$$

Next to start finding our desired coefficients c_{ij} multiply both sides by $\langle k|$ which is a member of the unperturbed basis. Note that $k \neq i$.

$$\begin{aligned}\epsilon_j^{(0)} \sum_j c_{ij} \langle k|j\rangle + \langle k|H^{(1)}|i\rangle &= \epsilon_i^{(0)} \sum_j c_{ij} \langle k|j\rangle + \epsilon_i^{(1)} \langle k|i\rangle \\ \epsilon_j^{(0)} \sum_j c_{ij} \langle k|j\rangle + \langle k|H^{(1)}|i\rangle &= \epsilon_i^{(0)} \sum_j c_{ij} \langle k|j\rangle\end{aligned}$$

Now due to the orthogonality between $|k\rangle$ and $|j\rangle$ the only terms in the above expression that survive are those where $k = j$.

$$\epsilon_k^{(0)} \sum_k c_{ik} + \langle k|H^{(1)}|i\rangle = \epsilon_i^{(0)} \sum_k c_{ik}$$

Consolidate our desired c_{ik} terms to get

$$\sum_k c_{ik}(\epsilon_k^{(0)} - \epsilon_i^{(0)}) = - \langle k|H^{(1)}|i\rangle$$

Therefore

$$c_{ik} = -\frac{\langle k|H^{(1)}|i\rangle}{(\epsilon_k^{(0)} - \epsilon_i^{(0)})}$$

$$\boxed{c_{ik} = \frac{\langle k|H^{(1)}|i\rangle}{(\epsilon_i^{(0)} - \epsilon_k^{(0)})}} \quad (7.12)$$

These are our desired correction coefficients. Also recall that $\Psi^{(1)} = \sum_j c_{ij}|j\rangle$ so you can see that the corrected wavefunction to first order is

$$\boxed{|\Psi\rangle = |i\rangle + \sum_k \frac{\langle k|H^{(1)}|i\rangle}{(\epsilon_i^{(0)} - \epsilon_k^{(0)})} |k\rangle} \quad (7.13)$$

The correction is

$$\boxed{\Psi^{(1)} = \sum_k \frac{\langle k|H^{(1)}|i\rangle}{(\epsilon_i^{(0)} - \epsilon_k^{(0)})} |k\rangle} \quad (7.14)$$

Now that we have the 1st order correction to the wavefunction, we can go back and express the 2nd order correction in energy in a more useful form.

Recall that

$$\epsilon^{(2)} = \langle \Psi_i^{(0)} | H^{(1)} | \Psi_i^{(1)} \rangle + \langle \Psi_i^{(0)} | H^{(2)} | \Psi_i^{(0)} \rangle$$

where we didn't know $|\Psi_i^{(1)}\rangle$ before. Now that we have an explicit form we can substitute it in. To make life even easier assume that the total perturbation goes only to first order, thus $H^{(2)} = 0$. We get

$$\begin{aligned} \epsilon^{(2)} &= \langle \Psi_i^{(0)} | H^{(1)} | \Psi_i^{(1)} \rangle \\ &= \sum_k \frac{\langle k|H^{(1)}|i\rangle \langle i|H^{(1)}|k\rangle}{(\epsilon_i^{(0)} - \epsilon_k^{(0)})} \end{aligned}$$

The final expression is

$$\boxed{\epsilon^{(2)} = \sum_k \frac{\langle k|H^{(1)}|i\rangle \langle i|H^{(1)}|k\rangle}{(\epsilon_i^{(0)} - \epsilon_k^{(0)})}} \quad (7.15)$$

This is a more user friendly 2nd order energy correction. Also note that $i \neq k$. This is called non-degenerate perturbation theory.

Summary

$$\boxed{\epsilon^{(1)} = \langle i | H^{(1)} | i \rangle} \quad (7.16)$$

$$\boxed{\epsilon^{(2)} = \sum_k \frac{\langle k | H^{(1)} | i \rangle \langle i | H^{(1)} | k \rangle}{(\epsilon_i^{(0)} - \epsilon_k^{(0)})} = \sum_k \frac{|\langle k | H^{(1)} | i \rangle|^2}{(\epsilon_i^{(0)} - \epsilon_k^{(0)})}} \quad (7.17)$$

$$\boxed{|\Psi \rangle = |i \rangle + \sum_k \frac{\langle k | H^{(1)} | i \rangle}{(\epsilon_i^{(0)} - \epsilon_k^{(0)})} |k \rangle} \quad (7.18)$$

So you see that we will typically go to second order corrections in the energies and only to 1st order in the wavefunctions.

What about 3rd order corrections?

Going back for the 3rd order non-degenerate energy correction $\epsilon^{(3)}$. Here is an alternative expression where we don't need to know $|\Psi_i^{(2)} \rangle$ only $|\Psi_i^{(1)} \rangle$.

Starting with

$$\epsilon_i^{(3)} = \langle \Psi_i^{(0)} | H^{(1)} | \Psi_i^{(2)} \rangle + \langle \Psi_i^{(0)} | H^{(2)} | \Psi_i^{(1)} \rangle + \langle \Psi_i^{(0)} | H^{(3)} | \Psi_i^{(0)} \rangle$$

and with

$$\begin{aligned} (H^{(0)} - \epsilon_i^{(0)}) \Psi_i^{(1)} + (H^{(1)} - \epsilon_i^{(1)}) \Psi_i^{(0)} &= 0 \\ (H^{(0)} - \epsilon_i^{(0)}) \Psi_i^{(2)} + (H^{(1)} - \epsilon_i^{(1)}) \Psi_i^{(1)} + (H^{(2)} - \epsilon_i^{(2)}) \Psi_i^{(0)} &= 0 \end{aligned}$$

Using the second and third equations we will try and get rid of the first term in the first equation. To do this, multiply the second expression by $\Psi_i^{*(2)}$ and integrate. Multiply the third expression by $\Psi_i^{*(1)}$ and integrate.

Second expression

From the second expression we get the following

$$\begin{aligned} \Psi_i^{*(2)} (H^{(0)} - \epsilon_i^{(0)}) \Psi_i^{(1)} + \Psi_i^{*(2)} (H^{(1)} - \epsilon_i^{(1)}) \Psi_i^{(0)} &= 0 \\ \langle \Psi_i^{(2)} | H^{(0)} - \epsilon_i^{(0)} | \Psi_i^{(1)} \rangle + \langle \Psi_i^{(2)} | H^{(1)} - \epsilon_i^{(1)} | \Psi_i^{(0)} \rangle &= 0 \end{aligned}$$

giving

$$\boxed{\langle \Psi_i^{(2)} | H^{(1)} - \epsilon_i^{(1)} | \Psi_i^{(0)} \rangle = - \langle \Psi_i^{(2)} | H^{(0)} - \epsilon_i^{(0)} | \Psi_i^{(1)} \rangle}$$

where since $H^{(1)}$ and $H^{(0)}$ and $\epsilon_i^{(1)}$ and $\epsilon_i^{(0)}$ are all Hermitian.

$$\begin{aligned} \langle \Psi_i^{(0)} | H^{(1)} - \epsilon_i^{(1)} | \Psi_i^{(2)} \rangle &= - \langle \Psi_i^{(1)} | H^{(0)} - \epsilon_i^{(0)} | \Psi_i^{(2)} \rangle \\ \langle \Psi_i^{(0)} | H^{(1)} | \Psi_i^{(2)} \rangle &= - \langle \Psi_i^{(1)} | H^{(0)} - \epsilon_i^{(0)} | \Psi_i^{(2)} \rangle \end{aligned}$$

therefore

$$\epsilon_i^{(3)} = - \langle \Psi_i^{(1)} | H^{(0)} - \epsilon_i^{(0)} | \Psi_i^{(2)} \rangle + \langle \Psi_i^{(0)} | H^{(2)} | \Psi_i^{(1)} \rangle + \langle \Psi_i^{(0)} | H^{(3)} | \Psi_i^{(0)} \rangle$$

Third expression

Now from the third expression we get

$$\begin{aligned} \Psi^{*(1)}(H^{(0)} - \epsilon_i^{(0)})\Psi_i^{(2)} + \Psi^{*(1)}(H^{(1)} - \epsilon_i^{(1)})\Psi_i^{(1)} + \Psi^{*(1)}(H^{(2)} - \epsilon_i^{(2)})\Psi_i^{(0)} &= 0 \\ \langle \Psi_i^{(1)} | H^{(0)} - \epsilon_i^{(0)} | \Psi_i^{(2)} \rangle + \langle \Psi_i^{(1)} | H^{(1)} - \epsilon_i^{(1)} | \Psi_i^{(1)} \rangle + \langle \Psi_i^{(1)} | H^{(2)} - \epsilon_i^{(2)} | \Psi_i^{(0)} \rangle &= 0 \end{aligned}$$

or

$$- \langle \Psi_i^{(1)} | H^{(0)} - \epsilon_i^{(0)} | \Psi_i^{(2)} \rangle = \langle \Psi_i^{(1)} | H^{(1)} - \epsilon_i^{(1)} | \Psi_i^{(1)} \rangle + \langle \Psi_i^{(1)} | H^{(2)} - \epsilon_i^{(2)} | \Psi_i^{(0)} \rangle$$

Finally using this last expression we can replace it into the result from the second expression above to get

$$\begin{aligned} \epsilon_i^{(3)} &= - \langle \Psi_i^{(1)} | H^{(0)} - \epsilon_i^{(0)} | \Psi_i^{(2)} \rangle + \langle \Psi_i^{(0)} | H^{(2)} | \Psi_i^{(1)} \rangle + \langle \Psi_i^{(0)} | H^{(3)} | \Psi_i^{(0)} \rangle \\ &= [\langle \Psi_i^{(1)} | H^{(1)} - \epsilon_i^{(1)} | \Psi_i^{(1)} \rangle + \langle \Psi_i^{(1)} | H^{(2)} - \epsilon_i^{(2)} | \Psi_i^{(0)} \rangle] \\ &\quad + \langle \Psi_i^{(0)} | H^{(2)} | \Psi_i^{(1)} \rangle + \langle \Psi_i^{(0)} | H^{(3)} | \Psi_i^{(0)} \rangle \end{aligned}$$

The only term that simplifies is the second one in brackets. It reduces to $\langle \Psi_i^{(1)} | H^{(2)} | \Psi_i^{(0)} \rangle$ since $|\Psi_i^{(1)}\rangle$ and $|\Psi_i^{(0)}\rangle$ are orthogonal. Now we get for the 3rd order correction

$$\epsilon_i^{(3)} = \langle \Psi_i^{(1)} | H^{(1)} - \epsilon_i^{(1)} | \Psi_i^{(1)} \rangle + \langle \Psi_i^{(1)} | H^{(2)} | \Psi_i^{(0)} \rangle + \langle \Psi_i^{(0)} | H^{(2)} | \Psi_i^{(1)} \rangle + \langle \Psi_i^{(0)} | H^{(3)} | \Psi_i^{(0)} \rangle$$

It is apparent that all we need to know now is $|\Psi_i^{(1)}\rangle$

Using

$$\begin{aligned} |\Psi_i^{(1)}\rangle &= \sum_k \frac{\langle k | H^{(1)} | i \rangle}{(\epsilon_i^{(0)} - \epsilon_k^{(0)})} |k\rangle \\ \langle \Psi_i^{(1)}| &= \sum_j \frac{\langle i | H^{(1)} | j \rangle}{(\epsilon_i^{(0)} - \epsilon_j^{(0)})} \langle j| \end{aligned}$$

Now assume (to make life simple) that the Hamiltonian only has a first order perturbation $H^{(2)} = 0$ and $H^{(3)} = 0$. This leaves behind

$$\begin{aligned}
\epsilon_i^{(3)} &= \langle \Psi_i^{(1)} | H^{(1)} - \epsilon_i^{(1)} | \Psi_i^{(1)} \rangle \\
&= \langle \Psi_i^{(1)} | H^{(1)} | \Psi_i^{(1)} \rangle - \langle \Psi_i^{(1)} | \epsilon_i^{(1)} | \Psi_i^{(1)} \rangle \\
&= \langle \Psi_i^{(1)} | H^{(1)} | \Psi_i^{(1)} \rangle - \epsilon_i^{(1)} \langle \Psi_i^{(1)} | \Psi_i^{(1)} \rangle \\
&= \left[\sum_j \frac{\langle i | H^{(1)} | j \rangle}{\langle j |} \right] H^{(1)} \left[\sum_k \frac{\langle k | H^{(1)} | i \rangle}{(\epsilon_i^{(0)} - \epsilon_k^{(0)})} | k \rangle \right] - \epsilon_i^{(1)} \langle \Psi_i^{(1)} | \Psi_i^{(1)} \rangle \\
&= \sum_{kj \neq i} \frac{\langle i | H^{(1)} | j \rangle \langle j | H^{(1)} | k \rangle \langle k | H^{(1)} | i \rangle}{(\epsilon_i^{(0)} - \epsilon_j^{(0)})(\epsilon_i^{(0)} - \epsilon_k^{(0)})} - \epsilon_i^{(1)} \langle \Psi_i^{(1)} | \Psi_i^{(1)} \rangle \\
&= \sum_{kj \neq i} \frac{\langle i | H^{(1)} | j \rangle \langle j | H^{(1)} | k \rangle \langle k | H^{(1)} | i \rangle}{(\epsilon_i^{(0)} - \epsilon_j^{(0)})(\epsilon_i^{(0)} - \epsilon_k^{(0)})} \\
&\quad - \epsilon_i^{(1)} \left[\sum_j \frac{\langle i | H^{(1)} | j \rangle}{(\epsilon_i^{(0)} - \epsilon_j^{(0)})} \langle j | \right] \left[\sum_k \frac{\langle k | H^{(1)} | i \rangle}{(\epsilon_i^{(0)} - \epsilon_k^{(0)})} | k \rangle \right] \\
&= \sum_{kj \neq i} \frac{\langle i | H^{(1)} | j \rangle \langle j | H^{(1)} | k \rangle \langle k | H^{(1)} | i \rangle}{(\epsilon_i^{(0)} - \epsilon_j^{(0)})(\epsilon_i^{(0)} - \epsilon_k^{(0)})} \\
&\quad - \epsilon_i^{(1)} \sum_{kj \neq i} \frac{\langle i | H^{(1)} | j \rangle \langle j | k \rangle \langle k | H^{(1)} | i \rangle}{(\epsilon_i^{(0)} - \epsilon_j^{(0)})(\epsilon_i^{(0)} - \epsilon_k^{(0)})}
\end{aligned}$$

Now since we have $\langle j | k \rangle$ the only terms that are non-zero are those where $j = k$

$$= \sum_{kj \neq i} \frac{\langle i | H^{(1)} | j \rangle \langle j | H^{(1)} | k \rangle \langle k | H^{(1)} | i \rangle}{(\epsilon_i^{(0)} - \epsilon_j^{(0)})(\epsilon_i^{(0)} - \epsilon_k^{(0)})} - \epsilon_i^{(1)} \sum_{k \neq i} \frac{\langle i | H^{(1)} | k \rangle \langle k | H^{(1)} | i \rangle}{(\epsilon_i^{(0)} - \epsilon_k^{(0)})^2}$$

Therefore

$$\boxed{\epsilon_i^{(3)} = \sum_{kj \neq i} \frac{\langle i | H^{(1)} | j \rangle \langle j | H^{(1)} | k \rangle \langle k | H^{(1)} | i \rangle}{(\epsilon_i^{(0)} - \epsilon_j^{(0)})(\epsilon_i^{(0)} - \epsilon_k^{(0)})} - \epsilon_i^{(1)} \sum_{k \neq i} \frac{\langle i | H^{(1)} | k \rangle \langle k | H^{(1)} | i \rangle}{(\epsilon_i^{(0)} - \epsilon_k^{(0)})^2}} \quad (7.19)$$

Alternatively, $\epsilon_i^{(1)} = \langle i | H^{(1)} | i \rangle$ so that

$$\boxed{\epsilon_i^{(3)} = \sum_{kj \neq i} \frac{\langle i | H^{(1)} | j \rangle \langle j | H^{(1)} | k \rangle \langle k | H^{(1)} | i \rangle}{(\epsilon_i^{(0)} - \epsilon_j^{(0)})(\epsilon_i^{(0)} - \epsilon_k^{(0)})} - \langle i | H^{(1)} | i \rangle \sum_{k \neq i} \frac{\langle i | H^{(1)} | k \rangle \langle k | H^{(1)} | i \rangle}{(\epsilon_i^{(0)} - \epsilon_k^{(0)})}} \quad (7.20)$$

Chapter 8

Density of states

The idea here is that while the confinement of an axis gives rise to discrete energies (as we saw in the previous section on confinement), in cases such as the quantum well and quantum wire, there are additional states along the unconfined axes (degrees of freedom). Here it is not practical to try and calculate the energies of each state since they are actually continuous bands or energies. As a consequence one way to get an idea of what the energies look like is to perform the following density of states argument. So when put together with the confined energies from the last section, these density of states calculations provide a more thorough description of what the electronic structure of 3D, 2D, 1D and 0D materials look like.

3 Dimensions (bulk)

Consider the volume in “k” space

$$V_k = \frac{4}{3}\pi k^3$$

($k^2 = k_x^2 + k_y^2 + k_z^2$) where for a particle in this sphere

$$\begin{aligned}k_x &= \frac{2\pi}{L_x} \\k_y &= \frac{2\pi}{L_y} \\k_z &= \frac{2\pi}{L_z}\end{aligned}$$

Note that the 2π arises from the constraints of a periodic boundary condition (Born-Von Karman conditions, named after Max Born and Theodore Von

Karman) as opposed to the more general $n\pi$ where $n = 0, 1, 2, 3, \dots$. The volume of a given mode is then $k_x k_y k_z$. The number of modes (N) in the sphere are then

$$N = \frac{V_k}{k_x k_y k_z} = \frac{\frac{4}{3}\pi k^3}{8\pi^3} L_x L_y L_z$$

Say the particle is an electron and we consider the spin degeneracy (up and down), then we multiply N by 2.

$$\begin{aligned} N' &= 2N = 2 \frac{\frac{4}{3}\pi k^3}{8\pi^3} L_x L_y L_z \\ &= \frac{k^3}{3\pi^2} L_x L_y L_z \text{ total number of states in sphere} \end{aligned}$$

Consider the density

$$\rho = \frac{N'}{L_x L_y L_z} = \frac{k^3}{3\pi^2} \text{ number of states/unit volume}$$

Now consider the energy density defined as

$$\rho' = \frac{d\rho}{d\varepsilon} = \frac{d\left(\frac{k^3}{3\pi^2}\right)}{d\varepsilon} = \frac{1}{3\pi^2} \frac{dk^3}{d\varepsilon} \quad (8.1)$$

where recall $k = \sqrt{\frac{2m\varepsilon}{\hbar^2}}$ (units of 1 over length) for a free electron or alternatively $k^3 = \left(\frac{2m\varepsilon}{\hbar^2}\right)^{\frac{3}{2}}$. The expression becomes

$$\rho' = \frac{1}{3\pi^2} \left(\frac{2m}{\hbar^2}\right)^{\frac{3}{2}} \frac{d\varepsilon^{\frac{3}{2}}}{d\varepsilon}$$

$$\rho' = \frac{1}{3\pi^2} \left(\frac{2m}{\hbar^2}\right)^{\frac{3}{2}} \frac{3}{2} \varepsilon^{\frac{1}{2}}$$

$$\boxed{\rho'_{3D} = \frac{1}{2\pi^2} \left(\frac{2m}{\hbar^2}\right)^{\frac{3}{2}} \sqrt{\varepsilon}} \quad (8.2)$$

This is the “density of states” in 3 dimensions. Note the square root dependence on energy. Note also that the energies of the system are

$$\varepsilon = \frac{\hbar^2 k^2}{2m} \quad (8.3)$$

where $k = \sqrt{k_x^2 + k_y^2 + k_z^2}$ and $k_{x,y,z} = \frac{2\pi n}{L_{x,y,z}}$.

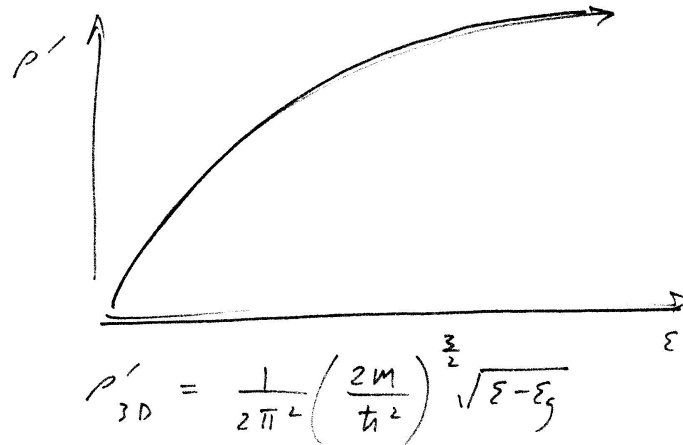


Figure 8.1: Density of states in a 3D material (bulk)

Exercises

1. Calculate the 3D density of states for free electrons with energy 0.1 eV. Express your answer in terms of eV and cm^3 .
2. Assume the electron resides in a non-zero potential V . Express the 3D density of states in this situation. Hint, just alter the expression for k below equation 5.1.

2 Dimensions (well)

Here we have 1 dimension that is quantized. Let's assume its the z direction. The total energy of this system is a sum of the energy along the quantized direction + the energy along the other 2 "free" directions. It is expressed as

$$\epsilon_{tot} = \frac{\hbar^2 k_z^2}{2m} + \frac{\hbar^2 k^2}{2m} = \epsilon_n + \epsilon_{x,y} \quad (8.4)$$

where $k^2 = k_x^2 + k_y^2$ and $k_z = \frac{n\pi}{L_z}$ Consider now an area in k space

$$A_k = \pi k^2$$

where for the particle

$$\begin{aligned}k_x &= \frac{2\pi}{L_x} \\k_y &= \frac{2\pi}{L_y}\end{aligned}$$

The area of a given mode is then $k_x k_y$ with the total number of modes (N) in the area being

$$N = \frac{\pi k^2}{4\pi^2} L_x L_y = \frac{k^2 L_x L_y}{4\pi}$$

Again if the particle is an electron and we consider spin, multiply by 2 to get

$$N' = 2N = \frac{k^2 L_x L_y}{2\pi} \text{ total number of states in area}$$

Consider now a density

$$\rho = \frac{N'}{L_x L_y} = \frac{k^2}{2\pi}$$

where recall that $k = \sqrt{\frac{2m\varepsilon}{\hbar^2}}$. Replacing this above gives

$$\rho = \frac{m\varepsilon}{\hbar^2 \pi} \text{ number of states per unit area}$$

Now consider the energy density

$$\rho' = \frac{d\rho}{d\varepsilon} = \frac{m}{\hbar^2 \pi}$$

This is the energy density of the subband for a given k_z (or ε_n). For each successive k_z there will be an additional $\frac{m}{\hbar^2 \pi}$ term and hence another subband. Therefore the density of states is written

$$\boxed{\rho'_{2D} = \frac{m}{\hbar^2 \pi} \sum_n \Theta(\varepsilon - \varepsilon_n)} \quad (8.5)$$

where Θ is the Heavyside function.

Exercises

1. Calculate the 2D density of states for free electrons with energy 0.1 eV. Consider only the lowest subband. Express your answer in eV and cm^3 units.

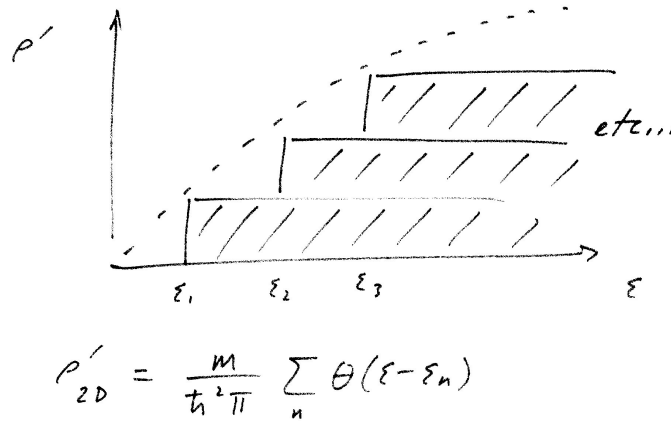


Figure 8.2: Density of states in a 2D material (quantum well)

1 Dimension (wire)

Consider now the situation where there are 2 dimensions confined and only 1 degree of freedom (say the x direction). The total energy of the system can be written as

$$\epsilon_{tot} = \frac{\hbar^2 k_z^2}{2m} + \frac{\hbar^2 k_y^2}{2m} + \frac{\hbar^2 k_x^2}{2m} = \epsilon_n + \epsilon_m + \epsilon_x \quad (8.6)$$

where $k = k_x = \frac{2\pi}{L_x}$. Furthermore along the confined directions, $k_z = \frac{n\pi}{L_z}$, $k_y = \frac{m\pi}{L_y}$ and m, n are integers. Consider a length $2k$. The number of modes along this length is

$$N = \frac{2k}{k_x} = \frac{2k}{\left(\frac{2\pi}{L_x}\right)} = \frac{kL_x}{\pi} \text{ number of state along the line}$$

Now if we consider an electron again, ensure to take into account the spin degeneracy

$$N' = 2N = \frac{2kL_x}{\pi}$$

Now a density is

$$\rho = \frac{N'}{L_x} = \frac{2k}{\pi} = \frac{2}{\pi} \sqrt{\frac{2m\epsilon}{\hbar^2}} \text{ number of states per unit length}$$

Consider the energy density

$$\begin{aligned}\rho' &= \frac{d\rho}{d\varepsilon} = \frac{2}{\pi} \sqrt{\frac{2m}{\hbar^2}} \frac{d\sqrt{\varepsilon}}{d\varepsilon} = \frac{2}{\pi} \sqrt{\frac{2m}{\hbar^2}} \frac{1}{2} \varepsilon^{-\frac{1}{2}} \\ &= \frac{1}{\pi} \sqrt{\frac{2m}{\hbar^2}} \frac{1}{\sqrt{\varepsilon}}\end{aligned}\quad (8.7)$$

This is the energy density for a given n, m value or $(\varepsilon_n, \varepsilon_m)$ combination). The complete expression taking into account all m, n combinations is

$$\boxed{\rho'_{1D} = \frac{1}{\pi} \sqrt{\frac{2m}{\hbar^2}} \sum_{n,m} \frac{1}{\sqrt{\varepsilon - \varepsilon_{n,m}}} \Theta(\varepsilon - \varepsilon_{n,m})} \quad (8.8)$$

where again, Θ is the Heavyside function. Notice the inverse square root dependence of the density of states with energy.

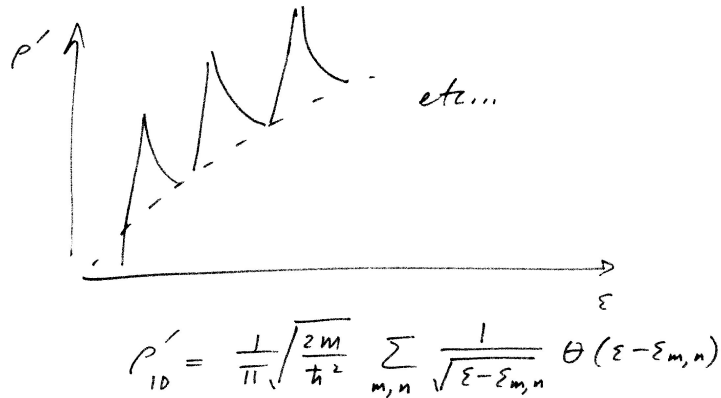


Figure 8.3: Density of states in a 1D material (quantum wire)

Exercises

1. Calculate the 1D density of states for free electrons with energy 0.1 eV above $\varepsilon_{n,m}$. Consider only the lowest subband. Express your result in units of eV and cm^3 .

0 Dimension (dot)

Here since all three dimensions are confined the density of states is basically a series of delta functions. The total energy of the system is

$$\varepsilon_{tot} = \frac{\hbar^2 k_x^2}{2m} + \frac{\hbar^2 k_y^2}{2m} + \frac{\hbar^2 k_z^2}{2m} = \varepsilon_m + \varepsilon_n + \varepsilon_o \quad (8.9)$$

where m, n, o are integers and $k_x = \frac{m\pi}{L_x}$, $k_y = \frac{n\pi}{L_y}$, $k_z = \frac{o\pi}{L_z}$. The density of states is

$$\rho'_{0D} = \delta(\varepsilon - \varepsilon_{m,n,o}) \quad (8.10)$$

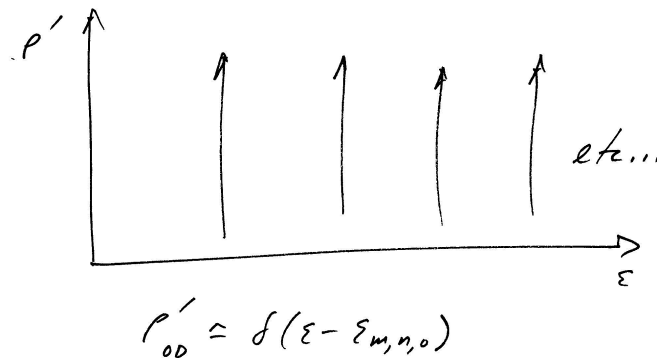


Figure 8.4: Density of states in a 0D material (quantum dot)

Exercises

1. Compare and contrast the density of states in the 0D case to the previous 3D, 2D and 1D cases.

Chapter 9

More density of states

Density of states in the conduction band

To do this calculation we need to know the probability that an electron will occupy a given state of energy ε . This probability, $P_e(\varepsilon)$, is referred to as the Fermi Dirac distribution. In addition, we need to know the density of states (ρ') which we calculated in the last section. Recall that this density of states has units of number per unit volume per unit energy. Therefore $\rho' d\varepsilon$ is the number of states per unit volume. The number of occupied states at a given energy per unit volume (or alternatively the concentration of electrons at a given energy) is therefore

$$n_e(\varepsilon) = P_e(\varepsilon)\rho'(\varepsilon)d\varepsilon$$

where the Fermi Dirac distribution is

$$P_e(\varepsilon) = \frac{1}{1 + e^{\frac{\varepsilon - \varepsilon_F}{kT}}} \quad (9.1)$$

Here ε_F is the Fermi energy (also called the chemical potential sometimes).

The total concentration of electrons in the conduction band is therefore the integral over all available energies

$$n_{tot} = \int_{\varepsilon_C}^{\infty} P_e(\varepsilon)\rho'(\varepsilon)d\varepsilon \quad (9.2)$$

where ε_C is the energy where the conduction band starts. Consider the density of states we just derived for a 3D (bulk) material in the previous section, ρ'_{3D} .

$$\rho'_{3D} = \frac{1}{2\pi^2} \left(\frac{2m_e}{\hbar^2} \right)^{\frac{3}{2}} \sqrt{\varepsilon}$$

Now rather than have the origin of the energy start at 0, start it where the conduction band begins (ε_C). This expression can then be recast as

$$\rho'_{3D} = \frac{1}{2\pi^2} \left(\frac{2m_e}{\hbar^2} \right)^{\frac{3}{2}} \sqrt{\varepsilon - \varepsilon_C} \quad (9.3)$$

Replace this and the Fermi Dirac distribution into the above expression for the total concentration of electrons in the conduction band to get

$$\begin{aligned} n_{tot} &= \int_{\varepsilon_C}^{\infty} \frac{1}{1 + e^{\frac{\varepsilon - \varepsilon_F}{kT}}} \frac{1}{2\pi^2} \left(\frac{2m_e}{\hbar^2} \right)^{\frac{3}{2}} \sqrt{\varepsilon - \varepsilon_C} d\varepsilon \\ &= \frac{1}{2\pi^2} \left(\frac{2m_e}{\hbar^2} \right)^{\frac{3}{2}} \int_{\varepsilon_C}^{\infty} \frac{1}{1 + e^{\frac{\varepsilon - \varepsilon_F}{kT}}} \sqrt{\varepsilon - \varepsilon_C} d\varepsilon \\ &= A \int_{\varepsilon_C}^{\infty} \frac{1}{1 + e^{\frac{\varepsilon - \varepsilon_F}{kT}}} \sqrt{\varepsilon - \varepsilon_C} d\varepsilon \\ &= A \int_{\varepsilon_C}^{\infty} \frac{1}{1 + e^{\frac{\varepsilon - \varepsilon_C + \varepsilon_C - \varepsilon_F}{kT}}} \sqrt{\varepsilon - \varepsilon_C} d\varepsilon \\ &= A \int_{\varepsilon_C}^{\infty} \frac{1}{1 + e^{\frac{\varepsilon - \varepsilon_C}{kT}} e^{\frac{\varepsilon_C - \varepsilon_F}{kT}}} \sqrt{\varepsilon - \varepsilon_C} d\varepsilon \end{aligned} \quad (9.4)$$

Here, let $\eta = \frac{\varepsilon - \varepsilon_C}{kT}$ and $\mu = \frac{-\varepsilon_C + \varepsilon_F}{kT}$ leading to

$$n_{tot} = A' \int_{\varepsilon_C}^{\infty} \frac{1}{e^{\eta - \mu}} \sqrt{\eta} d\eta$$

where $A = \frac{1}{2\pi^2} \left(\frac{2m_e}{\hbar^2} \right)^{\frac{3}{2}}$ and A' can be found from the above change of variables. Note at this point that someone has already solved this integral and one can just look it up. The integral is called the Fermi integral or Fermi Dirac integral. It is defined as follows and its solutions labeled $F_{\frac{1}{2}}(\eta)$ can be looked up.

$$F_{\frac{1}{2}}(\eta) = \int_0^{\infty} \frac{\sqrt{\eta}}{1 + e^{\eta - \mu}} d\eta$$

However, to stay instructive let's just consider the case where $\varepsilon - \varepsilon_F \gg kT$. In this case the exponential in the denominator of the Fermi Dirac distribution will dominate and the expression basically becomes

$$\frac{1}{1 + e^{\frac{\varepsilon - \varepsilon_F}{kT}}} \simeq \frac{1}{e^{\frac{\varepsilon - \varepsilon_F}{kT}}} = e^{-\frac{\varepsilon - \varepsilon_F}{kT}}$$

which has the Boltzman distribution form. Our expression for n_{tot} becomes

$$\begin{aligned} n_{tot} &= A \int_{\varepsilon_C}^{\infty} e^{-\frac{\varepsilon-\varepsilon_F}{kT}} \sqrt{\varepsilon - \varepsilon_C} d\varepsilon \\ &= A \int_{\varepsilon_0}^{\infty} e^{-\frac{[(\varepsilon-\varepsilon_C)+(\varepsilon_C-\varepsilon_F)]}{kT}} \sqrt{\varepsilon - \varepsilon_C} d\varepsilon \end{aligned} \quad (9.5)$$

Now change variables and let $x = \frac{\varepsilon-\varepsilon_C}{kT}$ such that $\varepsilon = \varepsilon_C + xkT$ and $d\varepsilon = kTdx$. Note also that the limits of integration will change accordingly. This leads to

$$\begin{aligned} n_{tot} &= A \int_0^{\infty} e^{-\frac{1}{kT}[(\varepsilon-\varepsilon_C)+(\varepsilon_C-\varepsilon_F)]} (xkT)^{\frac{1}{2}} kT dx \\ &= A(kT)^{\frac{3}{2}} e^{-\frac{\varepsilon_C-\varepsilon_F}{kT}} \int_0^{\infty} e^{-x} x^{\frac{1}{2}} dx \end{aligned}$$

The last integral is a common function called the Gamma function. It is defined as follows

$$\Gamma(n) = \int_0^{\infty} e^{-x} x^{n-1} dx$$

and its values can be looked up in most reference books. In our case we have $\Gamma\left(\frac{3}{2}\right)$. Therefore n_{tot} can be expressed as

$$n_C \equiv n_{tot} = A(kT)^{\frac{3}{2}} e^{-\frac{\varepsilon_C-\varepsilon_F}{kT}} \Gamma\left(\frac{3}{2}\right)$$

To be consistent with other texts, we use n_C in lieu of n_{tot} at this point and also define $N_C = A(kT)^{\frac{3}{2}} \Gamma\left(\frac{3}{2}\right)$. This leads to the common textbook expression

$$\boxed{n_C = N_C e^{-\frac{\varepsilon_C-\varepsilon_F}{kT}}} \quad (9.6)$$

This is the expression for the effective density of states of the conduction band.

Density of states in the valence band

The way this is calculated mirrors the approach for the conduction band with a few minor changes. Again, we need the probability of finding a hole at a given energy $P_h(\varepsilon)$. Likewise we need the density of states calculated

in the previous section. We will calculate the number of holes at a given energy per unit volume (or concentration of holes at a given energy)

$$n_h(\varepsilon) = P_h(\varepsilon)\rho'(\varepsilon)d\varepsilon$$

where we use the relationship

$$P_e(\varepsilon) + P_h(\varepsilon) = 1$$

or

$$P_h(\varepsilon) = 1 - P_e(\varepsilon)$$

This lead to

$$P_h(\varepsilon) = 1 - \frac{1}{1 + e^{\frac{\varepsilon - \varepsilon_F}{kT}}} \quad (9.7)$$

The density of states is

$$\rho'_{3D} = \frac{1}{2\pi^2} \left(\frac{2m_h}{\hbar^2} \right)^{\frac{3}{2}} \sqrt{\varepsilon_V - \varepsilon} \quad (9.8)$$

where ε_V is the energy where the valence band starts. Consequently the total concentration of holes in the valence band is the integral over all energies.

$$\begin{aligned} p_{tot} &= \int_{-\infty}^{\varepsilon_V} P_h(\varepsilon)\rho'(\varepsilon)d\varepsilon \\ &= \int_{-\infty}^{\varepsilon_V} \left(1 - \frac{1}{1 + e^{\frac{\varepsilon - \varepsilon_F}{kT}}} \right) \left(\frac{1}{2\pi^2} \right) \left(\frac{2m_h}{\hbar^2} \right)^{\frac{3}{2}} \sqrt{\varepsilon_V - \varepsilon} d\varepsilon \end{aligned}$$

For notational convenience let $B = \frac{1}{2\pi^2} \left(\frac{2m_h}{\hbar^2} \right)^{\frac{3}{2}}$. This leads to

$$p_{tot} = B \int_{-\infty}^{\varepsilon_V} \left(1 - \frac{1}{1 + e^{\frac{\varepsilon - \varepsilon_F}{kT}}} \right) \sqrt{\varepsilon_V - \varepsilon} d\varepsilon$$

Since, generally speaking $\varepsilon < \varepsilon_F$

$$p_{tot} = B \int_{-\infty}^{\varepsilon_V} \left(1 - \frac{1}{1 + e^{-\frac{\varepsilon_F - \varepsilon}{kT}}} \right) \sqrt{\varepsilon_V - \varepsilon} d\varepsilon$$

Approximate the stuff in parenthesis through the binomial expansion keeping only the first 2 terms (where $|x| < 1$).

$$\begin{aligned}\frac{1}{1+x} &\simeq 1 - x + x^2 - x^3 + x^4 + \dots \\ \frac{1}{1 + e^{-\frac{\varepsilon_F - \varepsilon}{kT}}} &\simeq 1 - e^{-\frac{\varepsilon_F - \varepsilon}{kT}} + \dots\end{aligned}$$

Therefore

$$\left(1 - \frac{1}{1 + e^{-\frac{\varepsilon_F - \varepsilon}{kT}}}\right) \simeq 1 - \left(1 - e^{-\frac{\varepsilon_F - \varepsilon}{kT}}\right) = e^{-\frac{\varepsilon_F - \varepsilon}{kT}}$$

Replacing this in the p_{tot} expression gives

$$p_{tot} = B \int_{-\infty}^{\varepsilon_V} e^{-\frac{\varepsilon_F - \varepsilon}{kT}} \sqrt{\varepsilon_V - \varepsilon} d\varepsilon$$

As with the conduction band case earlier, make a change of variables by letting $x = \frac{\varepsilon_V - \varepsilon}{kT}$ as well as $\varepsilon = \varepsilon_V - kTx$ and $d\varepsilon = -kTdx$. Note that the limits of integration change accordingly resulting in

$$\begin{aligned}p_{tot} &= B \int_{\infty}^0 e^{-\frac{1}{kT}[(\varepsilon_F - \varepsilon_V) + (\varepsilon_V - \varepsilon)]} \sqrt{kTx} (-kT) dx \\ &= B \int_0^{\infty} e^{-\frac{\varepsilon_F - \varepsilon_V}{kT}} e^{-\frac{\varepsilon_V - \varepsilon}{kT}} (kT)^{\frac{3}{2}} x^{\frac{1}{2}} dx \\ &= B(kT)^{\frac{3}{2}} e^{-\frac{\varepsilon_F - \varepsilon}{kT}} \int_0^{\infty} e^{-x} x^{\frac{1}{2}} dx\end{aligned}$$

Recall that the last integral above is the Gamma function. We get the following expression upon recognizing this

$$p_{tot} = B(kT)^{\frac{3}{2}} e^{-\frac{\varepsilon_F - \varepsilon}{kT}} \Gamma\left(\frac{3}{2}\right)$$

To be consistent with common notation and other texts p_{tot} can be expressed as

$$\boxed{p_V \equiv p_{tot} = N_v e^{-\frac{\varepsilon_F - \varepsilon_V}{kT}}} \quad (9.9)$$

where $N_v = B(kT)^{\frac{3}{2}} \Gamma\left(\frac{3}{2}\right)$. The last expression gives the effective density of states for the valence band.

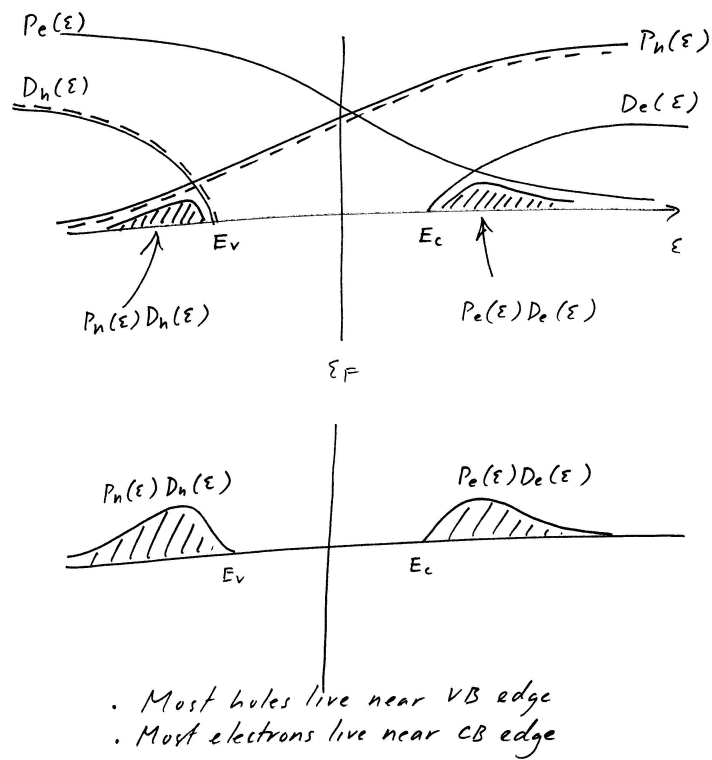


Figure 9.1: Sketch of the electron and hole distribution functions, their density of states and their population at the conduction and valence band edges.

Summary

Fermi level of an intrinsic semiconductor

If the bulk semiconductor is intrinsic, there has been no doping of the material and hence no extra electrons or holes anywhere. In this situation

$$n_C = p_V$$

where previously we found that

$$n_C = N_C e^{-\frac{\epsilon_C - \epsilon_F}{kT}} \text{ and } N_C = A(kT)^{\frac{3}{2}} \Gamma\left(\frac{3}{2}\right)$$

$$p_V = N_V e^{-\frac{\epsilon_F - \epsilon_V}{kT}} \text{ and } N_V = B(kT)^{\frac{3}{2}} \Gamma\left(\frac{3}{2}\right)$$

This leads to

$$N_C e^{-\frac{\epsilon_C - \epsilon_F}{kT}} = N_V e^{-\frac{\epsilon_F - \epsilon_V}{kT}}$$

or

$$A e^{-\frac{\epsilon_C - \epsilon_F}{kT}} = B e^{-\frac{\epsilon_F - \epsilon_V}{kT}}$$

where $A = \frac{1}{2\pi^2} \left(\frac{2m_e}{\hbar^2} \right)^{\frac{3}{2}}$ and $B = \frac{1}{2\pi^2} \left(\frac{2m_h}{\hbar^2} \right)^{\frac{3}{2}}$ and ultimately reduces to

$$\begin{aligned}
 m_e^{\frac{3}{2}} e^{-\frac{\varepsilon_C - \varepsilon_F}{kT}} &= m_h^{\frac{3}{2}} e^{-\frac{\varepsilon_F - \varepsilon_V}{kT}} \\
 e^{-\frac{\varepsilon_C - \varepsilon_F}{kT}} e^{\frac{\varepsilon_F - \varepsilon_V}{kT}} m_e^{\frac{3}{2}} &= m_h^{\frac{3}{2}} \\
 e^{\frac{1}{kT}(-\varepsilon_C + \varepsilon_F + \varepsilon_F - \varepsilon_V)} &= \left(\frac{m_h}{m_e} \right)^{\frac{3}{2}} \\
 e^{\frac{2\varepsilon_F}{kT}} e^{-\frac{\varepsilon_C + \varepsilon_V}{kT}} &= \left(\frac{m_h}{m_e} \right)^{\frac{3}{2}} \\
 e^{\frac{2\varepsilon_F}{kT}} &= \left(\frac{m_h}{m_e} \right)^{\frac{3}{2}} e^{\frac{\varepsilon_C + \varepsilon_V}{kT}} \\
 \ln \left(e^{\frac{2\varepsilon_F}{kT}} \right) &= \ln \left(\frac{m_h}{m_e} \right)^{\frac{3}{2}} + \ln \left(e^{\frac{\varepsilon_C + \varepsilon_V}{kT}} \right) \\
 \frac{2\varepsilon_F}{kT} &= \ln \left(\frac{m_h}{m_e} \right)^{\frac{3}{2}} + \frac{\varepsilon_C + \varepsilon_V}{kT} \\
 2\varepsilon_F &= kT \ln \left(\frac{m_h}{m_e} \right)^{\frac{3}{2}} + \varepsilon_C + \varepsilon_V \\
 \varepsilon_F &= \frac{kT}{2} \ln \left(\frac{m_h}{m_e} \right)^{\frac{3}{2}} + \frac{\varepsilon_C + \varepsilon_V}{2}
 \end{aligned}$$

This yields the final expression

$$\boxed{\varepsilon_F = \frac{\varepsilon_C + \varepsilon_V}{2} + \frac{3}{4} kT \ln \left(\frac{m_h}{m_e} \right)^{\frac{3}{2}}} \quad (9.10)$$

One can therefore see that at $T = 0$ the Fermi energy of an intrinsic semiconductor is at the halfway point between the top of the valence band and the bottom of the conduction band. Note also that since generally speaking $m_h > m_e$ the temperature dependent term is positive and slowly increases the Fermi level position with increasing temperature. To a good approximation however, the Fermi level is at the midway point between valence and conduction bands.

Exercises

- Calculate the intrinsic Fermi level of silicon at 0K, 10K, 77K, 300K and 600K. Note that $E_g = E_c + E_v$ and assume $m_e = 1.08m_0$ and

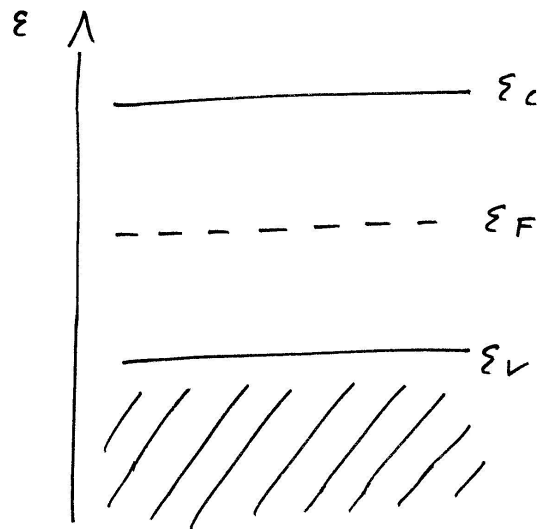


Figure 9.2: Sketch showing the Fermi level position at 0 degrees halfway between the conduction band and valence band positions.

$m_h = 0.55m_o$. Leave the answer in terms of E_g or if you desire look up the actual value of E_g and express the final answers in units of eV.

Chapter 10

Even more density of states

In the previous section we ran through the calculation for the conduction and valence band density of states for a 3D (bulk) material. In this section we repeat the same calculations for materials with lower dimensions. In particular we do this for 2D and 1D materials that are representative of quantum wells and quantum wires respectively.

Density of states in the conduction band: 2D

We start with the Fermi Dirac distribution for electrons and also the density of states (ρ'_{2D}) that we derived in an earlier section. Recall that

$$\rho'_{2D} = \frac{m_e}{\hbar^2 \pi} \sum_n \Theta(\varepsilon - \varepsilon_n)$$

Consider only one of the subbands. For example, we could choose the first band.

In this case the density of states simplifies to

$$\rho'_{2D} = \frac{m_e}{\hbar^2 \pi}$$

Now recall from the previous section that the number of states at a given energy per unit volume (or the concentration of electrons at a given energy) is

$$n_e(\varepsilon) = P_e(\varepsilon) \rho'(\varepsilon) d\varepsilon$$

The total concentration of electrons in this first subband is the integral over all available energies. Rather than use n_{tot} as before let's just stick to n_c

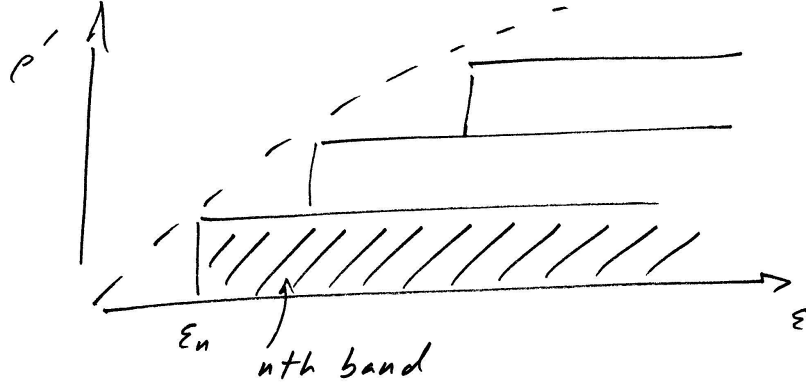


Figure 10.1: Conduction band density of states for a 2D material. The n th band is shaded to distinguish it from the other subbands.

from the start

$$n_c = \int_{\epsilon_C}^{\infty} P_e(\epsilon) \rho'(\epsilon) d\epsilon \quad (10.1)$$

where $P_e(\epsilon) = \frac{1}{1 + e^{\frac{\epsilon - \epsilon_F}{kT}}}$ is the Fermi Dirac distribution. Putting everything together we get

$$\begin{aligned} n_C &= \int_{\epsilon_C}^{\infty} \left(\frac{1}{1 + e^{\frac{\epsilon - \epsilon_F}{kT}}} \right) \frac{m_e}{\hbar^2 \pi} d\epsilon \\ &= \frac{m_e}{\hbar^2 \pi} \int_{\epsilon_C}^{\infty} \frac{1}{1 + e^{\frac{\epsilon - \epsilon_F}{kT}}} d\epsilon \end{aligned}$$

Since the band really begins at ϵ_n as opposed to ϵ_C like in the bulk the integral changes from

$$\int_{\epsilon_C}^{\infty} \rightarrow \int_{\epsilon_n}^{\infty}$$

leading to

$$n_C = \frac{m_e}{\hbar^2 \pi} \int_{\epsilon_n}^{\infty} \frac{d\epsilon}{1 + e^{\frac{\epsilon - \epsilon_F}{kT}}}$$

if now $\varepsilon - \varepsilon_F \gg kT$

$$n_C = \frac{m_e}{\hbar^2 \pi} \int_{\varepsilon_n}^{\infty} e^{-\frac{\varepsilon - \varepsilon_F}{kT}} d\varepsilon = - \frac{m_e kT}{\hbar^2 \pi} e^{-\frac{\varepsilon - \varepsilon_F}{kT}} \Big|_{\varepsilon_n}^{\infty}$$

This leads to the final expression for the carrier density of the nth subband, which in our example we assumed was the lowest one. Hence this is the carrier density at the conduction band edge of a 2D material.

$$\boxed{n_C = \frac{m_e kT}{\hbar^2 \pi} e^{-\frac{\varepsilon_n - \varepsilon_F}{kT}}} \quad (10.2)$$

Density of states in the valence band:2D

As with the conduction band case we need the probability of occupying a given state in the valence band. This is denoted $P_h(\varepsilon)$ and is evaluated from

$$\begin{aligned} P_e(\varepsilon) + P_h(\varepsilon) &= 1 \\ P_h(\varepsilon) &= 1 - P_e(\varepsilon) \end{aligned}$$

where $P_e(\varepsilon)$ is the Fermi Dirac distribution leading to

$$P_h(\varepsilon) = 1 - \frac{1}{1 + e^{\frac{\varepsilon - \varepsilon_F}{kT}}}$$

The number of states at a given energy per unit volume (concentration at a given energy) is

$$n_h(\varepsilon) = P_h(\varepsilon) \rho'(\varepsilon) d\varepsilon$$

If, as in the conduction band case, we consider only the nth subband $\rho' = \frac{m_h}{\hbar^2 \pi}$ and for simplicity the first.

The total concentration of holes in this first subband is the integral over all energies. We get

$$\begin{aligned} p_V &= \int_{-\infty}^{\varepsilon_V} P_h(\varepsilon) \rho'(\varepsilon) d\varepsilon & (10.3) \\ &= \int_{-\infty}^{\varepsilon_V} \left(1 - \frac{1}{1 + e^{\frac{\varepsilon - \varepsilon_F}{kT}}} \right) \left(\frac{m_h}{\hbar^2 \pi} \right) d\varepsilon \\ &= \frac{m_h}{\hbar^2 \pi} \int_{-\infty}^{\varepsilon_V} \left(1 - \frac{1}{1 + e^{\frac{\varepsilon - \varepsilon_F}{kT}}} \right) d\varepsilon \end{aligned}$$

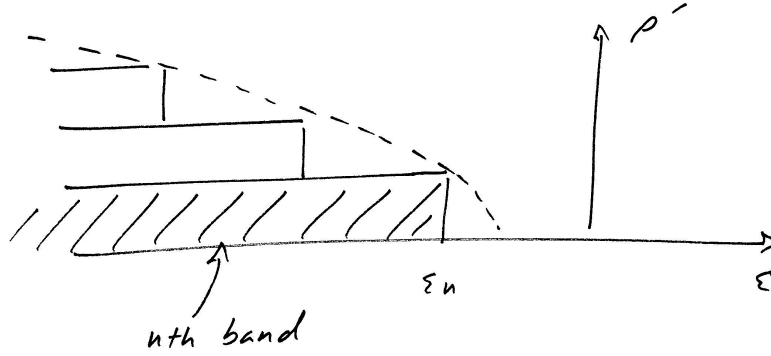


Figure 10.2: Valence band density of states for a 2D material. The n th band is shaded to distinguish it from the other subbands.

Since the n th (in our case the first) subband begins at ε_n rather than ε_V as in the bulk, the limits of the integral change

$$\int_{-\infty}^{\varepsilon_V} \rightarrow \int_{-\infty}^{\varepsilon_n}$$

The expression becomes

$$p_V = \frac{m_h}{\hbar^2 \pi} \int_{-\infty}^{\varepsilon_n} \left(1 - \frac{1}{1 + e^{\frac{\varepsilon - \varepsilon_F}{kT}}} \right) d\varepsilon$$

Since for the valence band $\varepsilon < \varepsilon_F$

$$p_V = \frac{m_h}{\hbar^2 \pi} \int_{-\infty}^{\varepsilon_n} \left(1 - \frac{1}{1 + e^{-\frac{\varepsilon_F - \varepsilon}{kT}}} \right) d\varepsilon$$

Apply the binomial expansion of the term in parenthesis to simplify, keeping only the first two terms. Recall that

$$\begin{aligned} \frac{1}{1+x} &= 1 - x + x^2 - x^3 + \dots \\ \frac{1}{1 + e^{-\frac{\varepsilon_F - \varepsilon}{kT}}} &= 1 - e^{-\frac{\varepsilon_F - \varepsilon}{kT}} + \dots \end{aligned}$$

such that

$$p_V = \frac{m_h}{\hbar^2 \pi} \int_{-\infty}^{\varepsilon_n} e^{-\frac{\varepsilon_F - \varepsilon}{kT}} d\varepsilon = \frac{m_h kT}{\hbar^2 \pi} e^{-\frac{\varepsilon_F - \varepsilon}{kT}} \Big|_{-\infty}^{\varepsilon_n}$$

This reduces to our desired final expression

$$\boxed{p_V = \frac{m_h kT}{\hbar^2 \pi} e^{-\frac{\varepsilon_F - \varepsilon_n}{kT}}} \quad (10.4)$$

Fermi level position:2D

The procedure for finding the Fermi level position is the same as in the 3D case. If we assume an intrinsic semiconductor with no doping such that there are no additional electrons or holes present

$$n_C = p_V$$

then

$$\frac{m_e kT}{\hbar^2 \pi} e^{-\frac{\varepsilon_{n1} - \varepsilon_F}{kT}} = \frac{m_h kT}{\hbar^2 \pi} e^{-\frac{\varepsilon_F - \varepsilon_{n2}}{kT}}$$

Note that n_1 and n_2 have been used to distinguish the start of the conduction band subband and the valence band subband respectively. The above expression reduces

$$\begin{aligned} e^{\frac{1}{kT}(-\varepsilon_{n1} + \varepsilon_F + \varepsilon_F - \varepsilon_{n2})} &= \left(\frac{m_h}{m_e}\right) \\ e^{-\frac{1}{kT}(\varepsilon_{n1} + \varepsilon_{n2}) + \frac{2\varepsilon_F}{kT}} &= \left(\frac{m_h}{m_e}\right) \\ e^{-\frac{\varepsilon_{n1} + \varepsilon_{n2}}{kT}} e^{\frac{2\varepsilon_F}{kT}} &= \left(\frac{m_h}{m_e}\right) \\ e^{\frac{2\varepsilon_F}{kT}} &= \left(\frac{m_h}{m_e}\right) e^{\frac{\varepsilon_{n1} + \varepsilon_{n2}}{kT}} \\ \ln\left(e^{\frac{2\varepsilon_F}{kT}}\right) &= \ln\left(\frac{m_h}{m_e}\right) + \ln\left(e^{\frac{\varepsilon_{n1} + \varepsilon_{n2}}{kT}}\right) \\ \frac{2\varepsilon_F}{kT} &= \ln\left(\frac{m_h}{m_e}\right) + \frac{\varepsilon_{n1} + \varepsilon_{n2}}{kT} \\ 2\varepsilon_F &= kT \ln\left(\frac{m_h}{m_e}\right) + \varepsilon_{n1} + \varepsilon_{n2} \end{aligned}$$

resulting in the final expression

$$\boxed{\varepsilon_F = \frac{\varepsilon_{n1} + \varepsilon_{n2}}{2} + \frac{kT}{2} \ln\left(\frac{m_h}{m_e}\right)} \quad (10.5)$$

As before, since $m_h > m_e$ the second term is positive and grows slowly with increasing temperature. To a first approximation however the Fermi level is midway between the valence band beginning and conduction band beginning.

Density of states in the conduction band:1D

We start with the Fermi Dirac distribution for electrons and the density of states (ρ'_{1D}) that we derived earlier. Recall that

$$\rho'_{1D} = \frac{1}{\pi} \sqrt{\frac{2m_e}{\hbar^2}} \sum_{m,n} \frac{1}{\sqrt{\varepsilon - \varepsilon_{m,n}}} \Theta(\varepsilon - \varepsilon_{m,n})$$

where Θ is the Heavyside function.

We consider only the (m,n) th subband beginning at energy $\varepsilon_{m,n}$. For convenience we could say this is the first subband.

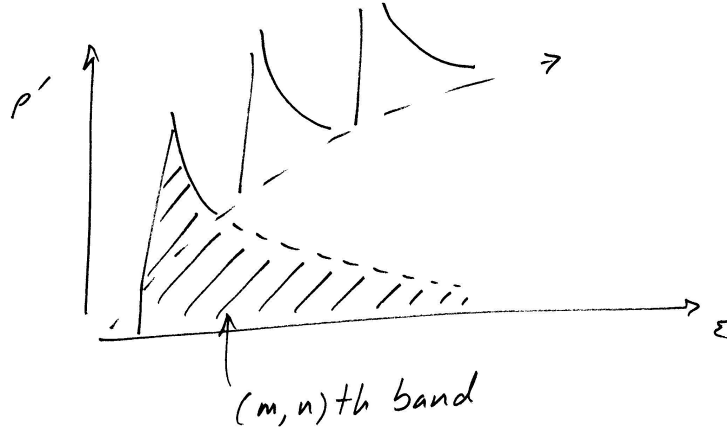


Figure 10.3: Conduction band density of states for a 1D material. The (m,n) th band is shaded to distinguish it from the other subbands.

The number of states at a given energy per unit volume (or the concentration at a given energy) is

$$n_e(\varepsilon) = P_e(\varepsilon) \rho'(\varepsilon) d\varepsilon$$

The total concentration in the first subband is the integral over all possible energies

$$n_c = \int_{\varepsilon_C}^{\infty} P_e(\varepsilon) \rho'(\varepsilon) d\varepsilon$$

where as before the Fermi Dirac distribution is

$$P_e(\varepsilon) = \frac{1}{1 + e^{\frac{\varepsilon - \varepsilon_F}{kT}}}$$

On introducing this, the total expression becomes

$$\begin{aligned} n_C &= \int_{\varepsilon_C}^{\infty} \frac{1}{1 + e^{\frac{\varepsilon - \varepsilon_F}{kT}}} \frac{1}{\pi} \sqrt{\frac{2m_e}{\hbar^2}} \frac{1}{\sqrt{\varepsilon - \varepsilon_{m,n}}} d\varepsilon \\ &= \frac{1}{\pi} \sqrt{\frac{2m_e}{\hbar^2}} \int_{\varepsilon_C}^{\infty} \frac{1}{1 + e^{\frac{\varepsilon - \varepsilon_F}{kT}}} \frac{1}{\sqrt{\varepsilon - \varepsilon_{m,n}}} d\varepsilon \end{aligned}$$

Since the band actually begins at $\varepsilon_{m,n}$ the limits of integration become

$$\int_{\varepsilon_C}^{\infty} \rightarrow \int_{\varepsilon_{m,n}}^{\infty}$$

Therefore

$$n_C = \frac{1}{\pi} \sqrt{\frac{2m_e}{\hbar^2}} \int_{\varepsilon_{m,n}}^{\infty} \left(\frac{1}{1 + e^{\frac{\varepsilon - \varepsilon_F}{kT}}} \right) \frac{1}{\sqrt{\varepsilon - \varepsilon_{m,n}}} d\varepsilon$$

If $\varepsilon - \varepsilon_F \gg kT$

$$\frac{1}{1 + e^{\frac{\varepsilon - \varepsilon_F}{kT}}} \rightarrow e^{-\frac{\varepsilon - \varepsilon_F}{kT}}$$

leading to

$$\begin{aligned} n_C &= \frac{1}{\pi} \sqrt{\frac{2m_e}{\hbar^2}} \int_{\varepsilon_{m,n}}^{\infty} e^{-\frac{\varepsilon - \varepsilon_F}{kT}} \frac{1}{\sqrt{\varepsilon - \varepsilon_{m,n}}} d\varepsilon \\ &= \frac{1}{\pi} \sqrt{\frac{2m_e}{\hbar^2}} \int_{\varepsilon_{m,n}}^{\infty} e^{-\frac{\varepsilon - \varepsilon_{m,n} + \varepsilon_{m,n} - \varepsilon_F}{kT}} \frac{1}{\sqrt{\varepsilon - \varepsilon_{m,n}}} d\varepsilon \\ &= \frac{1}{\pi} \sqrt{\frac{2m_e}{\hbar^2}} \int_{\varepsilon_{m,n}}^{\infty} e^{-\frac{\varepsilon - \varepsilon_{m,n}}{kT}} e^{-\frac{\varepsilon_{m,n} - \varepsilon_F}{kT}} \frac{1}{\sqrt{\varepsilon - \varepsilon_{m,n}}} d\varepsilon \\ &= \frac{1}{\pi} \sqrt{\frac{2m_e}{\hbar^2}} e^{-\frac{\varepsilon_{m,n} - \varepsilon_F}{kT}} \int_{\varepsilon_{m,n}}^{\infty} e^{-\frac{\varepsilon - \varepsilon_{m,n}}{kT}} \frac{1}{\sqrt{\varepsilon - \varepsilon_{m,n}}} d\varepsilon \end{aligned}$$

Make a change of variables and let $x = \frac{\varepsilon - \varepsilon_{m,n}}{kT}$ such that $\varepsilon = \varepsilon_{m,n} + kTx$ and $d\varepsilon = kT dx$. Also make corresponding changes to the limits of integration resulting in

$$n_C = \frac{\sqrt{kT}}{\pi} \sqrt{\frac{2m_e}{\hbar^2}} e^{-\frac{\varepsilon_{m,n} - \varepsilon_F}{kT}} \int_0^{\infty} e^{-x} x^{-\frac{1}{2}} dx$$

Recall that the last integral is the Gamma function

$$n_C = \frac{\sqrt{kT}}{\pi} \sqrt{\frac{2m_e}{\hbar^2}} e^{-\frac{\varepsilon_{m,n} - \varepsilon_F}{kT}} \Gamma\left(\frac{1}{2}\right)$$

The final expression for the carrier concentration of a 1D material at the conduction band edge is

$$\boxed{n_C = \frac{1}{\pi} \sqrt{\frac{2m_e kT}{\hbar^2}} e^{-\frac{\varepsilon_{m,n} - \varepsilon_F}{kT}} \Gamma\left(\frac{1}{2}\right)} \quad (10.6)$$

Density of states in the valence band:1D

The way this is calculated is very similar to that for the conduction band. However, a few of the terms change slightly in appearance. As before we need the probability of occupying a given state in the valence band. We also need the density of states calculated in the previous section. The occupation probability is calculated from

$$\begin{aligned} P_e(\varepsilon) + P_h(\varepsilon) &= 1 \\ P_h(\varepsilon) &= 1 - P_e(\varepsilon) \end{aligned}$$

Here $P_e(\varepsilon)$ is the Fermi Dirac distribution leading to

$$P_h(\varepsilon) = 1 - \frac{1}{1 + e^{\frac{\varepsilon - \varepsilon_F}{kT}}}$$

The density of states is

$$\rho'_{1D} = \frac{1}{\pi} \sqrt{\frac{2m_h}{\hbar^2}} \sum_{m,n} \frac{1}{\sqrt{\varepsilon_{m,n} - \varepsilon}} \Theta(\varepsilon_{m,n} - \varepsilon)$$

where Θ is the Heavyside function and the band explicitly starts at $\varepsilon_{m,n}$ causing its presence in the square root denominator. Now if we consider only the (m,n) th band, (and for convenience that could be the first one) the expression simplifies to

$$\rho'_{1D} = \frac{1}{\pi} \sqrt{\frac{2m_h}{\hbar^2}} \frac{1}{\sqrt{\varepsilon_{m,n} - \varepsilon}}$$

The number of states at a given energy per unit volume (or concentration at a given energy) is

$$n_h(\varepsilon) = P_h(\varepsilon) \rho'(\varepsilon) d\varepsilon$$

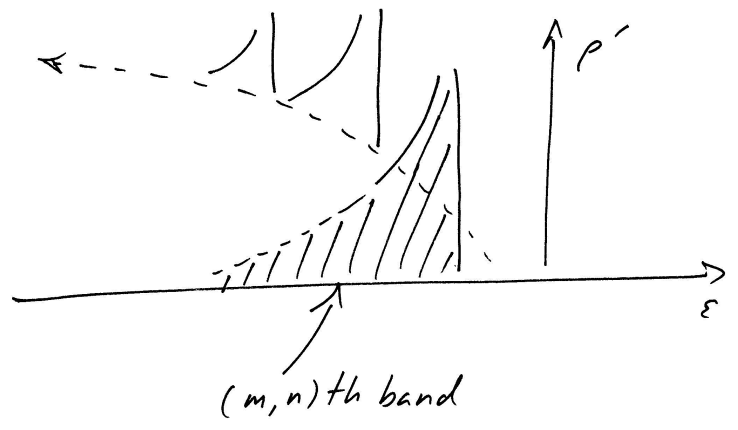


Figure 10.4: Valence band density of states for a 1D material. The (m,n)th band is shaded to distinguish it from the other subbands.

The total concentration in the subband is the integral over all energies

$$\begin{aligned}
 p_V &= \int_{-\infty}^{\varepsilon_V} P_h(\varepsilon) \rho'(\varepsilon) d\varepsilon & (10.7) \\
 &= \int_{-\infty}^{\varepsilon_V} \left(1 - \frac{1}{1 + e^{\frac{\varepsilon - \varepsilon_F}{kT}}} \right) \frac{1}{\pi} \sqrt{\frac{2m_h}{\hbar^2}} \frac{1}{\sqrt{\varepsilon_{m,n} - \varepsilon}} d\varepsilon \\
 &= \frac{1}{\pi} \sqrt{\frac{2m_h}{\hbar^2}} \int_{-\infty}^{\varepsilon_V} \left(1 - \frac{1}{1 + e^{\frac{\varepsilon - \varepsilon_F}{kT}}} \right) \frac{1}{\sqrt{\varepsilon_{m,n} - \varepsilon}} d\varepsilon
 \end{aligned}$$

Since the band begins at $\varepsilon_{m,n}$ rather than at ε_V as in the bulk, the limits of integration change

$$\int_{-\infty}^{\varepsilon_V} \rightarrow \int_{-\infty}^{\varepsilon_{m,n}} \quad (10.8)$$

The expression becomes

$$p_V = \frac{1}{\pi} \sqrt{\frac{2m_h}{\hbar^2}} \int_{-\infty}^{\varepsilon_{m,n}} \left(1 - \frac{1}{1 + e^{\frac{\varepsilon - \varepsilon_F}{kT}}} \right) \frac{1}{\sqrt{\varepsilon_{m,n} - \varepsilon}} d\varepsilon$$

Furthermore, since $\varepsilon < \varepsilon_F$

$$p_V = \frac{1}{\pi} \sqrt{\frac{2m_h}{\hbar^2}} \int_{-\infty}^{\varepsilon_{m,n}} \left(1 - \frac{1}{1 + e^{-\frac{\varepsilon_F - \varepsilon}{kT}}} \right) \frac{1}{\sqrt{\varepsilon_{m,n} - \varepsilon}} d\varepsilon$$

Using the binomial expansion and keeping only the first two terms

$$\begin{aligned}\frac{1}{1+x} &= 1 - x + x^2 - x^3 + \dots \\ \frac{1}{1 + e^{-\frac{\varepsilon_F - \varepsilon}{kT}}} &= 1 - e^{-\frac{\varepsilon_F - \varepsilon}{kT}}\end{aligned}$$

the term in parenthesis becomes

$$\left(1 - \frac{1}{1 + e^{-\frac{\varepsilon_F - \varepsilon}{kT}}}\right) = e^{-\frac{\varepsilon_F - \varepsilon}{kT}}$$

yielding

$$\begin{aligned}p_V &= \frac{1}{\pi} \sqrt{\frac{2m_h}{\hbar^2}} \int_{-\infty}^{\varepsilon_{m,n}} e^{-\frac{\varepsilon_F - \varepsilon}{kT}} \frac{1}{\sqrt{\varepsilon_{m,n} - \varepsilon}} d\varepsilon \\ &= \frac{1}{\pi} \sqrt{\frac{2m_h}{\hbar^2}} \int_{-\infty}^{\varepsilon_{m,n}} e^{-\frac{\varepsilon_F - \varepsilon_{m,n} + \varepsilon_{m,n} - \varepsilon}{kT}} \frac{1}{\sqrt{\varepsilon_{m,n} - \varepsilon}} d\varepsilon \\ &= \frac{1}{\pi} \sqrt{\frac{2m_h}{\hbar^2}} \int_{-\infty}^{\varepsilon_{m,n}} e^{-\frac{\varepsilon_F - \varepsilon_{m,n}}{kT}} e^{-\frac{\varepsilon_{m,n} - \varepsilon}{kT}} \frac{1}{\sqrt{\varepsilon_{m,n} - \varepsilon}} d\varepsilon \\ &= \frac{1}{\pi} \sqrt{\frac{2m_h}{\hbar^2}} e^{-\frac{\varepsilon_F - \varepsilon_{m,n}}{kT}} \int_{-\infty}^{\varepsilon_{m,n}} e^{-\frac{\varepsilon_{m,n} - \varepsilon}{kT}} \frac{1}{\sqrt{\varepsilon_{m,n} - \varepsilon}} d\varepsilon\end{aligned}\tag{10.9}$$

To simplify, make a change of variables. Let $x = \frac{\varepsilon_{m,n} - \varepsilon}{kT}$ or $\varepsilon = \varepsilon_{m,n} - kTx$ and $d\varepsilon = -kTdx$. Note also the corresponding changes in the limits of integration.

$$\begin{aligned}p_V &= \frac{1}{\pi} \sqrt{\frac{2m_h}{\hbar^2}} e^{-\frac{\varepsilon_F - \varepsilon_{m,n}}{kT}} \int_{\infty}^0 e^{-x} \frac{1}{\sqrt{kT}\sqrt{x}} (-kT) dx \\ &= \frac{\sqrt{kT}}{\pi} \sqrt{\frac{2m_h}{\hbar^2}} e^{-\frac{\varepsilon_F - \varepsilon_{m,n}}{kT}} \int_0^{\infty} e^{-x} x^{-\frac{1}{2}} dx\end{aligned}$$

The resulting integral is the now familiar Gamma function. This lead to our final expression for the total concentration of holes in the valence band

$$\boxed{p_V = \frac{1}{\pi} \sqrt{\frac{2m_h kT}{\hbar^2}} e^{-\frac{\varepsilon_F - \varepsilon_{m,n}}{kT}} \Gamma\left(\frac{1}{2}\right)}\tag{10.10}$$

Fermi level position:1D

This evaluation goes the same was as for the 2D or 3D material. If we are dealing with intrinsic semiconductors with no additional doping

$$n_C = p_V$$

From what we have just evaluated

$$\frac{1}{\pi} \sqrt{\frac{2m_e kT}{\hbar^2}} \Gamma\left(\frac{1}{2}\right) e^{-\frac{\varepsilon_{m1,n1} - \varepsilon_F}{kT}} = \frac{1}{\pi} \sqrt{\frac{2m_h kT}{\hbar^2}} \Gamma\left(\frac{1}{2}\right) e^{-\frac{\varepsilon_F - \varepsilon_{m2,n2}}{kT}}$$

Here note that $\varepsilon_{m1,n1}$ and $\varepsilon_{m2,n2}$ are used to refer to the energies where the conduction and valence bands begin respectively. This expression reduces as follows

$$\begin{aligned} \sqrt{m_e} e^{-\frac{\varepsilon_{m1,n1} - \varepsilon_F}{kT}} &= \sqrt{m_h} e^{-\frac{\varepsilon_F - \varepsilon_{m2,n2}}{kT}} \\ e^{-\frac{(-\varepsilon_{m1,n1} + \varepsilon_F + \varepsilon_F - \varepsilon_{m2,n2})}{kT}} &= \sqrt{\frac{m_h}{m_e}} \\ e^{-\frac{\varepsilon_{m1,n1} + \varepsilon_{m2,n2}}{kT}} e^{\frac{2\varepsilon_F}{kT}} &= \sqrt{\frac{m_h}{m_e}} \\ e^{\frac{2\varepsilon_F}{kT}} &= \sqrt{\frac{m_h}{m_e}} e^{\frac{\varepsilon_{m1,n1} + \varepsilon_{m2,n2}}{kT}} \\ \ln\left(e^{\frac{2\varepsilon_F}{kT}}\right) &= \ln\left(\frac{m_h}{m_e}\right)^{\frac{1}{2}} + \ln\left(e^{\frac{\varepsilon_{m1,n1} + \varepsilon_{m2,n2}}{kT}}\right) \\ \frac{2\varepsilon_F}{kT} &= \ln\left(\frac{m_h}{m_e}\right)^{\frac{1}{2}} + \frac{\varepsilon_{m1,n1} + \varepsilon_{m2,n2}}{kT} \\ 2\varepsilon_F &= \frac{kT}{2} \ln\left(\frac{m_h}{m_e}\right) + \varepsilon_{m1,n1} + \varepsilon_{m2,n2} \end{aligned}$$

This leads to our final expression

$$\boxed{\varepsilon_F = \frac{\varepsilon_{m1,n1} + \varepsilon_{m2,n2}}{2} + \frac{kT}{4} \ln\left(\frac{m_h}{m_e}\right)} \quad (10.11)$$

Since $m_h > m_e$ the second term is positive and as a consequence the Fermi energy has a slight temperature dependence. However, as before, to a first approximation, it can be taken as the midpoint between the conduction band and valence band energies (the zero temperature limit).

Chapter 11

Joint density of states

In the previous sections we have calculated the “density of states” in both the conduction band and valence band separately. Interband transitions occur between both bands giving rise to optical transitions of the semiconductor. As a consequence it is instructive to calculate the “joint” density of states which is proportional to the absorption coefficient of the material.

3D bulk

We derive the density of states again as done previously. Consider a spherical volume of

$$V_k = \frac{4}{3}\pi k^3$$

The volume of a given mode was, recall $k_x k_y k_z$ where

$$\begin{aligned}k_x &= \frac{2\pi}{L_x} \\k_y &= \frac{2\pi}{L_y} \\k_z &= \frac{2\pi}{L_z}\end{aligned}$$

The number of modes or states in the given sphere is then

$$N = \frac{V_k}{k_x k_y k_z} = \frac{\frac{4}{3}\pi k^3}{8\pi^3} L_x L_y L_z$$

For an electron, multiply this by 2 to account for spin

$$\begin{aligned} N' &= 2N = 2 \frac{\frac{4}{3}\pi k^3}{8\pi^3} L_x L_y L_z \\ &= \frac{k^3}{3\pi^2} L_x L_y L_z \text{ total number of states in sphere} \end{aligned}$$

Consider as before, the density

$$\rho = \frac{N'}{L_x L_y L_z} = \frac{k^3}{3\pi^2} \text{ number of states/unit volume}$$

Now consider the energy density

$$\rho' = \frac{d\rho}{d\varepsilon} \text{ alternatively } \frac{d\rho}{dk}$$

Note that previously we solved for $\frac{d\rho}{d\varepsilon}$. This time however, let's solve for $\frac{d\rho}{dk}$.

$$\begin{aligned} \frac{d\rho}{d\varepsilon} &= \frac{1}{2\pi^2} \left(\frac{2m}{\hbar^2} \right)^{\frac{3}{2}} \sqrt{\varepsilon} \text{ (previously)} \\ \frac{d\rho}{dk} &= \frac{k^2}{\pi^2} \text{ (this time)} \end{aligned}$$

Starting with the energy density

$$\rho'_2 = \frac{d\rho}{dk} = \frac{k^2}{\pi^2} \quad (11.1)$$

divide by 2 to go back to only 1 spin orientation since in an optical transition spin flips are generally forbidden.

$$\rho'_1 = \frac{\rho'_2}{2} = \frac{k^2}{2\pi^2} \quad (11.2)$$

This expression applies to either conduction band or valence band. Applying the following equivalence

$$\rho_j(\varepsilon)d\varepsilon = \rho'_1(k)dk \quad (11.3)$$

we obtain

$$\begin{aligned} \rho_j(\varepsilon) &= \rho'_1(k) \frac{dk}{d\varepsilon} \\ &= \frac{k^2}{2\pi^2} \frac{dk}{d\varepsilon} \end{aligned} \quad (11.4)$$

where ρ_j is the desired joint density of states. Now from the conservation of momentum, transitions in k are vertical such that the initial k value in the valence band is the same k value as in the conduction band ($k_a = k_b = k$, where k_a is the k value in the valence band and k_b is the value in the conduction band).

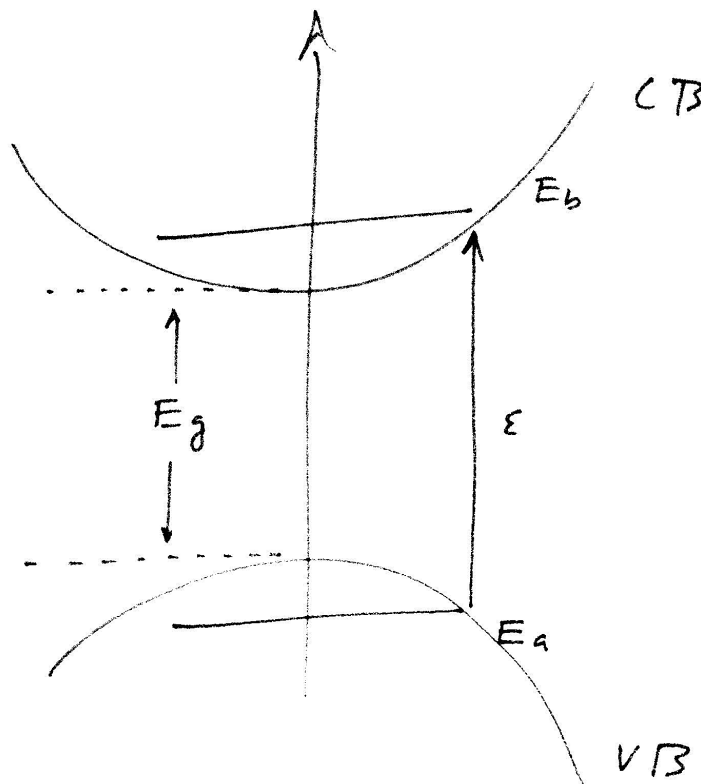


Figure 11.1: Cartoon showing momentum conserving vertical transition between the valence band and conduction band in k space

The energy of the initial state in the valence band is

$$\varepsilon_a = \varepsilon_v - \frac{\hbar^2 k_a^2}{2m_h}$$

Likewise the energy of the final state in the conduction band is

$$\varepsilon_b = \varepsilon_c + \frac{\hbar^2 k_b^2}{2m_e}$$

The energy of the transition is

$$\begin{aligned}
 \varepsilon &= \varepsilon_b - \varepsilon_a \\
 &= \left(\varepsilon_c + \frac{\hbar^2 k^2}{2m_e} \right) - \left(\varepsilon_v - \frac{\hbar^2 k^2}{2m_h} \right) \\
 &= (\varepsilon_c - \varepsilon_v) + \frac{\hbar^2 k^2}{2m_e} + \frac{\hbar^2 k^2}{2m_h} \\
 &= \varepsilon_g + \frac{\hbar^2 k^2}{2} \left(\frac{1}{m_e} + \frac{1}{m_h} \right)
 \end{aligned} \tag{11.5}$$

where $\varepsilon_g = \varepsilon_c - \varepsilon_v$. Now from above

$$\frac{d\varepsilon}{dk} = \hbar^2 k \frac{m_e + m_h}{m_e m_h}$$

leading to the desired expression

$$\frac{dk}{d\varepsilon} = \frac{1}{\hbar^2 k} \left(\frac{m_e m_h}{m_e + m_h} \right) \tag{11.6}$$

Since $\rho_j(\varepsilon) = \frac{k^2}{2\pi^2} \left(\frac{dk}{d\varepsilon} \right)$

$$\begin{aligned}
 \rho_j(\varepsilon) &= \frac{k^2}{2\pi^2} \left(\frac{1}{\hbar^2 k} \right) \left(\frac{m_e m_h}{m_e + m_h} \right) = \frac{k}{2\pi^2 \hbar^2} \left(\frac{m_e m_h}{m_e + m_h} \right) \\
 &= \frac{k\mu}{2\pi^2 \hbar^2}
 \end{aligned} \tag{11.7}$$

where for notational simplicity we have used the reduced mass $\mu = \frac{m_e m_h}{m_e + m_h}$. Now to continue towards our final expression, explicitly describe what k is

$$\begin{aligned}
 \varepsilon &= \varepsilon_g + \frac{\hbar^2 k^2}{2\mu} \\
 &\downarrow \\
 k^2 &= \frac{2\mu(\varepsilon - \varepsilon_g)}{\hbar^2}
 \end{aligned}$$

or

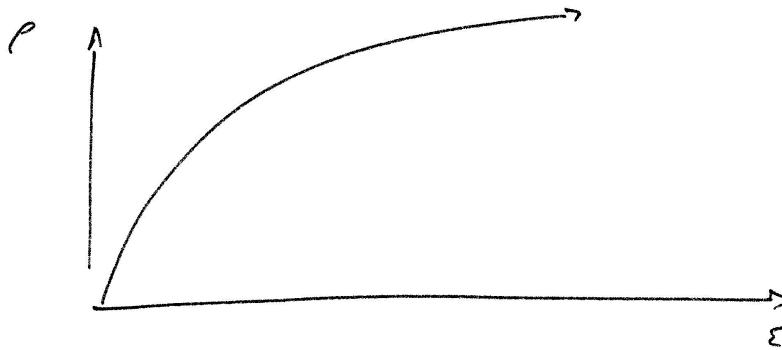
$$k = \frac{\sqrt{2\mu(\varepsilon - \varepsilon_g)}}{\hbar} \tag{11.8}$$

Replacing this into our main expression for ρ_j gives

$$\begin{aligned}
 \rho_j(\varepsilon) &= \frac{\mu}{2\pi^2 \hbar^2} k \\
 &= \left(\frac{\mu}{2\pi^2 \hbar^2} \right) \sqrt{\frac{2\mu(\varepsilon - \varepsilon_g)}{\hbar^2}}
 \end{aligned}$$

which after some simplification gives

$$\rho_j(\varepsilon) = \frac{1}{4\pi^2} \left(\frac{2\mu}{\hbar^2} \right)^{\frac{3}{2}} \sqrt{\varepsilon - \varepsilon_g} \quad (11.9)$$



$$\rho_{\text{joint}}(\varepsilon) = \frac{1}{4\pi^2} \left(\frac{2\mu}{\hbar^2} \right)^{\frac{3}{2}} \sqrt{\varepsilon - \varepsilon_g}$$

Figure 11.2: Joint density of states for a 3D material (bulk)

2D well

As we did before, consider an area in k -space of

$$A_k = \pi k^2$$

where the area occupied by a given mode or state is $k_y k_z$ (we implicitly assume that k_x represents the confined direction)

$$k_y = \frac{2\pi}{L_y}$$

$$k_z = \frac{2\pi}{L_z}$$

Together, the number of modes in the area is

$$N = \frac{A_k}{k_y k_z} = \frac{\pi k^2}{4\pi^2} L_y L_z = \frac{k^2}{4\pi} L_y L_z$$

Multiply by 2 to account for spin

$$N' = 2N = \frac{k^2}{2\pi} L_y L_z$$

Now consider the density

$$\rho = \frac{N'}{L_y L_z} = \frac{k^2}{2\pi} \quad (11.10)$$

with the energy density given by

$$\rho' = \frac{d\rho}{d\varepsilon} \text{ or alternatively } \frac{d\rho}{dk}$$

Note that previously we solved for $\frac{d\rho}{d\varepsilon}$. This time, however, let's consider $\frac{d\rho}{dk}$.

$$\begin{aligned} \frac{d\rho}{d\varepsilon} &= \frac{m}{\pi \hbar^2} \text{ (previously)} \\ \frac{d\rho}{dk} &= \frac{k}{\pi} \text{ (this time)} \end{aligned}$$

Starting with the energy density

$$\rho_2' = \frac{d\rho}{dk} = \frac{k}{\pi}$$

divide by 2 to get rid of the spin since formally speaking, spin flip optical transitions are forbidden.

$$\rho_1' = \frac{\rho_2'}{2} = \frac{k}{2\pi} \quad (11.11)$$

Now applying the following equivalence

$$\rho_j(\varepsilon) d\varepsilon = \rho_2'(k) dk \quad (11.12)$$

one obtains

$$\begin{aligned} \rho_j(\varepsilon) &= \rho_1'(k) \frac{dk}{d\varepsilon} \\ &= \frac{k}{2\pi} \frac{dk}{d\varepsilon} \end{aligned} \quad (11.13)$$

where ρ_j is the desired joint density of states. As before in the 3D case, the conservation of momentum means that transitions in k-space are “vertical”.

That is, the initial k value in the valence band is the same as the final k value in the conduction band ($k_a = k_b = k$) where $k_a(k_b)$ is the valence (conduction) band k values.

The energy of the initial state in the valence band is

$$\varepsilon_a = \varepsilon_{n2} - \frac{\hbar^2 k^2}{2m_h}$$

Likewise the energy of the final state in the conduction band is

$$\varepsilon_b = \varepsilon_{n1} + \frac{\hbar^2 k^2}{2m_e}$$

The transition energy is

$$\begin{aligned} \varepsilon &= \varepsilon_b - \varepsilon_a \\ &= \varepsilon_{n1} - \varepsilon_{n2} + \frac{\hbar^2 k^2}{2} \left(\frac{1}{m_e} + \frac{1}{m_h} \right) \\ &= \varepsilon_g + \frac{\hbar^2 k^2}{2\mu} \end{aligned} \quad (11.14)$$

where ε_g (effective band gap, not to be confused with the bulk band gap) $= \varepsilon_{n1} - \varepsilon_{n2}$ and μ is the reduced mass $\mu = \frac{m_e m_h}{m_e + m_h}$. This leads to

$$\frac{d\varepsilon}{dk} = \frac{\hbar^2 k}{\mu}$$

or

$$\frac{dk}{d\varepsilon} = \frac{\mu}{\hbar^2 k} \quad (11.15)$$

such that when replaced into our main expression the desired expression for the joint density of states is

$$\boxed{\rho_j(\varepsilon) = \frac{k}{2\pi} \left(\frac{\mu}{\hbar^2 k} \right) = \frac{\mu}{2\hbar^2 \pi}} \quad (11.16)$$

1D wire

Consider the length in k -space

$$L_k = 2k$$

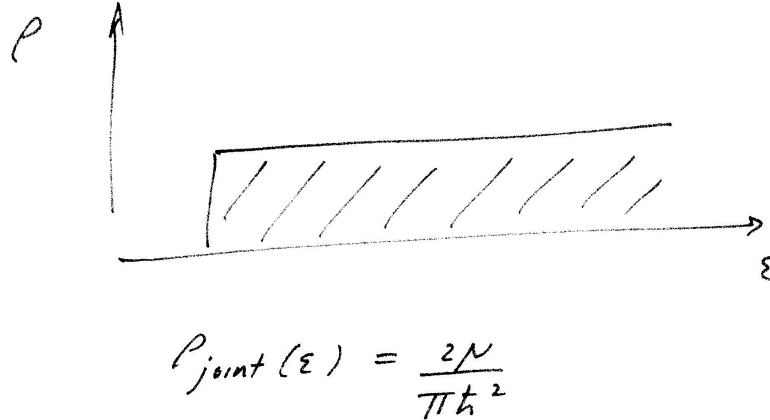


Figure 11.3: Joint density of states for a 2D material (quantum well)

The length occupied by a given mode or state is k_x where

$$k_x = \frac{2\pi}{L_x}$$

The number of states in the given length is

$$N = \frac{L_k}{k_x} = \frac{2k}{2\pi} L_x = \frac{kL_x}{\pi}$$

Multiply this by 2 to account for spin

$$N' = 2 \frac{kL_x}{\pi} \text{ total number of states}$$

Consider the density

$$\rho = \frac{N'}{L_x} = \frac{2k}{\pi} \text{ number of states per unit length}$$

Then the energy density is

$$\rho' = \frac{d\rho}{d\epsilon} \text{ or alternatively } \frac{d\rho}{dk}$$

Previously we solved for $\frac{d\rho}{d\epsilon}$

$$\frac{d\rho}{d\epsilon} = \frac{1}{\pi} \sqrt{\frac{2m}{\hbar^2}} \frac{1}{\sqrt{\epsilon}} \text{ (previously)}$$

$$\frac{d\rho}{dk} = \frac{2}{\pi} \text{ (this time)}$$

Starting with the energy density

$$\rho_2' = \frac{d\rho}{dk} = \frac{2}{\pi} \quad (11.17)$$

divide by 2 to consider only one spin orientation since spin flip transitions are generally forbidden.

$$\rho_1' = \frac{\rho_2'}{2} = \frac{1}{\pi}$$

Now apply the following equivalence

$$\rho_j(\varepsilon)d\varepsilon = \rho_1'(k)dk$$

leading to

$$\begin{aligned} \rho_j(\varepsilon) &= \rho_1'(k) \frac{dk}{d\varepsilon} \\ &= \frac{1}{\pi} \frac{dk}{d\varepsilon} \end{aligned} \quad (11.18)$$

where $\rho_j(\varepsilon)$ is the desired joint density of states. As in the other two cases, 3D and 2D, the conservation of momentum means that transitions in k-space are vertical so that $k_a = k_b = k$. Here k_a is the k value in the valence band and k_b is the k value in the conduction band.

The energy of the initial state in the valence band is

$$\varepsilon_a = \varepsilon_{m2,n2} - \frac{\hbar^2 k^2}{2m_h}$$

Likewise the energy of the final state in the conduction band is

$$\varepsilon_b = \varepsilon_{m1,n1} + \frac{\hbar^2 k^2}{2m_e}$$

The transition energy is

$$\begin{aligned} \varepsilon &= \varepsilon_b - \varepsilon_a \\ &= \varepsilon_{m1,n1} - \varepsilon_{m2,n2} + \frac{\hbar^2 k^2}{2} \left(\frac{1}{m_e} + \frac{1}{m_h} \right) \\ &= \varepsilon_g + \frac{\hbar^2 k^2}{2\mu} \end{aligned} \quad (11.19)$$

where $\varepsilon_g = \varepsilon_{m_1, n_1} - \varepsilon_{m_2, n_2}$ and the reduced mass $\mu = \frac{m_e m_h}{m_e + m_h}$. This leads to

$$\frac{d\varepsilon}{dk} = \frac{\hbar^2 k}{\mu}$$

or

$$\frac{dk}{d\varepsilon} = \frac{\mu}{\hbar^2 k} \quad (11.20)$$

Since $\rho_j(\varepsilon) = \frac{1}{\pi} \frac{dk}{d\varepsilon}$ the joint density becomes

$$\rho_j(\varepsilon) = \frac{\mu}{\pi \hbar^2} \frac{1}{k}$$

Now to continue towards our final expression we express k fully. Since $\varepsilon = \varepsilon_g + \frac{\hbar^2 k^2}{2\mu}$ we get

$$\begin{aligned} k^2 &= \frac{2\mu(\varepsilon - \varepsilon_g)}{\hbar^2} \\ &\downarrow \\ k &= \sqrt{\frac{2\mu}{\hbar^2}(\varepsilon - \varepsilon_g)} \\ &\downarrow \\ \frac{1}{k} &= \frac{\hbar}{\sqrt{2\mu}\sqrt{\varepsilon - \varepsilon_g}} \end{aligned}$$

This leads to the final expression for the joint density of states

$$\boxed{\rho_j(\varepsilon) = \frac{1}{\pi} \sqrt{\frac{\mu}{2\hbar^2}} \frac{1}{\sqrt{\varepsilon - \varepsilon_g}}} \quad (11.21)$$

or more directly analogous to last time

$$\boxed{= \frac{1}{2\pi} \sqrt{\frac{2\mu}{\hbar^2}} \frac{1}{\sqrt{\varepsilon - \varepsilon_g}}} \quad (11.22)$$

Note that the singularities are referred to as Van Hove singularities. In the case of real quantum wires, they are probably smeared out by Coulomb effects.

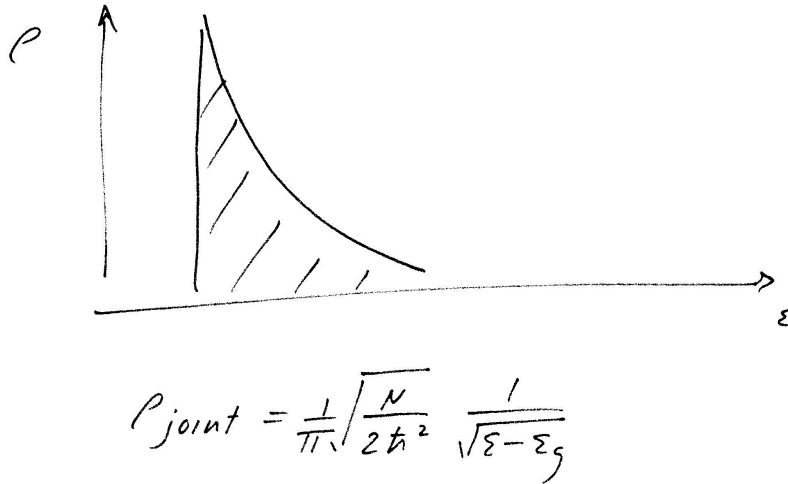


Figure 11.4: Joint density of states for a 1D material (quantum wire)

Summary

Now that we have the explicit joint density of states for 3D, 2D and 1D materials we can summarize their implications for the absorption spectra of these materials. It can be shown that the absorption coefficient and joint density of states are proportional to each other. The calculation to achieve this involves Fermi's golden rule and will not be discussed at this point. To summarize

- 3D: $\rho_j \propto \sqrt{\epsilon - \epsilon_g}$
- 2D: $\rho_j \propto$ constant depending on the reduced mass
- 1D: $\rho_j \propto \frac{1}{\sqrt{\epsilon - \epsilon_g}}$
- 0D: $\rho_j \propto \delta(\epsilon - \epsilon_n)$

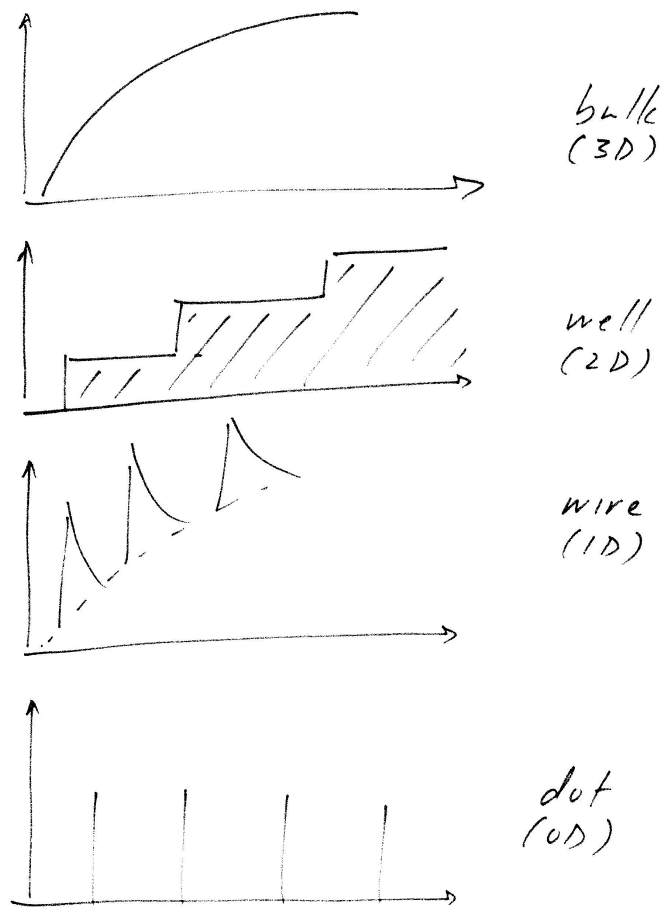


Figure 11.5: Summary of the joint density of states for 3D, 2D, 1D, and 0D materials

ov

Chapter 12

Absorption

A geometric derivation of exponential attenuation

Here we derive the standard exponential attenuation law of light being absorbed by a solid. Picture an incident beam of light with an initial intensity of I_o . Picture next a bulk solid which can be divided into very small slices or “slabs” of thickness Δx . The whole solid consists of n slabs placed back to back such that the total thickness of the solid is x .

Next, each slab absorbs some light and transmits the rest. In each slab a fraction $(1 - f)$ of the impinging light is transmitted. Here $0 \leq f \leq 1$. So initially we have I_o . After the first slab you have a residual intensity of $I = (1 - f)I_o$. After the second slab you have a residual intensity of $I = (1 - f)^2 I_o$ and so on.

The general expression for the residual intensity after n slabs is therefore

$$I = (1 - f)^n I_o$$

To derive the exponential attenuation law, re-arrange this into the following

$$\frac{I}{I_o} = (1 - f)^n$$

This is the fraction of the incident intensity that makes it through n slabs. Recall now that $x = n\Delta x$ giving

$$\frac{I}{I_o} = (1 - f)^{\frac{x}{\Delta x}}$$

Next, to simplify the notation let $p = (1 - f)$ yielding

$$\frac{I}{I_o} = p^{\frac{x}{\Delta x}}$$

where $0 < p < 1$. This can be rearranged as

$$\begin{aligned} \frac{I}{I_o} &= \left(p^{\frac{1}{\Delta x}} \right)^x \\ &= \left(\frac{1}{p} \right)^{-x} \end{aligned}$$

where note that $\frac{1}{p} > 1$. Now note that p is a constant. In addition, Δx is a constant. Therefore, call the whole thing some other constant, say y .

$$\frac{I}{I_o} = y^{-x}$$

where recall that y can be expressed alternatively as $y = e^{\ln y}$. This gives

$$\frac{I}{I_o} = \left(e^{\ln y} \right)^{-x}$$

Next, since y is a constant, so is $\ln y$. Call it α giving

$$\frac{I}{I_o} = e^{-\alpha x}$$

where $\alpha > 0$. This leads to the popular expression for the exponential attenuation of light in a solid.

$$\boxed{I = I_o e^{-\alpha x}} \tag{12.1}$$

where α is called the absorption coefficient.

Alternative derivation

There is an alternative derivation of the exponential attenuation law. The argument goes as follows:

(a) The change in intensity of the light is proportional to the thickness of the solid.

$$\Delta I \propto -\Delta x$$

The negative sign is to indicate attenuation of the light in the solid through absorption.

(b) Next, the absolute change in intensity on attenuation is also proportional to the intensity of the light.

$$\Delta I \propto I$$

(c) As a consequence, the change in intensity is also proportional to the product of both independent variables

$$\Delta I \propto -I\Delta x$$

Call the proportionality constant α giving

$$\begin{aligned}\Delta I &= -\alpha I\Delta x \\ \frac{\Delta I}{\Delta x} &= -\alpha I\end{aligned}$$

Next, if one lets ΔI and Δx become infinitesimally small one gets

$$\boxed{\frac{dI}{dx} = -\alpha I} \quad (12.2)$$

which is a first order differential equation that is easily solved.

$$\begin{aligned}\frac{dI}{I} &= -\alpha dx \\ \ln I &= \alpha x + \text{const}\end{aligned}$$

or

$$\boxed{I = I_0 e^{-\alpha x}} \quad (12.3)$$

This is our exponential attenuation law again.

Conventions

In dealing with the absorption of materials there are various conventions.

In bulk solids it is common to see

$$I = I_0 e^{-\alpha l} \quad (12.4)$$

where α is the “absorption coefficient” and is generally expressed in units of cm^{-1} when l is expressed in cm units.

In more molecular-like systems it is common to see

$$I = I_o e^{-\sigma n l} \quad (12.5)$$

where σ is the “cross section” having units of cm^2 and n is the concentration of absorbers in the sample in units of number/per cm^3 and l is the pathlength expressed in units of cm.

By comparison one can see that if one needs to relate α and σ

$$\alpha = \sigma n \quad (12.6)$$

but this shouldn't be necessary since in one case one has a bulk solid and in the other case the assumption is of a molecular-like material (quantum dots for example).

Finally, when dealing with molecular-like materials it is common to see the everyone's favorite Beer's law.

$$A = \epsilon c l \quad (12.7)$$

where A is the “absorbance” or “optical density (OD)”, ϵ is the molar extinction coefficient in units of ($M^{-1}cm^{-1}$), c is the concentration in units of (M) and finally l is the pathlength in units of cm.

Sometimes it is of interest to convert between extinction coefficient (ϵ) and the absorption cross section (σ). This can be achieved in the following fashion

$$A = \epsilon c l$$

where we also know that $A = \log(\frac{1}{T})$ and where $T = \frac{I}{I_o}$ is the transmittance of the sample. Thus

$$A = \log\left(\frac{I_o}{I}\right) = \epsilon c l$$

and from the exponential attenuation law we know that

$$\frac{I}{I_o} = e^{-\sigma n l}$$

Note that c is in units of M and n is in units of number/ cm^3 .

$$\begin{aligned} \log e^{\sigma n l} &= \epsilon c l \\ (\sigma n l) \log e &= \epsilon c l \end{aligned}$$

Drop the l terms on either side since they are both the same.

$$(\sigma n)l \log e = \epsilon c$$

Let's now convert c to the same units as n which is in terms of number of absorbers per cm^3 .

$$c(\text{mol/l}) \rightarrow \frac{cN_a}{dm^3}$$

where recall that moles times Avogadro's number is the number of absorbers and that $1l = dm^3$. Next convert from dm to cm by recalling that $1dm = 10cm$.

$$c(\text{mol/l}) \rightarrow \frac{cN_a}{1000cm^3} = n$$

therefore

$$c = \frac{n1000}{N_a}$$

This means that our previous equivalence becomes

$$(\sigma n)l \log e = \epsilon n \frac{1000}{N_a}$$

or

$$\sigma = \frac{\epsilon(1000)}{\log e(N_a)}$$

$$\boxed{\sigma = \frac{(2.303)(1000)\epsilon}{N_a}} \quad (12.8)$$

where the 2.303 comes from $\frac{1}{\log e}$.

Exercises

1. A solution with a dye concentration of $c = 4 \times 10^{-6}$ M has a molar absorptivity of $\epsilon = 1.5 \times 10^5 \frac{1}{M \cdot cm}$. It is illuminated with green laser light at $\lambda = 514.5$ nm and with a power of $P = 10nW$. (a) What is the percentage of light absorbed by the sample after 1 micron. (b) Calculate the number of photons per second absorbed by the sample. (c) Calculate the cross section of the molecule in units of cm^2 and \AA^2 .

2. *diIC*₁₂ is a popular dye in single molecule experiments. If one wants to image individual diI molecules one needs sufficiently low coverages on a microscope coverslip. Let's say no more than 10 molecules per square micron. (a) If 10 microliters of a diI solution spread evenly on a 4 mm diameter region achieves this coverage, what is the molar concentration of the diI solution. (b) This diI sample is then illuminated with 543.5 nm (a green HeNe). A uniform 100 nm diameter spot is illuminated with an excitation rate of 3×10^{10} photons per second. What is the incident excitation power, P? (c) Next, a single diI molecule being illuminated absorbs 2.3×10^5 p/s. Calculate the absorption cross section and the molar extinction coefficient of diI (a 1 cm cell may be assumed).
3. Colloidal CdSe QDs overcoated with ZnS are all the rage today. Look up a paper by XG Peng in Chemistry of Materials and from his data provide the range of absorption cross sections (σ) at the band edge for QD sizes between 2 nm and 6 nm in diameter. Express the same range in \AA^2 and then compare this to the physical size of the dot.

Chapter 13

Interband transitions

Interband absorption in semiconductors

Although we haven't discussed the concept of bands yet, there are so-called bands in semiconductors, namely the conduction band and the valence band. These bands are composed of accessible states to the electron and hole. We generally speak of electrons in the conduction band and holes in the valence band. The transition which promotes an electron from the valence band to the conduction band through the absorption of a photon is referred to as an "interband" transition. Alternatively, transitions within a given band are called "intra-band" transitions. Now in what follows we will assume that we are dealing with direct gap semiconductors.

The transition rate for absorption is given by Fermi's (second) Golden Rule. Note this is also sometimes abbreviated and expressed as the "Golden rule". Its basically a quantum mechanical way to derive a rate for a process and is analogous to a classical rate. The initial state (i) to final state (f) transition rate, $R_{i \rightarrow f}$, given by Fermi's Golden Rule is

$$\boxed{R_{i \rightarrow f} = \frac{2\pi}{\hbar} |M|^2 \rho_f(\epsilon)} \quad (13.1)$$

where $|M|$ is a matrix element linking the initial and final states and $\rho_f(\epsilon)$ is the joint density of states. Here is where your density of states calculations start to pay off. Notice that the transition rate is proportional to the joint density of states which you have just calculated for bulk, 0D, 1D and 2D materials.

Time dependent perturbation theory

We will work in what is referred to as the Schrodinger representation. Note for future reference that there are three picture of quantum mechanics, the Schrodinger representation, the Heisenberg representation and the interaction representation. Usually when doing time dependent perturbation theory, the interaction representation is preferred. But we will get the same expressions this way as well.

The general Schrodinger equation is actually written as

$$i\hbar \frac{\partial}{\partial t} \Psi(r, t) = H\Psi(r, t) \quad (13.2)$$

where $H = -\frac{\hbar^2 \nabla^2}{2m} + V$ and V is a potential perturbation which can be time dependent. Note that the wavefunctions depend on time now. We look for solutions of the form

$$\Psi(r, t) = \Psi(r)f(t) \quad (13.3)$$

To keep the notation more compact, reduce this to $\Psi(r, t) = \Psi f$. Replace this into the general time dependent Schrodinger equation.

$$\begin{aligned} i\hbar \frac{\partial}{\partial t}(\Psi f) &= \left(-\frac{\hbar^2 \nabla^2}{2m} + V \right) \Psi f \\ i\hbar \Psi \frac{\partial f}{\partial t} &= -\frac{\hbar^2 \nabla^2}{2m} \Psi f + V \Psi f \\ i\hbar \Psi f' &= -\frac{\hbar^2}{2m} f \Psi'' + V \Psi f \end{aligned}$$

Consolidate the f and Ψ terms giving

$$i\hbar \Psi \frac{f'}{f} = -\frac{\hbar^2}{2m} \Psi'' + V \Psi$$

or

$$\boxed{i\hbar \frac{f'}{f} = \frac{1}{\Psi} \left[-\frac{\hbar^2}{2m} \Psi'' + V \Psi \right]} \quad (13.4)$$

You will note that the left hand side is dependent only on time and that the right hand side is dependent only on position. To be equivalent both sides must equal a constant which we call ϵ . (turns out to be the eigenenergy or eigenvalue)

Left hand side

$$i\hbar \frac{f'}{f} = \epsilon$$

$$\frac{f'}{f} = -\frac{i\epsilon}{\hbar}$$

which we integrate to get

$$\ln f = -\frac{i\epsilon t}{\hbar}$$

The desired time dependent function is

$$\boxed{f(t) = e^{-\frac{i\epsilon t}{\hbar}}} \quad (13.5)$$

Right hand side

$$\frac{1}{\Psi} \left[-\frac{\hbar^2}{2m} \Psi'' + V\Psi \right] = \epsilon$$

$$-\frac{\hbar^2}{2m} \Psi'' + V\Psi = \epsilon\Psi$$

$$\left(-\frac{\hbar^2}{2m} \nabla^2 + V \right) \Psi = \epsilon\Psi$$

or

$$\boxed{H_0\Psi = \epsilon\Psi} \quad (13.6)$$

This is our usual eigenvalue and eigenvector problem. Solving this we get ϵ and $\Psi(r)$.

Putting everything together, the general form of the wavefunction is

$$\Psi(r, t) = \Psi(r)f(t)$$

or more explicitly

$$\boxed{\Psi(r, t) = \Psi(r)e^{-\frac{i\epsilon t}{\hbar}}} \quad (13.7)$$

Time dependent perturbation theory now

The basic idea of time-dependent perturbation theory is simple. First understand that if the perturbation (call it V) were absent, the eigenfunctions appropriate to the problem would be given by the equation

$$H^{(0)}\Psi^{(0)} = \epsilon^{(0)}\Psi^{(0)}$$

If the initial state is expanded in terms of these unperturbed eigenfunctions

$$\Psi(t=0) = \sum_n C_n e^{-\frac{i\epsilon_n^{(0)}t}{\hbar}} \Psi_n^{(0)}(r)$$

where C_n is a constant, then in the absence of a perturbation we would have for all times

$$\Psi(t > 0) = \sum_n C_n e^{-\frac{i\epsilon_n^{(0)}t}{\hbar}} \Psi_n^{(0)}(r)$$

where again C_n is a constant.

If the perturbation is present however, the above linear combination is no longer a valid solution to the Schrodinger equation.

But we can still express whatever true solution exists as a linear combination of the un-perturbed zero order wavefunctions provided that the coefficients are now time dependent.

$$\begin{aligned} \Psi(r, t) &= \sum_n C_n(t) e^{-\frac{i\epsilon_n^{(0)}t}{\hbar}} \Psi_n^{(0)}(r) \\ &= \sum_n C_n(t) \Psi_n(r, t) \end{aligned}$$

So let's go ahead and start deriving some results of the time dependent perturbation theory.

Plug our time dependent form of the wavefunction back into the general time dependent Schrodinger equation.

$$i\hbar \frac{d\Psi(r, t)}{dt} = H\Psi(r, t)$$

where $H = H^{(0)} + V(t)$ or $H^{(0)} + H^{(1)}(t)$. Use whatever notation that suits you. We then have

$$i\hbar \frac{d\Psi(r, t)}{dt} = \left(H^{(0)} + H^{(1)} \right) \Psi(r, t)$$

Ok since the notation gets wild, we're going to drop the (t) in $C_n(t)$. Just understand that the time dependence is implied. Furthermore, just work with $\Psi_n(r, t)$ instead of the full expression.

$$\begin{aligned} \left[H^{(0)} + H^{(1)} \right] \sum_n C_n \Psi_n(r, t) &= i\hbar \frac{d}{dt} \left[\sum_n C_n \Psi_n(r, t) \right] \\ H^{(0)} \sum_n C_n \Psi_n(r, t) + H^{(1)} \sum_n C_n \Psi_n(r, t) &= i\hbar \frac{d}{dt} \left[\sum_n C_n \Psi_n(r, t) \right] \end{aligned}$$

Expand the right hand side using the Chain rule.

$$\begin{aligned} &H^{(0)} \sum_n C_n \Psi_n(r, t) + H^{(1)} \sum_n C_n \Psi_n(r, t) \\ &= i\hbar \left[\sum_n C_n \frac{d\Psi_n(r, t)}{dt} + \sum_n \frac{dC_n}{dt} \Psi_n(r, t) \right] \\ &H^{(0)} \sum_n C_n \Psi_n(r, t) + H^{(1)} \sum_n C_n \Psi_n(r, t) \\ &= i\hbar \sum_n C_n \frac{d\Psi_n(r, t)}{dt} + i\hbar \sum_n \frac{dC_n}{dt} \Psi_n(r, t) \end{aligned}$$

The first terms on the left and right side of the last expression will cancel. To illustrate let $n = 1$ for example.

$$\begin{aligned} H^{(0)} C_1 \Psi_1(r, t) &= i\hbar C_1 \frac{d\Psi_1(r, t)}{dt} \\ H^{(0)} C_1 \Psi_1(r) e^{-\frac{i\epsilon_1^{(0)} t}{\hbar}} &= i\hbar C_1 \Psi_1(r) e^{-\frac{i\epsilon_1^{(0)} t}{\hbar}} \left(-\frac{i\epsilon_1^{(0)}}{\hbar} \right) \end{aligned}$$

cancel common terms

$$H^{(0)} \Psi_1(r) = \epsilon_1^{(0)} \Psi_1(r)$$

which is our favorite eigenvalue and eigenvector relation. The same applies for all n so that we are left with

$$\boxed{H^{(1)} \sum_n C_n \Psi_n(r, t) = i\hbar \sum_n \frac{dC_n}{dt} \Psi_n(r, t)} \quad (13.8)$$

Now we desire to solve this expression for the time dependent coefficients of some state say k .

Multiply both sides by $\Psi_k^{(0)*}(r)$ and recall that all states $n \neq k$ of the same basis are orthogonal.

$$\begin{aligned} \Psi_k^{(0)*}(r) H^{(1)} \sum_n C_n \Psi_n(r, t) &= i\hbar \sum_n \frac{dC_n}{dt} \Psi_k^{(0)*}(r) \Psi_n(r, t) \\ \sum_n C_n \Psi_k^{(0)*}(r) H^{(1)} \Psi_n(r, t) &= i\hbar \sum_n \frac{dC_n}{dt} \Psi_k^{(0)*}(r) \Psi_n(r, t) \\ \sum_n C_n \Psi_k^{(0)*}(r) H^{(1)} \Psi_n^{(0)}(r) e^{-\frac{i\epsilon_n^{(0)}t}{\hbar}} &= i\hbar \sum_n \frac{dC_n}{dt} \Psi_k^{(0)*}(r) \Psi_n^{(0)}(r) e^{-\frac{i\epsilon_n^{(0)}t}{\hbar}} \\ &\text{where } n = k \text{ on right side} \\ \sum_n C_n \langle k | H^{(1)} | n \rangle e^{-\frac{i\epsilon_n^{(0)}t}{\hbar}} &= i\hbar \frac{dC_n}{dt} e^{-\frac{i\epsilon_k^{(0)}t}{\hbar}} \end{aligned}$$

Multiply both sides by $e^{\frac{i\epsilon_k^{(0)}t}{\hbar}}$ to simplify.

$$\begin{aligned} \sum_n C_n \langle k | H^{(1)} | n \rangle e^{-\frac{i\epsilon_n^{(0)}t}{\hbar}} e^{\frac{i\epsilon_k^{(0)}t}{\hbar}} &= i\hbar \frac{dC_k}{dt} \\ \sum_n C_n \langle k | H^{(1)} | n \rangle e^{\frac{i(\epsilon_k^{(0)} - \epsilon_n^{(0)})t}{\hbar}} &= i\hbar \frac{dC_k}{dt} \end{aligned}$$

Let $\omega_{kn} = \frac{\epsilon_k^{(0)} - \epsilon_n^{(0)}}{\hbar}$ since $\Delta\epsilon = \hbar\omega$ yielding

$$\sum_n C_n \langle k | H^{(1)} | n \rangle e^{i\omega_{kn}t} = i\hbar \frac{dC_k}{dt}$$

or better yet

$$\boxed{i\hbar \frac{dC_k}{dt} = \sum_n C_n \langle k | H^{(1)} | n \rangle e^{i\omega_{kn}t}} \quad (13.9)$$

This is basically a system of simultaneous linear differential equations.

In matrix notation

$$i\hbar \frac{d}{dt} \begin{pmatrix} C_1 \\ C_2 \\ C_3 \\ \dots \\ C_k \end{pmatrix} = \begin{pmatrix} V_{11} & V_{12}e^{i\omega_{12}t} & \dots \\ V_{21}e^{-i\omega_{12}t} & V_{22} & \dots \\ \dots & \dots & \dots \\ \dots & \dots & \dots \\ \dots & \dots & \dots \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \\ C_3 \\ \dots \\ C_k \end{pmatrix}$$

Now what? This expression is pretty tough to solve. So to overcome this problem you apply time dependent perturbation theory (which is basically making some assumptions about the size of the perturbation and in turn what C_1, C_2, C_3, \dots are).

To illustrate, suppose the system starts out in state 1. Then

$$\begin{aligned} C_1(t=0) &= 1 \\ C_2(t=0) &= 0 \\ C_3(t=0) &= 0 \end{aligned}$$

and so forth.

If there are no perturbations, then the system stays in state 1 forever.

Our zeroth order approximation is therefore

$$\begin{aligned} C_1(t) &= 1 \\ C_2(t) &= 0 \\ C_3(t) &= 0 \\ &\dots \\ C_n(t) &= 0 \end{aligned}$$

To get our first order approximation insert these zeroth order values into our matrix yielding

$$\begin{aligned} i\hbar \frac{dC_k}{dt} &= C_{n=1} \langle k|H^{(1)}|n=1 \rangle e^{i\omega_{kn=1}t} \\ \frac{dC_k}{dt} &= \frac{1}{i\hbar} C_{n=1} \langle k|H^{(1)}|n=1 \rangle e^{i\omega_{kn=1}t} \\ &\text{where } c_{n=1} \approx 1 \end{aligned}$$

yielding

$$\boxed{C_k(t) = \frac{1}{i\hbar} \int_0^t \langle k|H^{(1)}|n=1 \rangle e^{i\omega_{kn=1}t'} dt'} \quad (13.10)$$

Note however that when integrating $\frac{dC_k}{dt}$ there is actually a constant of integration that comes out.

$$C_k(t) = const + \frac{1}{i\hbar} \int_0^t \langle k|H^{(1)}|n \rangle e^{i\omega_{kn}t'} dt'$$

where $const = C_k(t=0) = \delta_{kn}$. Replacing this into the above expression gives

$$C_k(t) = C_k(0) + \frac{1}{i\hbar} \int_0^t \langle k|H^{(1)}|n \rangle e^{i\omega_{kn}t'} dt'$$

or your standard textbook expression

$$\boxed{C_k(t) = \delta_{kn} + \frac{1}{i\hbar} \int_0^t \langle k|H^{(1)}|n \rangle e^{i\omega_{kn}t'} dt'} \quad (13.11)$$

Summary

$$\boxed{C_k^{(0)}(t) = \delta_{kn}} \quad (13.12)$$

$$\boxed{C_k^{(1)}(t) = \frac{1}{i\hbar} \int_0^t \langle k|H^{(1)}|n \rangle e^{i\omega_{kn}t'} dt'} \quad (13.13)$$

Two level system, sinusoidal perturbation, Schrodinger representation

Let's illustrate our results with a more concrete example and for simplicity keep it to two levels (a two level system, TLS).

Recall that our time dependent Schrodinger equation is

$$\begin{aligned} i\hbar \frac{d}{dt} \Psi(r, t) &= H \Psi(r, t) \\ &\text{where } H = H^{(0)} + H^{(1)} \\ i\hbar \frac{d}{dt} \Psi(r, t) &= (H^{(0)} + H^{(1)}) \Psi(r, t) \end{aligned}$$

Now recall that a general solution to this time dependent Schrodinger equation can be expressed as a linear combination of the unperturbed zero order wavefunctions, provided time dependent coefficients.

$$\boxed{\Psi(r, t) = C_1(t) \Psi_1(r, t) + C_2(t) \Psi_2(r, t)}$$

where

$$\begin{aligned} \Psi_1(r, t) &= \Psi_1(r) e^{-\frac{i\epsilon_1^{(0)} t}{\hbar}} \\ \Psi_2(r, t) &= \Psi_2(r) e^{-\frac{i\epsilon_2^{(0)} t}{\hbar}} \end{aligned}$$

Now let's be more specific with what $H^{(1)}(t)$ looks like. We want to consider the interaction of a molecule with dipole moment μ with light where the electric field of light can be written as

$$\epsilon = E_0 \cos 2\pi \nu t$$

where ν is the frequency of the radiation. Alternatively

$$\boxed{\epsilon = E_0 \cos \omega t} \quad (13.14)$$

This is going to be our sinusoidal perturbation. You can see that it will oscillate periodically with time.

The interaction between a molecule's dipole moment and the electric field is then

$$H^{(1)} = -\mu \cdot \epsilon$$

$$\boxed{H^{(1)} = -\mu \cdot E_0 \cos \omega t} \quad (13.15)$$

Such that

$$\boxed{H = H^{(0)} - \mu \cdot E_0 \cos \omega t} \quad (13.16)$$

We are now going to re-derive our previous expression for the coefficients. Insert $\Psi(r, t)$ into the Schrodinger equation. Also for shorthand let the time dependence of the coefficients be understood $C_1(t) \rightarrow C_1$ and $C_2(t) \rightarrow C_2$.

$$\begin{aligned} & i\hbar \frac{d}{dt} [C_1 \Psi_1(r, t) + C_2 \Psi_2(r, t)] \\ = & (H^{(0)} + H^{(1)}) [C_1 \psi_1(r, t) + C_2 \Psi_2(r, t)] \\ & i\hbar \frac{d}{dt} (C_1 \Psi_1(r, t)) + i\hbar \frac{d}{dt} (C_2 \Psi_2(r, t)) \\ = & (H^{(0)} + H^{(1)}) [C_1 \Psi(r, t) + C_2 \Psi_2(r, t)] \\ & i\hbar C_1 \frac{d\Psi_1(r, t)}{dt} + i\hbar \Psi_1(r, t) \frac{dC_1}{dt} + i\hbar C_2 \frac{d\Psi_2(r, t)}{dt} + i\hbar \Psi_2(r, t) \frac{dC_2}{dt} \\ = & H^{(0)} (C_1 \Psi_1(r, t) + C_2 \Psi_2(r, t)) + H^{(1)} (C_1 \Psi_1(r, t) + C_2 \Psi_2(r, t)) \\ & C_1 i\hbar \frac{d\Psi_1(r, t)}{dt} + C_2 i\hbar \frac{d\Psi_2(r, t)}{dt} + i\hbar \Psi_1(r, t) \frac{dC_1}{dt} + i\hbar \Psi_2(r, t) \frac{dC_2}{dt} \\ = & C_1 H^{(0)} \Psi_1(r, t) + C_2 H^{(0)} \Psi_2(r, t) + C_1 H^{(1)} \Psi_1(r, t) + C_2 H^{(1)} \Psi_2(r, t) \end{aligned}$$

We've seen previously that the first terms on either side of the last equation will cancel since they are equivalent.

$$\begin{aligned} H^{(0)} \Psi_1(r, t) &= i\hbar \frac{d\Psi_1(r, t)}{dt} \\ H^{(0)} \Psi_2(r, t) &= i\hbar \frac{d\Psi_2(r, t)}{dt} \end{aligned}$$

leaving

$$\boxed{i\hbar\Psi_1(r,t)\frac{dC_1}{dt} + i\hbar\Psi_2(r,t)\frac{dC_2}{dt} = C_1H^{(1)}\Psi_1(r,t) + C_2H^{(1)}\Psi_2(r,t)} \quad (13.17)$$

Now like before, multiply both sides by $\Psi_2^{(0)*}(r)$ and integrate.

$$\begin{aligned} & i\hbar\Psi_2^{(0)*}(r)\Psi_1(r,t)\frac{dC_1}{dt} + i\hbar\Psi_2^{(0)*}(r)\Psi_2(r,t)\frac{dC_2}{dt} \\ &= C_1\Psi_2^{(0)*}(r)H^{(1)}\Psi_1(r,t) + C_2\Psi_2^{(0)*}(r)H^{(1)}\Psi_2(r,t) \\ & i\hbar\langle 2|1\rangle e^{-\frac{i\epsilon_1^{(0)}t}{\hbar}}\frac{dC_1}{dt} + i\hbar\langle 2|2\rangle e^{-\frac{i\epsilon_2^{(0)}t}{\hbar}}\frac{dC_2}{dt} \\ &= C_1\langle 2|H^{(1)}|1\rangle e^{-\frac{i\epsilon_1^{(0)}t}{\hbar}} + C_2\langle 2|H^{(1)}|2\rangle e^{-\frac{i\epsilon_2^{(0)}t}{\hbar}} \\ i\hbar e^{-\frac{i\epsilon_2^{(0)}t}{\hbar}}\frac{dC_2}{dt} &= C_1\langle 2|H^{(1)}|1\rangle e^{-\frac{i\epsilon_1^{(0)}t}{\hbar}} + C_2\langle 2|H^{(1)}|2\rangle e^{-\frac{i\epsilon_2^{(0)}t}{\hbar}} \\ & \text{multiply both sides by } e^{\frac{i\epsilon_2^{(0)}t}{\hbar}} \\ i\hbar\frac{dC_2}{dt} &= C_1\langle 2|H^{(1)}|1\rangle e^{\frac{i(\epsilon_2^{(0)}-\epsilon_1^{(0)})t}{\hbar}} + C_2\langle 2|H^{(1)}|2\rangle \end{aligned}$$

as before let $\omega_{21} = \frac{\epsilon_2^{(0)} - \epsilon_1^{(0)}}{\hbar}$. The exact expression we get then is

$$\boxed{i\hbar\frac{dC_2}{dt} = C_1\langle 2|H^{(1)}|1\rangle e^{i\omega_{21}t} + C_2\langle 2|H^{(1)}|2\rangle} \quad (13.18)$$

Now we invoke time dependent perturbation theory and assume that since $H^{(1)}$ is small

$$\begin{aligned} C_1(t) &\simeq 1 \\ C_2(t) &\simeq 0 \end{aligned}$$

giving

$$\boxed{i\hbar\frac{dC_2}{dt} = \langle 2|H^{(1)}|1\rangle e^{i\omega_{21}t}} \quad (13.19)$$

What now? Well, we haven't really dealt with $H^{(1)}$ yet. Recall that

$$H^{(1)} = -\mu \cdot E_o \cos\omega t$$

For convenience, choose the z direction

$$H_z^{(1)} = -\mu E_o \cos \omega t$$

or

$$\boxed{H_z^{(1)} = -\mu_z E_o \left(\frac{e^{i\omega t} + e^{-i\omega t}}{2} \right)} \quad (13.20)$$

Ok, let's throw this into our previous expression and see what happens.

$$\begin{aligned} i\hbar \frac{dC_2}{dt} &= \langle 2 | -\frac{\mu_z E_o}{2} (e^{i\omega t} + e^{-i\omega t}) | 1 \rangle e^{i\omega_{21} t} \\ i\hbar \frac{dC_2}{dt} &= -\frac{E_o}{2} \langle 2 | \mu_z (e^{i\omega t} + e^{-i\omega t}) | 1 \rangle e^{i\omega_{21} t} \\ i\hbar \frac{dC_2}{dt} &= -\frac{E_o}{2} \langle 2 | \mu_z | 1 \rangle (e^{i\omega t + i\omega_{21} t} + e^{-i\omega t + i\omega_{21} t}) \end{aligned} \quad (13.21)$$

The term $\langle 2 | \mu_z | 1 \rangle$ is the transition dipole moment.

$$\begin{aligned} i\hbar \frac{dC_2}{dt} &= -\frac{E_o}{2} \langle 2 | \mu_z | 1 \rangle (e^{i(\omega + \omega_{21})t} + e^{-i(\omega - \omega_{21})t}) \\ \frac{dC_2}{dt} &= \frac{iE_o}{2\hbar} \langle 2 | \mu_z | 1 \rangle (e^{i(\omega + \omega_{21})t} + e^{-i(\omega - \omega_{21})t}) \end{aligned}$$

Integrate this to find $C_2(t)$ but before we do this, let's flip back to ϵ notation so that our final expression will be consistent with everyone else.

$$\frac{dC_2}{dt} = \frac{iE_o}{2\hbar} \langle 2 | \mu_z | 1 \rangle (e^{i\frac{\hbar\omega + \hbar\omega_{21}}{\hbar}t} + e^{-i\frac{\hbar\omega - \hbar\omega_{21}}{\hbar}t})$$

where $\hbar\omega = h\nu$ and $\hbar\omega_{21} = \hbar\omega_2 - \hbar\omega_1 = \epsilon_2 - \epsilon_1$

$$\frac{dC_2}{dt} = \frac{iE_o}{2\hbar} \langle 2 | \mu_z | 1 \rangle (e^{i\frac{h\nu + (\epsilon_2 - \epsilon_1)}{\hbar}t} + e^{-i\frac{h\nu - \epsilon_2 + \epsilon_1}{\hbar}t})$$

$$\boxed{\frac{dC_2}{dt} = \frac{iE_o}{2\hbar} \langle 2 | \mu_z | 1 \rangle (e^{i\frac{(\epsilon_2 - \epsilon_1 + h\nu)}{\hbar}t} + e^{i\frac{(\epsilon_2 - \epsilon_1 - h\nu)}{\hbar}t})} \quad (13.22)$$

Now integrate this to get $C_2(t)$. Remember that when you integrate $\frac{dC_2}{dt}$ you get an additional constant of integration which is $C_2(0)$.

$$C_2(t) = C_2(0) + \frac{iE_o \langle 2 | \mu_z | 1 \rangle}{2\hbar} \int_0^t e^{\frac{i(\epsilon_2 - \epsilon_1 + h\nu)t'}{\hbar}} + e^{\frac{i(\epsilon_2 - \epsilon_1 - h\nu)t'}{\hbar}} dt'$$

where $C_2(0) = \delta_{12}$

$$C_2(t) = \delta_{12} + \frac{iE_o \langle 2 | \mu_z | 1 \rangle}{2\hbar} \int_0^t e^{\frac{i(\epsilon_2 - \epsilon_1 + h\nu)t'}{\hbar}} + e^{\frac{i(\epsilon_2 - \epsilon_1 - h\nu)t'}{\hbar}} dt'$$

but since states $1 \neq 2$ we are left in practice with the 1st order contribution. So let's just work with it.

$$\boxed{C_2^{(1)}(t) = \frac{iE_o \langle 2|\mu_z|1 \rangle}{2\hbar} \int_0^t e^{\frac{i(\epsilon_2 - \epsilon_1 + h\nu)t'}{\hbar}} + e^{\frac{i(\epsilon_2 - \epsilon_1 - h\nu)t'}{\hbar}} dt'} \quad (13.23)$$

So go on and integrate this expression.

$$C_2^{(1)}(t) = \frac{E_o \langle 2|\mu_z|1 \rangle}{2} \left[\frac{(e^{\frac{i(\epsilon_2 - \epsilon_1 + h\nu)t}{\hbar}} - 1)\hbar}{i(\epsilon_2 - \epsilon_1 + h\nu)} + \frac{(e^{\frac{i(\epsilon_2 - \epsilon_1 - h\nu)t}{\hbar}} - 1)\hbar}{i(\epsilon_2 - \epsilon_1 - h\nu)} \right]$$

or

$$\boxed{C_2^{(1)}(t) = \frac{E_o \langle 2|\mu_z|1 \rangle}{2} \left[\frac{(e^{\frac{i(\epsilon_2 - \epsilon_1 + h\nu)t}{\hbar}} - 1)}{(\epsilon_2 - \epsilon_1 + h\nu)} + \frac{(e^{\frac{i(\epsilon_2 - \epsilon_1 - h\nu)t}{\hbar}} - 1)}{(\epsilon_2 - \epsilon_1 - h\nu)} \right]} \quad (13.24)$$

Now at this point, another approximation is made here.

You have two terms. The one with $(\epsilon_2 - \epsilon_1 - h\nu)$ in the denominator is called the resonant term. The one with $(\epsilon_2 - \epsilon_1 + h\nu)$ is called the non-resonant term.

When $h\nu = (\epsilon_2 - \epsilon_1)$ the resonant term will dominate. Under this resonance condition where the monochromatic light energy equals the energy difference between states 1 and 2 you get

$$\boxed{C_2^{(1)}(t) \simeq \frac{E_o \langle 2|\mu_z|1 \rangle}{2} \left[\frac{e^{\frac{i(\epsilon_2 - \epsilon_1 - h\nu)t}{\hbar}} - 1}{(\epsilon_2 - \epsilon_1 - h\nu)} \right]} \quad (13.25)$$

Note that this approximation is also called the "Rotating Wave Approximation".

Since we like math, let's keep simplifying. Extract out $e^{\frac{i(\epsilon_2 - \epsilon_1 - h\nu)t}{2\hbar}}$

$$C_2^{(1)}(t) \simeq \frac{E_o \langle 2|\mu_z|1 \rangle}{2} e^{\frac{i(\epsilon_2 - \epsilon_1 - h\nu)t}{2\hbar}} \left(\frac{e^{\frac{i(\epsilon_2 - \epsilon_1 - h\nu)t}{2\hbar}} - e^{\frac{-i(\epsilon_2 - \epsilon_1 - h\nu)t}{2\hbar}}}{\epsilon_2 - \epsilon_1 - h\nu} \right)$$

$$C_2^{(1)}(t) \simeq \frac{iE_o \langle 2|\mu_z|1 \rangle}{(\epsilon_2 - \epsilon_1 - h\nu)} e^{\frac{i(\epsilon_2 - \epsilon_1 - h\nu)t}{2\hbar}} \left(\frac{e^{\frac{i(\epsilon_2 - \epsilon_1 - h\nu)t}{2\hbar}} - e^{\frac{-i(\epsilon_2 - \epsilon_1 - h\nu)t}{2\hbar}}}{2i} \right)$$

resulting in

$$\boxed{C_2^{(1)}(t) \simeq \frac{iE_o \langle 2|\mu_z|1 \rangle}{(\epsilon_2 - \epsilon_1 - h\nu)} e^{\frac{i(\epsilon_2 - \epsilon_1 - h\nu)t}{2\hbar}} \sin\left(\frac{\epsilon_2 - \epsilon_1 - h\nu}{2\hbar} t\right)} \quad (13.26)$$

Ok, let's keep going. Recall that the probability of being in a particular state k is basically

$$P_k = |C_k|^2$$

Therefore the transition probability into the final state 2 is

$$\begin{aligned} P_2 &= |C_2(t)|^2 \\ &= \frac{E_o^2 |\langle 2|\mu_z|1 \rangle|^2}{(\epsilon_2 - \epsilon_1 - h\nu)^2} \sin^2 \left(\frac{\epsilon_2 - \epsilon_1 - h\nu}{2\hbar} t \right) \end{aligned}$$

The final transition probability into state 2 is

$$P_2 = E_o^2 |\langle 2|\mu_z|1 \rangle|^2 \left(\frac{\sin^2 \left(\frac{\epsilon_2 - \epsilon_1 - h\nu}{2\hbar} t \right)}{(\epsilon_2 - \epsilon_1 - h\nu)^2} \right) \quad (13.27)$$

Note that P_2 is $P_2(h\nu, t)$. It depends on both time and frequency.

In general express this as

$$P_k = \frac{E_o^2 |\langle k|\mu_z|n \rangle|^2}{\hbar^2} \frac{\sin^2 \frac{\omega_{kn} - \omega}{2} t}{(\omega_{kn} - \omega)^2} \quad (13.28)$$

since $\hbar\omega = h\nu$.

Fermi's Golden Rule

Recall that we are dealing with a harmonic perturbation of the form

$$H^{(1)} = -\mu_z E_o \cos \omega t \quad (13.29)$$

$$H^{(1)} = -\mu_z E_o \sin \omega t \quad (13.30)$$

Note that it doesn't matter which one we use, both work fine.

From our previous work we derived

$$P_k = \frac{E_o^2 |\langle k|\mu_z|n \rangle|^2}{\hbar^2} \frac{\sin^2 \frac{\omega_{kn} - \omega}{2} t}{(\omega_{kn} - \omega)^2}$$

where again $\omega_{kn} = (\omega_k - \omega_n)$.

Like above, we want to consider the transition probability and more specifically the rate of transition into a group of final states k (not just into one state).

So we integrate over the spread of final states dictated by the density $\rho(\epsilon_k)$ or $\rho(\omega_k)$.

$$\boxed{P_{k,group} = \int_{-\infty}^{\infty} P_k(\omega_k, t) \rho(\omega_k) d\omega_k} \quad (13.31)$$

$$\begin{aligned} P_{k,group} &= \int_{-\infty}^{\infty} \frac{E_o |\langle k | \mu_z | n \rangle|^2}{\hbar^2} \frac{\sin^2 \frac{\omega_{kn} - \omega}{2}}{(\omega_{kn} - \omega)^2} \rho(\omega_k) d\omega_k \\ &= \int_{-\infty}^{\infty} \frac{E_o |\langle k | \mu_z | n \rangle|^2}{\hbar^2} \frac{\sin^2 \frac{\omega_k - \omega_n - \omega}{2}}{(\omega_k - \omega_n - \omega)^2} \rho(\omega_k) d\omega_k \end{aligned}$$

Now as before, assume that $\rho(\omega_k)$ is constant or rather that the distribution is uniform. It can be removed from the integral.

$$P_{k,group} = \frac{E_o^2 \rho(\omega_k) |\langle k | \mu_z | n \rangle|^2}{\hbar^2} \int_{-\infty}^{\infty} \frac{\sin^2 \frac{\omega_k - \omega_n - \omega}{2} t}{(\omega_k - \omega_n - \omega)^2} d\omega_k$$

Let $x = (\omega_k - \omega_n - \omega)$ and $dx = d\omega_k$. Also remember to change the limits of integration. We then get

$$P_{k,group} = \frac{E_o^2 \rho(\omega_k) |\langle k | \mu_z | n \rangle|^2}{\hbar^2} \int_{-\infty}^{\infty} \frac{\sin^2 \frac{xt}{2}}{x^2} dx$$

As before, from a table of integrals the definite integral is

$$\int_{-\infty}^{\infty} \frac{\sin^2 \frac{xt}{2}}{x^2} dx = \frac{\pi t}{2}$$

The resulting transition probability is then

$$P_{k,group} = \frac{E_o^2 |\langle k | \mu_z | n \rangle|^2}{\hbar^2} \frac{\pi t}{2}$$

Also like before, rather than stop here, flip $\rho(\omega_k)$ into $\rho(\epsilon_k)$ to keep the notation consistent with other texts ($\rho(\omega_k) = \hbar \rho(\epsilon_k)$).

$$\boxed{P_{k,group} = \frac{\pi E_o^2 \rho(\epsilon_k)}{2\hbar} |\langle k | \mu_z | n \rangle|^2 t} \quad (13.32)$$

This is the desired form of the transition probability.

Now the desired rate is $R = \frac{dP_{k,group}}{dt}$.

$$\boxed{R = \frac{\pi}{2\hbar} E_o^2 \rho(\epsilon_k) |\langle k | \mu_z | n \rangle|^2} \quad (13.33)$$

Again, the rate is constant with respect to time. It is the quantum mechanical version of a classical rate. This is called Fermi's Golden Rule.

Putting it all together Interband transitions (or a whole lot of work)

There are various ways to proceed with this. We carry out the following derivations in a solid state physics approach to stay consistent with existing texts.

For the interaction of light and matter we have seen the Hamiltonian

$$H^{(1)} = -\mu \cdot E$$

One can also express this interaction through the following Hamiltonian

$$H^{(1)} = -\frac{q}{m}(A \cdot P)$$

where A is the vector potential associated with the magnetic field of the electromagnetic wave and P is the quantum mechanical operator.

Proceed using the second Hamiltonian. It can be shown that

$$A = A_o \cos(\omega t - kr)$$

where we ignore a potential phase factor, ϕ , and $A = A_o \hat{e}$ where the second A_o is a scalar amplitude and \hat{e} is a unit vector showing the orientation of the electric field polarization.

Then

$$\begin{aligned} H^{(1)} &= -\frac{q}{m} A_o \hat{e} \cos(\omega t - kr) P \\ &= -\frac{q}{m} A_o \cos(\omega t - kr) (\hat{e} \cdot P) \\ &= -\frac{q A_o}{2m} \left[e^{i(\omega t - kr)} + e^{-i(\omega t - kr)} \right] (\hat{e} \cdot P) \end{aligned}$$

Next it can be shown that generally the transition probability can be expressed as (this is best done through the interaction representation)

$$P_k = \frac{1}{\hbar^2} \left| \int_{t_o}^t dt' e^{i\omega_{kn}t'} \langle k | H^{(1)} | n \rangle \right|^2 \quad (13.34)$$

where k is the final state (the conduction band in our case) and n is the initial state (the valence band in our state).

We will shove into this transition probability the above expression for $H^{(1)}$ to get

$$\begin{aligned} P_k &= \frac{1}{\hbar^2} \left| \int_{t_0}^t dt' e^{i\omega_{kn}t'} \langle k | \left(e^{i(\omega t' - kr)} + e^{-i(\omega t' - kr)} \right) (\hat{e} \cdot P) | n \rangle \right|^2 \\ &= \frac{1}{\hbar^2} \left(\frac{q^2 A_0^2}{4m^2} \right) \left| \int_{t_0}^t dt' e^{i\omega_{kn}t'} \langle k | \left(e^{i(\omega t' - kr)} + e^{-i(\omega t' - kr)} \right) (\hat{e} \cdot P) | n \rangle \right|^2 \end{aligned}$$

Next there are two terms sandwiched by $\langle k |$ and $| n \rangle$. These are the terms $e^{i(\omega t' - kr)}$ and $e^{-i(\omega t' - kr)}$. The one on the left will eventually lead to a “non-resonant” term related to stimulated emission. The second term on the right will lead to the absorption term that we are interested in. This is the “resonant” term. So to simplify life just drop the first term. We now have

$$P_k \simeq \frac{1}{\hbar^2} \left(\frac{q^2 A_0^2}{4m^2} \right) \left| \int_{t_0}^t dt' e^{i\omega_{kn}t'} \langle k | \left(e^{-i(\omega t' - kr)} \right) (\hat{e} \cdot P) | n \rangle \right|^2$$

Pull out the time dependent term from the bra-ket since this is only a spatial integral.

$$\begin{aligned} P_k &\simeq \frac{1}{\hbar^2} \left(\frac{q^2 A_0^2}{4m^2} \right) \left| \int_{t_0}^t dt' e^{i\omega_{kn}t' - i\omega t'} \langle k | e^{ikr} (\hat{e} \cdot P) | n \rangle \right|^2 \\ &\simeq \frac{1}{\hbar^2} \left(\frac{q^2 A_0^2}{4m^2} \right) \left| \int_{t_0}^t dt' e^{i(\omega_{kn} - \omega)t'} \langle k | e^{ikr} (\hat{e} \cdot P) | n \rangle \right|^2 \end{aligned}$$

Next we have the Bloch forms of the wavefunctions

$$\begin{aligned} |k\rangle &= \frac{1}{\sqrt{N}} u_k(r) e^{ikr} \\ |n\rangle &= \frac{1}{\sqrt{N}} u_n(r) e^{ik_n r} \end{aligned}$$

where $\frac{1}{\sqrt{N}}$ is a normalization constant reflecting normalization of both wavefunctions over the entire volume of the solid being considered. N is the number of unit cells in the crystal considered.

$$P_k \simeq \frac{1}{\hbar^2} \left(\frac{q^2 A_0^2}{4m^2} \right) \left| \int_{t_0}^t dt' e^{i(\omega_{kn} - \omega)t'} \frac{1}{N} \int_V u_k e^{-ik_k r} e^{ikr} (\hat{e} \cdot P) u_n e^{ik_n r} d^3r \right|^2$$

where the interior integral is over the entire volume of the solid considered.

$$P_j \simeq \frac{1}{\hbar^2} \left(\frac{q^2 A_0^2}{4m^2} \right) \left(\frac{1}{N^2} \right) \left| \int_{t_0}^t dt' e^{i(\omega_{kn} - \omega)t'} \int_V u_k e^{-ik_k r} e^{ikr} (\hat{e} \cdot P) u_n e^{ik_n r} d^3r \right|^2$$

Now in principle, P operates on both u_n and $e^{ik_n r}$ but since we understand that $e^{ik_n r}$ varies very slowly, we (as an approximation) ignore P operating on $e^{ik_n r}$.

$$P_k \simeq \frac{1}{\hbar^2} \left(\frac{q^2 A_o^2}{4m^2} \right) \left(\frac{1}{N^2} \right) \left| \int_{t_o}^t dt' e^{i(\omega_{kn} - \omega)t'} \int_V e^{i(k+k_n - k_r)r} u_k(\hat{e} \cdot P) u_n d^3 r \right|^2$$

Next $e^{i(k+k_n - k_r)r}$ is a slowly varying term that can be approximated by a constant in a given unit cell. We approximate this now as

$$P_k \simeq \frac{1}{\hbar^2} \left(\frac{q^2 A_o^2}{4m^2} \right) \left(\frac{1}{N^2} \right) \left| \int_{t_o}^t dt' e^{i(\omega_{kn} - \omega)t'} \sum_N e^{i(k+k_n - k_r)r} \int_V u_k(\hat{e} \cdot P) u_n d^3 r \right|^2$$

We take what is sometimes called the long wavelength approximation $e^{ikr} \simeq 1$ and get

$$P_k \simeq \frac{1}{\hbar^2} \left(\frac{q^2 A_o^2}{4m^2} \right) \left(\frac{1}{N^2} \right) \left| dt' e^{i(\omega_{kn} - \omega)t'} N \int_V u_k(\hat{e} \cdot P) u_n d^3 r \right|^2$$

Next cancel common terms and choose a projection for P , say $P_x(\hat{e} \cdot P) = P_x$ giving

$$P_k \simeq \frac{1}{\hbar^2} \left(\frac{q^2 A_o^2}{4m^2} \right) \left| \int_{t_o}^t dt' e^{i(\omega_{kn} - \omega)t'} \langle k | P_x | n \rangle \right|^2$$

Extract the spatial integral since it does not depend on time.

$$P_k \simeq \frac{1}{\hbar^2} \left(\frac{q^2 A_o^2}{4m^2} \right) | \langle k | P_x | n \rangle |^2 \left| \int_{t_o}^t dt' e^{i(\omega_{kn} - \omega)t'} \right|^2$$

where

$$\int_{t_o}^t dt' e^{i(\omega_{kn} - \omega)t'} = \frac{e^{i(\omega_{kn} - \omega)t'} \Big|_{t_o}^t}{i(\omega_{kn} - \omega)}$$

Let $t_o = 0$ to get

$$\frac{e^{i(\omega_{kn} - \omega)t} - 1}{i(\omega_{kn} - \omega)}$$

Re-express this as

$$\frac{e^{\frac{i(\omega_{kn} - \omega)t}{2}} \left(e^{\frac{i(\omega_{kn} - \omega)t}{2}} - e^{-\frac{i(\omega_{kn} - \omega)t}{2}} \right)}{i(\omega_{kn} - \omega)}$$

or as

$$2e^{\frac{i(\omega_{kn}-\omega)t}{2}} \frac{\sin\left(\frac{(\omega_{kn}-\omega)t}{2}\right)}{(\omega_{kn}-\omega)}$$

Put it all together now to get

$$P_k \simeq \frac{1}{\hbar^2} \left(\frac{q^2 A_o^2}{4m_o^2} \right) | \langle k | P_x | n \rangle |^2 \left| 2e^{\frac{i(\omega_{kn}-\omega)t}{2}} \frac{\sin\left(\frac{(\omega_{kn}-\omega)t}{2}\right)}{(\omega_{kn}-\omega)} \right|^2$$

This simplifies to

$$\boxed{P_k \simeq \frac{1}{\hbar^2} \left(\frac{q^2 A_o^2}{m^2} \right) | \langle k | P_x | n \rangle |^2 \frac{\sin^2\left(\frac{(\omega_{kn}-\omega)t}{2}\right)}{(\omega_{kn}-\omega)^2}} \quad (13.35)$$

Now to account for the available states we eventually integrate over the joint density of states. However, first convert the transition probability over from frequency to energy using $\epsilon = \hbar\omega$

$$P_k(\epsilon) \simeq \frac{1}{\hbar^2} \left(\frac{q^2 A_o^2}{m^2} \right) | \langle k | P_x | n \rangle |^2 \frac{\sin^2\left(\frac{(\epsilon_{kn}-\epsilon)t}{2\hbar}\right)}{(\epsilon_{kn}-\epsilon)^2} \hbar^2$$

giving

$$P_k(\epsilon) \simeq \left(\frac{q^2 A_o^2}{m^2} \right) | \langle k | P_x | n \rangle |^2 \frac{\sin^2\left(\frac{(\epsilon_{kn}-\epsilon)t}{2\hbar}\right)}{(\epsilon_{kn}-\epsilon)^2} \quad (13.36)$$

Next at long enough times ($t \rightarrow \infty$), we can simplify the last term above into a delta function using

$$\lim_{t \rightarrow \infty} \frac{\sin^2 at}{\pi a^2 t} = \delta(a)$$

So that

$$\left(\frac{\pi t}{4\hbar^2} \right) \left[\frac{\sin^2\left(\frac{(\epsilon_{kn}-\epsilon)t}{2\hbar}\right)}{\pi \left(\frac{(\epsilon_{kn}-\epsilon)}{2\hbar}\right)^2 t} \right] \rightarrow \frac{\pi t}{4\hbar^2} \delta\left(\frac{\epsilon_{kn}-\epsilon}{2\hbar}\right) \quad (13.37)$$

giving

$$P_k(\epsilon) \simeq \left(\frac{q^2 A_o^2}{m^2} \right) \left(\frac{\pi t}{4\hbar^2} \right) | \langle k | P_x | n \rangle |^2 \delta\left(\frac{\epsilon_{kn}-\epsilon}{2\hbar}\right)$$

where $\delta\left(\frac{\epsilon_{kn}-\epsilon}{2\hbar}\right) = 2\hbar\delta(\epsilon_{kn} - \epsilon)$.

The desired transition probability is

$$\boxed{P_k \simeq \frac{2\pi}{\hbar} \left(\frac{q^2 A_o^2}{4m^2}\right) |\langle k|P_x|n \rangle|^2 \delta(\epsilon_{kn} - \epsilon)t} \quad (13.38)$$

The transition rate, R , is the time derivative of the probability, $R = \frac{dP_k}{dt}$ giving

$$\boxed{R = \frac{2\pi}{\hbar} \left(\frac{q^2 A_o^2}{4m^2}\right) |\langle k|P_x|n \rangle|^2 \delta(\epsilon_{kn} - \epsilon)} \quad (13.39)$$

Finally, we integrate the rate expression over the joint density of states

$$\boxed{R_{tot} = \frac{2\pi}{\hbar} \left(\frac{q^2 A_o^2}{4m^2}\right) |\langle k|P_x|n \rangle|^2 \int \rho_j(\epsilon) \delta(\epsilon_{kn} - \epsilon) d\epsilon} \quad (13.40)$$

In the 3D bulk case we have

$$\rho_{j,3D}(\epsilon) = \frac{1}{4\pi^2} \left(\frac{2\mu}{\hbar^2}\right)^{\frac{3}{2}} \sqrt{\epsilon - \epsilon_g}$$

If we account for the degeneracy of the spin, multiply the above expression by two

$$\rho_{j,3D}(\epsilon) = \frac{1}{2\pi^2} \left(\frac{2\mu}{\hbar^2}\right)^{\frac{3}{2}} \sqrt{\epsilon - \epsilon_g}$$

Shove this into the above expression for the total transition rate (per unit volume) to get

$$\begin{aligned} R_{tot} &\simeq \frac{2\pi}{\hbar} \left(\frac{q^2 A_o^2}{4m_o^2}\right) |\langle k|P_x|n \rangle|^2 \int \frac{1}{2\pi^2} \left(\frac{2\mu}{\hbar^2}\right)^{\frac{3}{2}} \sqrt{\epsilon - \epsilon_g} \delta(\epsilon_{kn} - \epsilon) d\epsilon \\ &\simeq \frac{2\pi}{\hbar^2} \left(\frac{q^2 A_o^2}{3m_o^2}\right) |\langle k|P_x|n \rangle|^2 \left(\frac{1}{2\pi^2}\right) \left(\frac{2\mu}{\hbar^2}\right)^{\frac{3}{2}} \int \sqrt{\epsilon - \epsilon_g} \delta(\epsilon_{kn} - \epsilon) d\epsilon \end{aligned}$$

The final expression we have been waiting for is

$$\boxed{R_{tot} = \frac{2\pi}{\hbar} \left(\frac{q^2 A_o^2}{4m_o^2}\right) |\langle k|P_x|n \rangle|^2 \left[\left(\frac{1}{2\pi^2}\right) \left(\frac{2\mu}{\hbar^2}\right)^{\frac{3}{2}} \sqrt{\epsilon - \epsilon_g} \right]} \quad (13.41)$$

This is the total transition rate per unit volume of a bulk 3D material including spin degeneracy.

The absorption coefficient

Now let's close the loop and connect this total rate to the absorption coefficient.

The number of photons per unit area per unit time is

$$n_p = \frac{I}{\hbar\omega} \quad (13.42)$$

and the absorption coefficient can be expressed as

$$\alpha = \frac{R_{tot}}{n_p} = \frac{\hbar\omega R_{tot}}{I} \quad (13.43)$$

Next we need an alternative expression for the incident intensity, I .

$$I = \frac{n_r c \epsilon_o \omega^2 A_o^2}{2} = \frac{n_r c \epsilon_o E^2}{2} \quad (13.44)$$

This was obtained from

$$I = \left\langle \frac{1}{\mu_o} E \times B \right\rangle$$

but in many cases (nonmagnetic materials) $\mu \simeq \mu_o$ giving

$$I = \left\langle \frac{1}{\mu} E \times B \right\rangle$$

where $B = \frac{n}{c} E$. This leads to

$$\begin{aligned} I &= \left\langle \frac{1}{\mu} \left(\frac{n}{c} \right) E^2 \right\rangle \\ &= \frac{1}{\mu} \left(\frac{n}{c} \right) \frac{1}{2} E_o^2 \end{aligned}$$

and since $c^2 = \frac{1}{\epsilon_o \mu_o}$ we have $c = \frac{1}{c \epsilon_o \mu_o}$ or $c \epsilon_o = \frac{1}{c \mu_o} \simeq \frac{1}{c \mu}$ and finally $c \mu = \frac{1}{c \epsilon_o}$. Replace the last expression into our main equation to get

$$\boxed{I = \frac{n c \epsilon_o \omega^2 A_o^2}{2}} \quad (13.45)$$

Then using this alternative expression for I , replace it into the expression for α above.

$$\alpha = \frac{2 \hbar \omega}{n c \epsilon_o \omega^2 A_o^2} R_{tot}$$

After simplifying this you get

$$\alpha = \left(\frac{q^2}{2\pi m_o^2 c \epsilon_o} \right) \left(\frac{1}{n\omega} \right) \left(\frac{2\mu}{\hbar^2} \right)^{\frac{3}{2}} | \langle k | P_x | n \rangle |^2 \sqrt{\epsilon - \epsilon_g} \quad (13.46)$$

Often there is a change of notation. Let

$$\boxed{E_p = \frac{2 | \langle k | P_x | n \rangle |^2}{m_o}} \quad (13.47)$$

which has a value of approximately 20 eV for many semiconductors. The final expression is

$$\boxed{\alpha_{3D}(\epsilon) = \left(\frac{q^2 \hbar}{4\pi m_o c \epsilon_o n} \right) \left(\frac{E_p}{\hbar\omega} \right) \left(\frac{2\mu}{\hbar^2} \right)^{\frac{3}{2}} \sqrt{\epsilon - \epsilon_g}} \quad (13.48)$$

The absorption spectrum therefore is shown to be proportional to the joint density of states. This is the 3D case.

Chapter 14

Emission

In the last section, the absorption coefficient of materials with different dimensionality were shown to be proportional to the calculated joint density of states. In turn, one could predict that the absorption spectrum of these materials would look like. Here we work out the background behind spontaneous emission, a complementary process to absorption.

Preliminaries: Einstein A and B coefficients

Picture a two level system like that shown in the figure.

Here

- N_1 =population in the ground state or alternatively probability of being in the ground state
- N_2 =population in the excited state, alternatively probability of being in the excited state
- g_1 =degeneracy of ground state
- g_2 =degeneracy of excited state
- ρ =the energy density containing thermal and/or external contributions

Three processes were considered

1. absorption with total rate: $B_{12}\rho N_1$
2. stimulated emission with total rate: $B_{21}\rho N_2$

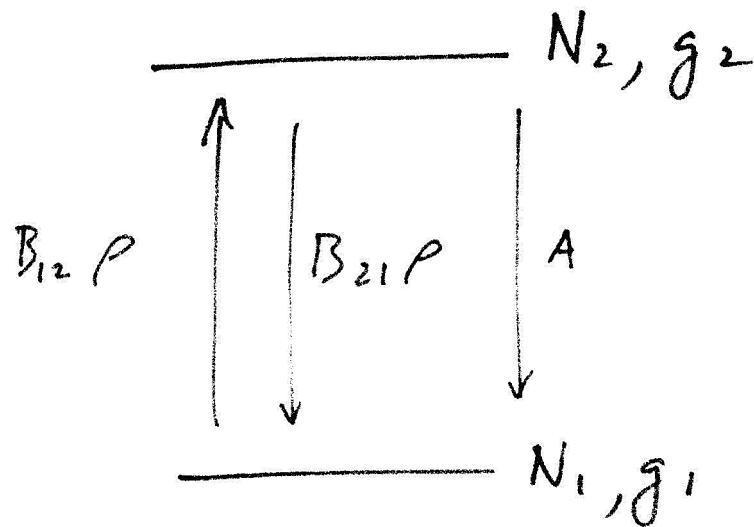


Figure 14.1: Two level system showing various transitions considered for Einstein A and B coefficients

3. spontaneous emission with total rate: AN_2

Putting everything together one obtains the following rate equations for populating either the ground or excited states

$$\begin{aligned}\frac{dN_1}{dt} &= -B_{12}\rho N_1 + B_{21}\rho N_2 + AN_2 \\ \frac{dN_2}{dt} &= B_{12}\rho N_1 - B_{21}\rho N_2 - AN_2\end{aligned}$$

In equilibrium, the upward and downward rates are equivalent ($\frac{dN_1}{dt} = \frac{dN_2}{dt} = 0$). So using either of the above expressions

$$B_{12}\rho N_1 - B_{21}\rho N_2 - AN_2 = 0$$

or

$$B_{12}\rho N_1 \text{ (upward)} = B_{21}\rho N_2 + AN_2 \text{ (downward)}$$

solving for ρ gives

$$\begin{aligned}\rho &= \frac{AN_2}{B_{12}N_1 - B_{21}N_2} \\ &= \frac{A}{B_{12}\frac{N_1}{N_2} - B_{21}}\end{aligned}$$

where N_1 and N_2 are Boltzman distributed

$$\begin{aligned}N_1 &= \frac{1}{N} \sum_{i=1}^{g_1} e^{-\frac{\varepsilon_1 - \varepsilon_F}{kT}} = \frac{g_1}{N} e^{-\frac{\varepsilon_1 - \varepsilon_0}{kT}} \\ N_2 &= \frac{1}{N} \sum_{i=1}^{g_2} e^{-\frac{\varepsilon_2 - \varepsilon_F}{kT}} = \frac{g_2}{N} e^{-\frac{\varepsilon_2 - \varepsilon_0}{kT}}\end{aligned}$$

and N is a normalization constant. This leads to

$$\frac{N_1}{N_2} = \frac{g_1}{g_2} e^{-\frac{\varepsilon_1 - \varepsilon_2}{kT}} = \frac{g_1}{g_2} e^{\frac{h\nu}{kT}}$$

where $h\nu = \varepsilon_2 - \varepsilon_1$ is the energy of the photon or transition. Replace this into the above expression for ρ giving

$$\rho(\nu) = \frac{A}{B_{12}\frac{g_1}{g_2} e^{\frac{h\nu}{kT}} - B_{21}} \quad (14.1)$$

This is the expression for the energy density derived by Einstein. Next he realized that this expression had to be equivalent to the Planck distribution for blackbody radiation. Expressed in terms of wavelength, the Planck distribution is (standard textbook expression)

$$\rho_p(\lambda) = \frac{8\pi hc}{\lambda^5} \left(\frac{1}{e^{\frac{hc}{\lambda kT}} - 1} \right) \quad (14.2)$$

Units: J/(unit volume unit wavelength). This expression is interesting from a historical point of view because when the derivative of this expression is set to zero one obtains the Wein displacement law for blackbody radiation. Likewise the integral of this expression gives what's called the Stefan Boltzman law for blackbody radiation. In a sense, knowing or unknowingly, Planck basically explained everything in one shot. Apparently he nearly suffered a nervous breakdown doing it.

Now back to the main discussion. Expressed in terms of ν rather than λ , the Planck energy density is (derived in the next section)

$$\rho_p(\nu) = \frac{8\pi h\nu^3}{c^3} \left(\frac{1}{e^{\frac{h\nu}{kT}} - 1} \right) \quad (14.3)$$

Units: J/(unit volume unit frequency). Note that in general, in the absence of vacuum, the index of refraction must be taken into account when considering the speed of light.

Interlude: Derivation of Planck energy density

Just like in the density of states section, imagine a sphere of radius k with volume

$$V_k = \frac{4}{3}\pi k^3$$

where the volume of a given mode is $k_x k_y k_z$ (assume periodic boundary conditions, though not necessary, just for convenience) and

$$\begin{aligned} k_x &= \frac{2\pi}{L_x} \\ k_y &= \frac{2\pi}{L_y} \\ k_z &= \frac{2\pi}{L_z} \end{aligned}$$

The number of modes in the given sphere is then (look familiar?)

$$N = \frac{V_k}{k_x k_y k_z} = \frac{\frac{4}{3}\pi k^3}{8\pi^3} L_x L_y L_z$$

For a photon now, multiply this by 2 to account for two possible polarizations

$$\begin{aligned} N' &= 2N = 2 \frac{\frac{4}{3}\pi k^3}{8\pi^3} L_x L_y L_z \\ &= \frac{k^3}{3\pi^2} L_x L_y L_z \text{ total number of modes} \end{aligned}$$

Consider the volume density

$$\rho = \frac{N'}{L_x L_y L_z} = \frac{k^3}{3\pi^2} = \frac{k^3}{3\pi^2} \text{ number of modes/unit volume} \quad (14.4)$$

Here is where we diverge a little from the past. To get the frequency density, first let $k = \frac{2\pi}{\lambda}$ where $\lambda = \frac{c}{\nu}$. Therefore

$$k = \frac{2\pi\nu}{c}$$

alternatively $2\pi\nu\frac{n}{c}$ if the index of refraction is different from 1. Then ρ becomes

$$\rho = \frac{k^3}{3\pi^2} = \frac{8\pi}{3} \frac{\nu^3}{c^3}$$

Now like before, to get the frequency density (alternatively called the mode density)

$$\rho' = \frac{d\rho}{d\nu} = \frac{8\pi\nu^2}{c^3} \text{ number per unit volume per unit frequency} \quad (14.5)$$

Planck next showed that the average energy per mode is

$$\langle \varepsilon \rangle = \frac{h\nu}{e^{\frac{h\nu}{kT}} - 1}$$

Note the Bose Einstein form of the equation (photons are bosons). This leads to the Planck distribution

$$\begin{aligned} \rho_p(\nu) &= \rho' \langle \varepsilon \rangle \\ &= \frac{8\pi\nu^2}{c^3} \left(\frac{h\nu}{e^{\frac{h\nu}{kT}} - 1} \right) \\ &= \frac{8\pi h\nu^3}{c^3} \left(\frac{1}{e^{\frac{h\nu}{kT}} - 1} \right) \end{aligned}$$

Alternative derivation

Consider the volume of a thin shell in k space. The volume is

$$V = 4\pi k^2 dk$$

where k_x, k_y, k_z all equal $\frac{n\pi}{L_{x,y,z}}$. The number of modes, N is then

$$\begin{aligned} N &= \frac{4\pi k^2 dk}{k_x k_y k_z} \\ &= \frac{4\pi k^2 dk}{\pi^3} L_x L_y L_z \end{aligned}$$

Since there are two polarizations, multiply by 2.

$$N = \frac{8\pi k^2 dk}{\pi^3} L_x L_y L_z$$

Now divide by 8 since each corner of an imaginary cube is shared by 8 other neighboring cubes.

$$N = \frac{k^2}{\pi^2} dk L_x L_y L_z$$

Now get the volume density by dividing by $L_x L_y L_z$ giving

$$\rho = \frac{N'}{L_x L_y L_z} = \frac{k^2 dk}{\pi^2}$$

where $k = \frac{2\pi}{\lambda} = \frac{2\pi\nu}{c}$ such that $k^2 = \frac{4\pi^2\nu^2}{c^2}$ and also $dk = \frac{2\pi}{c} d\nu$. We get

$$\begin{aligned} \rho &= \frac{4\pi\nu^2}{c^3} (2\pi) d\nu \\ &= \frac{8\pi\nu^2}{c^3} d\nu \end{aligned}$$

Now consider the frequency density

$$\rho' d\nu = \rho$$

such that

$$\rho' = \frac{8\pi\nu^2}{c^3}$$

Now back to where we left off. Equate the Einstein and Planck relations

$$\frac{A}{B_{12} \frac{g_1}{g_2} e^{\frac{h\nu}{kT}} - B_{21}} = \frac{8\pi h\nu^3}{c^3} \left(\frac{1}{e^{\frac{h\nu}{kT}} - 1} \right)$$

where note again that if the index is not 1 then replace c with $\frac{c}{n}$. Upon examination of the above equation, to achieve equivalence

$$B_{12} \frac{g_1}{g_2} = B_{21} \quad (14.6)$$

More usually you will see $g_1 = g_2$ such that the expression reduces to $B_{12} = B_{21}$. Furthermore

$$\frac{A}{B_{21}} = \frac{8\pi h\nu^3}{c^3}$$

yielding the standard textbook expression

$$\begin{aligned} A &= \frac{8\pi h\nu^3}{c^3} B_{21} \\ &= \frac{8\pi h\nu^3}{c^3} \frac{g_1}{g_2} B_{12} \end{aligned} \quad (14.7)$$

Again, note that if $g_1 = g_2$ you will commonly see written

$$\boxed{A = \frac{8\pi h\nu^3}{c^3} B_{12}} \quad (14.8)$$

Also if the index is not 1 (not vacuum) then replace c with $\frac{c}{n}$ in the above expressions. The above final expression are what are referred to as the relationship between Einstein A and B coefficients. Remember that A is the one associated with spontaneous emission.

Word of caution

The Planck distribution is often written a number of ways. Different texts will have what on the surface appear to be completely different expressions. These difference actually arises because of differing definitions for the “density” being used by the various authors. Previously we saw two expression for the Planck distribution. The first, in terms of wavelength, has units of: number per unit volume per unit wavelength. The second, in terms of frequency, has units of: number per unit volume per unit frequency. Alternatively sometimes what people mean is the same expression but in units

of: number per unit volume per unit energy. This is potentially very confusing. One should look very carefully at what is meant by “density”. Here we derive this third form of the Planck distribution with units: number per unit volume per unit energy.

Starting with

$$\rho = \frac{8\pi\nu^3}{3} \left(\frac{n}{c}\right)^3 \quad (14.9)$$

where this expression comes from the previous section where we derived the Planck distribution and where also the index of the medium has been explicitly considered (recall, replace c with $\frac{c}{n}$). Rearrange the expression to

$$\rho = \frac{8\pi n^3 \nu^3}{3c^3} \left(\frac{h^3}{h^3}\right) = \frac{8\pi n^3 (h\nu)^3}{3c^3 h^3} = \frac{8\pi n^3 \varepsilon^3}{3c^3 h^3}$$

Now as before take the derivative with respect to energy to get the energy density

$$\begin{aligned} \rho' &= \frac{d\rho}{d\varepsilon} = \frac{8\pi n^3 (3\varepsilon^2)}{3c^3 h^3} \\ &= \frac{8\pi n^3 \varepsilon^2}{c^3 h^3} \\ &= \frac{8\pi n^3 (h\nu)^2}{c^3 h^3} \\ &= \frac{8\pi n^3 \nu^2}{c^3 h} \end{aligned}$$

Some authors will leave it at the second step which is just $\rho' = \frac{8\pi n^3 \varepsilon^2}{c^3 h^3}$. We will go with

$$\rho' = \frac{8\pi n^3 \nu^2}{c^3 h} \quad (14.10)$$

Now earlier we had the Planck derived average energy to be

$$\langle \varepsilon \rangle = \frac{h\nu}{e^{\frac{h\nu}{kT}} - 1}$$

This leads to an expression for the average number of photons which is

$$\langle p \rangle = \frac{\langle \varepsilon \rangle}{h\nu} = \frac{1}{e^{\frac{h\nu}{kT}} - 1} \quad (14.11)$$

The resulting Planck photon density is

$$\rho_p = \frac{8\pi n^3 \nu^2}{c^3 h} \left(\frac{1}{e^{\frac{h\nu}{kT}} - 1} \right) \text{ number per unit volume per unit energy} \quad (14.12)$$

This then leads to another set of relationships between Einstein A and B coefficients

$$\boxed{B_{12} = B_{21}} \quad (14.13)$$

$$\boxed{A = \frac{8\pi n^3 \nu^2}{c^3 h} B_{12} \text{ or } = \frac{8\pi n^3 \varepsilon^2}{c^3 h^3} B_{12}} \quad (14.14)$$

The only difference between these relations and the previous ones was how the Planck distribution was defined. Number per unit volume per unit energy (as done here) or number per unit volume per unit frequency (as done previously).

Einstein A and B coefficients revisited

In the last section we derived the relationships between the Einstein coefficients for absorption, stimulated emission and spontaneous emission. In this section, let's rederive the expression in a slightly different manner, but in a way that will be useful a little later on.

Let R_{12} be the unit transition rate from the ground state to the excited state (basically the rate constant)

$$\begin{aligned} R_{12} &= P_{abs} \rho d\varepsilon \\ &\equiv B_{12} \rho \end{aligned}$$

where

- ρ is the number of photons per unit volume per unit energy (note the units!)
- $\rho d\varepsilon$ is the number of photons per unit volume
- P_{abs} is the probability for absorption per unit time
- R_{12} is the absorptions per unit volume per unit time

and where $B_{12} = P_{abs} d\varepsilon$

As before set up the rate equations except now consider explicitly the probability of occupied and unoccupied states in the valence and conduction bands. Let

- f_1 = probability of occupied valence band state
- f_2 = probability of occupied conduction band state
- $1 - f_1$ = probability of unoccupied valence band state
- $1 - f_2$ = probability of unoccupied conduction band state

1 → 2 transition (absorption) requires

- valence band state occupied (f_1)
- conduction band state empty ($1 - f_2$)

resulting in the joint probability being $f_1(1 - f_2)$ such that

$$R_{12,abs} = B_{12}\rho f_1(1 - f_2)$$

2 → 1 transition (stimulated emission) requires

- valence band state empty ($1 - f_1$)
- conduction band state occupied (f_2)

resulting in the joint probability being $f_2(1 - f_1)$ such that

$$R_{21,stim} = B_{21}\rho f_2(1 - f_1)$$

2 → 1 transition (spontaneous emission) requires

- valence band state empty ($1 - f_1$)
- conduction band state occupied (f_2)

resulting in the joint probability being $f_2(1 - f_1)$ such that

$$R_{21,spont} = A f_2(1 - f_1)$$

In all three cases f_1 and f_2 are Fermi Dirac distributions

$$f_1 = \frac{1}{e^{\frac{\varepsilon_1 - \varepsilon_F}{kT}} + 1}$$

$$f_2 = \frac{1}{e^{\frac{\varepsilon_2 - \varepsilon_F}{kT}} + 1}$$

At this point, for simplicity, assume that $g_1 = g_2$. At equilibrium the upward and downward rates equal resulting in

$$\begin{aligned} R_{12,abs} &= R_{21,stim} + R_{21,spont} \\ B_{12}\rho f_1(1-f_2) &= B_{21}\rho f_2(1-f_1) + Af_2(1-f_1) \end{aligned}$$

Rearrange this to solve for ρ

$$\rho[B_{12}f_1(1-f_2) - B_{21}f_2(1-f_1)] = Af_2(1-f_1)$$

which gives

$$\begin{aligned} \rho &= \frac{Af_2(1-f_1)}{B_{12}f_1(1-f_2) - B_{21}f_2(1-f_1)} \\ &= \frac{A}{B_{12}\frac{f_1(1-f_2)}{f_2(1-f_1)} - B_{21}} \end{aligned}$$

Now introduce the explicit expression for f_1 and f_2 . To simplify

$$\begin{aligned} \frac{f_1(1-f_2)}{f_2(1-f_1)} &= \frac{\left(\frac{1}{e^{\frac{\varepsilon_1 - \varepsilon_F}{kT}} + 1}\right) \left(1 - \frac{1}{e^{\frac{\varepsilon_2 - \varepsilon_F}{kT}} + 1}\right)}{\left(\frac{1}{e^{\frac{\varepsilon_2 - \varepsilon_F}{kT}} + 1}\right) \left(1 - \frac{1}{e^{\frac{\varepsilon_1 - \varepsilon_F}{kT}} + 1}\right)} \\ &= e^{\frac{\varepsilon_2 - \varepsilon_1}{kT}} \\ &= e^{\frac{h\nu}{kT}} \end{aligned}$$

Replace this into the main expression for ρ to get

$$\rho = \frac{A}{B_{12}\frac{f_1(1-f_2)}{f_2(1-f_1)} - B_{21}} = \frac{A}{B_{12}e^{\frac{h\nu}{kT}} - B_{21}}$$

But ρ equals the Planck distribution for photon density

$$\rho = \frac{A}{B_{12}e^{\frac{h\nu}{kT}} - B_{21}} = \frac{8\pi n^3 \nu^2}{c^3 h} \left(\frac{1}{e^{\frac{h\nu}{kT}} - 1} \right)$$

meaning that for this to be true

$$B_{12} = B_{21}$$

and

$$A = \frac{8\pi n^3 \nu^2}{c^3 h} B_{12}$$

which are exactly the same Einstein A and B relations we found before.

Emission spectrum

Here we will calculate the emission spectrum of a 3D material using the Einstein A and B coefficients. Define the (net) unit transition rate from the ground state to the excited state.

$$R_{net} = P_{abs}\rho d\varepsilon \quad (14.15)$$

where as before

- ρ is the number of photons per unit volume per unit energy
- $\rho d\varepsilon$ is the number of photons per unit volume
- P_{abs} is the probability for absorption per unit time
- R_{net} is the absorptions per unit volume per unit time

Now consider the net rate upwards from 1 to 2 including stimulated emission

$$\begin{aligned} R_{12,abs} &= B_{12}\rho f_1(1 - f_2) \text{ (transition rate per unit volume)} \\ R_{21,stim} &= B_{21}\rho f_2(1 - f_1) \text{ (transition rate per unit volume)} \end{aligned}$$

such that the net upwards rate is

$$\begin{aligned} R_{net} &= R_{12,abs} - R_{21,stim} \text{ (transition rate per unit volume)} \\ &= B_{12}\rho f_1(1 - f_2) - B_{21}\rho f_2(1 - f_1) \end{aligned}$$

Since we have already shown that $B_{12} = B_{21}$

$$R_{net} = B_{12}\rho[f_1(1 - f_2) - f_2(1 - f_1)] \quad (14.16)$$

Equating the two expressions for ρ gives

$$\begin{aligned} P_{abs}\rho d\varepsilon &= B_{12}\rho[f_1(1 - f_2) - f_2(1 - f_1)] \\ P_{abs}d\varepsilon &= B_{12}[f_1(1 - f_2) - f_2(1 - f_1)] \end{aligned}$$

Here we relate P_{abs} to the absorption coefficient as follows

$$P_{abs} = \frac{c}{n}\alpha$$

How? If P is the absolute probability for an absorption event and $\frac{dP}{dt} \equiv P_{abs}$

$$\begin{aligned} \frac{dP}{dt} &= \frac{dP}{dz} \left(\frac{dz}{dt} \right) \\ P_{abs} &= \alpha \left(\frac{c}{n} \right) \end{aligned}$$

Insert this P_{abs} expression into the above equivalence to get

$$\alpha \left(\frac{c}{n} \right) d\varepsilon = B_{12} [f_1(1 - f_2) - f_2(1 - f_1)]$$

Rearrange this to solve for B_{12}

$$B_{12} = \alpha \left(\frac{c}{n} \right) \frac{d\varepsilon}{[f_1(1 - f_2) - f_2(1 - f_1)]}$$

Relate B_{12} to A using our derived Einstein A and B coefficients.

$$\begin{aligned} A &= \frac{8\pi n^3 \nu^2}{c^3 h} B_{12} = \frac{8\pi \nu^2}{h} \left(\frac{n}{c} \right)^3 \alpha \left(\frac{c}{n} \right) \frac{d\varepsilon}{[f_1(1 - f_2) - f_2(1 - f_1)]} \\ &= \frac{8\pi \nu^2}{h} \left(\frac{n}{c} \right)^2 \alpha \frac{d\varepsilon}{[f_1(1 - f_2) - f_2(1 - f_1)]} \end{aligned}$$

Now recall that

$$R_{21,spont} = A f_2(1 - f_1)$$

such that

$$\begin{aligned} R_{21,spont} &= \alpha \frac{8\pi \nu^2}{h} \left(\frac{n}{c} \right)^2 \frac{f_2(1 - f_1) d\varepsilon}{[f_1(1 - f_2) - f_2(1 - f_1)]} \\ &= \alpha \frac{8\pi \nu^2}{h} \left(\frac{n}{c} \right)^2 \frac{d\varepsilon}{\frac{f_1(1 - f_2)}{f_2(1 - f_1)} - 1} \end{aligned}$$

Now for convenience define

$$r_{spont}(\varepsilon) d\varepsilon = R_{21}$$

where $r_{spont}(\varepsilon)$ is the transition rate per unit volume per unit energy. By inspection

$$r_{spont}(\varepsilon) = \alpha \frac{8\pi \nu^2}{h} \left(\frac{n}{c} \right)^2 \frac{1}{\frac{f_1(1 - f_2)}{f_2(1 - f_1)} - 1}$$

and

$$\begin{aligned} f_1 &= \frac{1}{e^{\frac{\varepsilon_1 - \varepsilon_F}{kT}} + 1} \\ f_2 &= \frac{1}{e^{\frac{\varepsilon_2 - \varepsilon_F}{kT}} + 1} \end{aligned}$$

We've solved the ratio in the denominator before giving

$$\frac{f_1(1-f_2)}{f_2(1-f_1)} = e^{\frac{h\nu}{kT}}$$

Putting everything together

$$r_{spont}(\varepsilon) = \alpha \left(\frac{8\pi\nu^2}{h} \right) \left(\frac{n}{c} \right)^2 \left(\frac{1}{e^{\frac{h\nu}{kT}} - 1} \right)$$

If $h\nu > kT$ the term in parenthesis $\sim e^{-\frac{h\nu}{kT}}$ and the main expression becomes

$$r_{spont}(\varepsilon) = \alpha \left(\frac{8\pi\nu^2}{h} \right) \left(\frac{n}{c} \right)^2 e^{-\frac{h\nu}{kT}} \quad (14.17)$$

We can stop here or continue to express this as an intensity.

$$I_{3D}(\varepsilon) = \varepsilon \frac{r_{spont}(\varepsilon)}{\alpha_{exc}(\varepsilon)}$$

where $\alpha_{exc}(\varepsilon)$ is the absorption coefficient at the particular excitation position. This value is a constant.

$$\begin{aligned} I_{3D}(\varepsilon) &= (h\nu) \frac{\alpha \left(\frac{8\pi\nu^2}{h} \right) \left(\frac{n}{c} \right)^2 e^{-\frac{h\nu}{kT}}}{\alpha_{exc}(\varepsilon)} \\ &= \frac{\alpha}{\alpha_{exc}} (8\pi\nu^3) \left(\frac{n}{c} \right)^2 e^{-\frac{h\nu}{kT}} \end{aligned} \quad (14.18)$$

Finally, for a 3D material the absorption coefficient is proportional to the joint density of states

$$\alpha \propto A \sqrt{\varepsilon - \varepsilon_g}$$

where A is a constant. This leads to our final expression for the emission spectrum

$$I_{3D} = \frac{A \sqrt{\varepsilon - \varepsilon_g}}{\alpha_{exc}} (8\pi\nu^3) \left(\frac{n}{c} \right)^2 e^{-\frac{h\nu}{kT}}$$

$$\boxed{I_{3D} = A' \sqrt{\varepsilon - \varepsilon_g} e^{-\frac{h\nu}{kT}}} \quad (14.19)$$

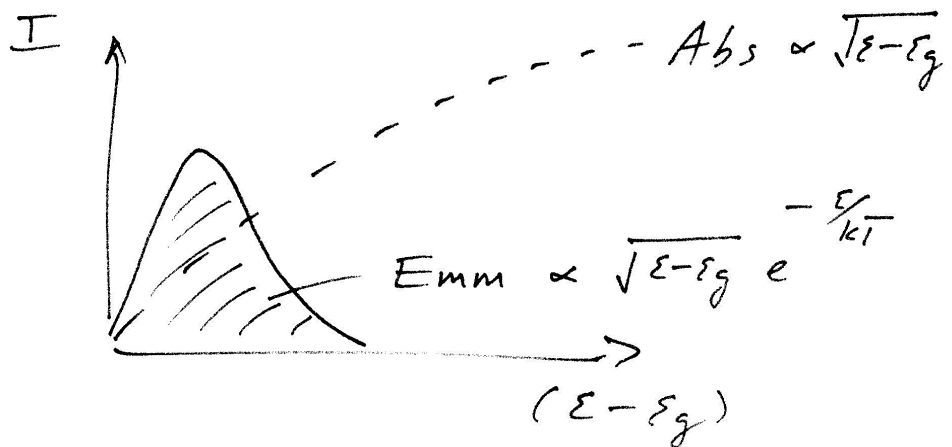


Figure 14.2: Derived emission spectrum for a 3D material using Einstein A and B coefficients

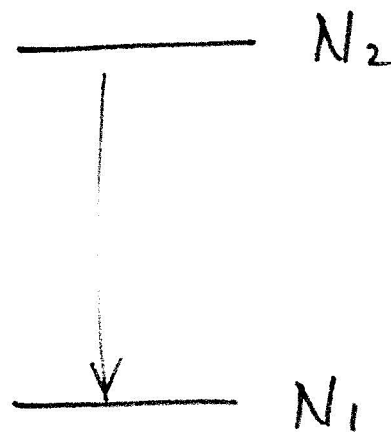


Figure 14.3: Two level system considered in pulsed laser experiment. Only the (downward) radiative transition is considered.

Quantum yields and lifetimes

Here we discuss some complementary aspects to the emission. Imagine populating the excited state of the system and immediately (i.e. instantaneously) turning off the light. Basically this is a pulsed experiment.

The depopulation of the excited state occurs by spontaneous emission since stimulated emission only occurs in the presence of the excitation. The relevant rate equation is

$$\frac{dN_2}{dt} = -AN_2$$

or if $A = k_{rad}$

$$\begin{aligned} \frac{dN_2}{dt} &= k_{rad}N_2 \\ \frac{dN_2}{N_2} &= -k_{rad}dt \\ \ln N_2 &= -k_{rad}t + (\text{const}) \\ N_2 &= Ce^{-k_{rad}t} \end{aligned}$$

At $t = 0$ $C = N_2(0)$ resulting in

$$N_2(t) = N_2(0)e^{-k_{rad}t} \quad (14.20)$$

The decay will be exponential and $\tau = \frac{1}{k_{rad}}$ is called the lifetime of the excited state.

In general, however, since we don't live in a perfect world, there are other de-excitation pathways. These include energy transfer or non-radiative decay through defect states. So in general the total decay rate out of the excited state is the sum of all rates

$$k_{tot} = k_{rad} + k_1 + k_2 + k_3 + \dots$$

giving

$$\begin{aligned} \frac{dN_2}{dt} &= -(k_{rad} + k_1 + k_2 + k_3 + \dots)N_2 \\ &= -\left(k_{rad} + \sum_i^n k_i\right)N_2 \\ &= -k_{tot}N_2 \end{aligned}$$

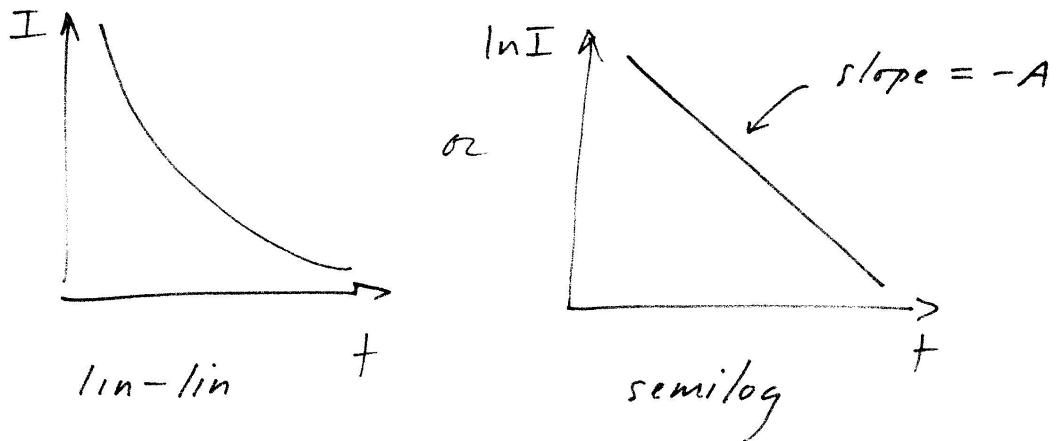


Figure 14.4: Linear and semilogarithmic plots of the excited state decay profile.

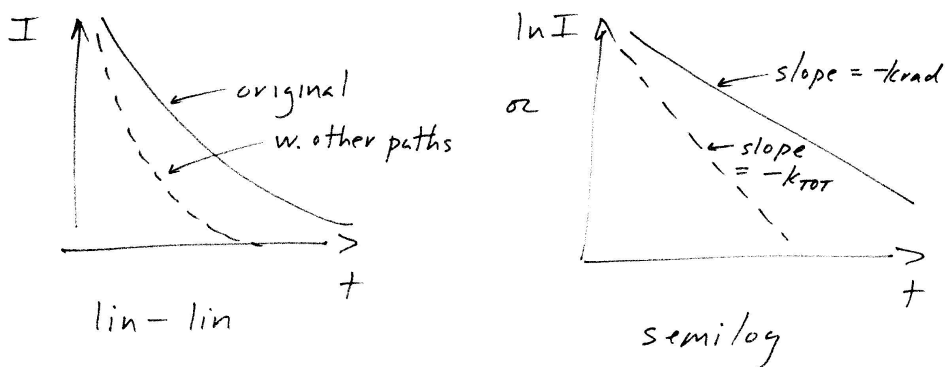


Figure 14.5: Linear and semilogarithmic plots of the excited state decay in the absence and presence of multiple decay pathways.

Only the k_{rad} pathway gives you emission. Therefore, the efficiency of emission is called the quantum yield (QY) and is defined by

$$QY = \frac{k_{rad}}{k_{tot}} \text{ or } \frac{\frac{1}{\tau_{rad}}}{\frac{1}{\tau_{tot}}} = \frac{\tau_{tot}}{\tau_{rad}} \quad (14.21)$$

For most applications, one desires a QY that is as close to unity as possible ($QY = 1$). For example, such applications could involve lasers, light emitting diodes, fluorescent tags and so forth.

Exercises

1. The first part of this chapter described three processes: (stimulated) absorption, stimulated emission, and spontaneous emission. Rationalize why there isn't a fourth process called spontaneous absorption.
2. Spontaneous emission competes with thermally stimulated emission (Planck radiation is the source). Show that at 300 K, thermal stimulation dominates for frequencies well below 5×10^{12} Hz whereas spontaneous emission dominates for frequencies well above 5×10^{12} Hz. Which mechanism dominates for visible light.
3. Come up with an alternative means of measuring the quantum yield of a system based on its fluorescence decay profile. See Figure 9.5. Assume an exponential decay and consider its pure radiative decay as well as its radiative decay in the presence of other non-radiative recombination pathways.
4. Show mathematically how one extracts the average decay time of an exponential process. Compare this to the $1/e$ lifetime.
5. Derive a relation between the half life (50% mark) and the lifetime of the state.
6. Explain how one measures the quantum yield of something in real life.
7. Show how the Plank expression reduces to both the Wein displacement law and the Stefan Boltzman law for blackbody radiation. What are the Wein and Stefan Boltzman laws anyway?

Chapter 15

Bands

Kronig Penney Approximation with Rectangular Barriers

An approximation to a crystalline solid is a periodic potential consisting of rectangular barriers. This is called the Kronig-Penny model and has analytical solutions. Although this is a crude approximation to the true potential experienced by the electron, the importance of the model lies in the idea that the periodicity of the potential is the root cause for the emergence of what are called band gaps. There are discontinuities in the energies that an electron can have. This ultimately leads to the distinction between metals, semiconductors and insulators.

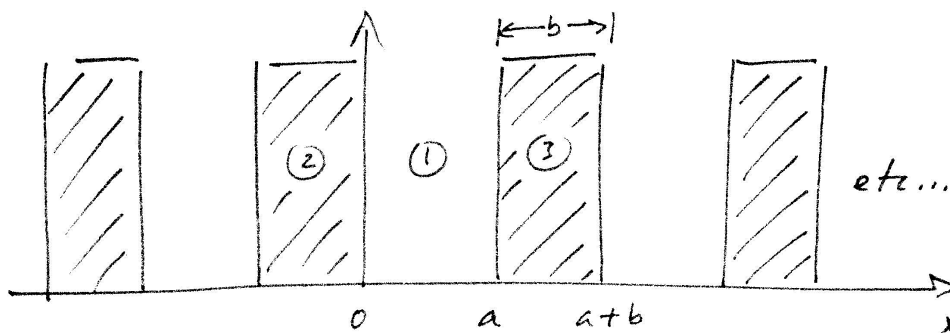


Figure 15.1: Periodic potential of rectangular barriers

Region 1:

$$\begin{aligned}
-\frac{\hbar^2}{2m} \frac{d^2\Psi}{dx^2} &= \varepsilon\Psi & (15.1) \\
\frac{d^2\Psi}{dx^2} &= -\frac{2m\varepsilon}{\hbar^2}\Psi \\
\frac{d^2\Psi}{dx^2} + \frac{2m\varepsilon}{\hbar^2}\Psi &= 0
\end{aligned}$$

Therefore

$$\frac{d^2\Psi}{dx^2} + k^2\Psi = 0 \quad (15.2)$$

where

$$k^2 = \frac{2m\varepsilon}{\hbar^2}$$

Region 2:

$$\begin{aligned}
-\frac{\hbar^2}{2m} \frac{d^2\Psi}{dx^2} + V\Psi &= \varepsilon\Psi \\
-\frac{\hbar^2}{2m} \frac{d^2\Psi}{dx^2} &= (\varepsilon - V)\Psi \\
\frac{d^2\Psi}{dx^2} &= \frac{2m(V - \varepsilon)}{\hbar^2}\Psi \\
\frac{d^2\Psi}{dx^2} - \frac{2m(V - \varepsilon)}{\hbar^2}\Psi &= 0
\end{aligned}$$

Therefore

$$\frac{d^2\Psi}{dx^2} - \phi^2\Psi = 0 \quad (15.3)$$

where

$$\phi^2 = \frac{2m(V - \varepsilon)}{\hbar^2}$$

The wavefunctions in regions 1 and 2 are:

$$\Psi_1 = Ae^{ikx} + Be^{-ikx} \text{ region 1} \quad (15.4)$$

$$\Psi_2 = Ce^{\phi x} + De^{-\phi x} \text{ region 2} \quad (15.5)$$

Now apply Bloch's theorem where

$$\Psi(x+a) = \Psi(x)e^{ila} \text{ for delta function potential with period } a$$

or alternatively

$$\Psi(x+(a+b)) = \Psi(x)e^{il(a+b)} \text{ for periodic potential with period } a+b$$

leading to

Region 3:

$$\Psi_3 = \Psi_2 e^{il(a+b)} \quad (15.6)$$

or

$$\Psi_3 = (Ce^{\phi x} + De^{-\phi x}) e^{il(a+b)} \quad (15.7)$$

Now apply the matching conditions for wavefunctions

$$\Psi_1(0) = \Psi_2(0) \rightarrow A + B = C + D$$

$$\Psi_1'(0) = \Psi_2'(0) \rightarrow ik(A - B) = \phi(C - D)$$

$$\Psi_1(a) = \Psi_3(a) \rightarrow Ae^{ika} + Be^{-ika} = (Ce^{-\phi b} + De^{\phi b})e^{il(a+b)}$$

$$\Psi_1'(a) = \Psi_3'(a) \rightarrow ikAe^{ika} - ikBe^{-ika} = (\phi Ce^{-\phi b} - D\phi e^{\phi b})e^{il(a+b)}$$

where we used the fact that

$$\Psi_3(a) = \Psi_2(-b)e^{il(a+b)}$$

$$\Psi_3'(a) = \Psi_2'(-b)e^{il(a+b)}$$

This leads to a system of four equations with four unknowns (A, B, C, D). The system can be recast into matrix form. However, to give a non-trivial solution (A=B=C=D=0) one must have the determinant of prefactors equal zero. This means solving the following equation.

$$\begin{vmatrix} 1 & 1 & -1 & -1 \\ ik & -ik & -\phi & \phi \\ e^{ika} & e^{-ika} & -e^{-\phi b + il(a+b)} & -e^{\phi b + il(a+b)} \\ ik e^{ika} & -ik e^{-ika} & -\phi e^{-\phi b + il(a+b)} & \phi e^{\phi b + il(a+b)} \end{vmatrix} = 0$$

Most texts will skip the tedious steps to the final expression. Here we outline what needs to be done. The solution can be approached a number

of ways. This is just one. Apply the following steps to generate successively simpler determinants.

$$-ik(\text{row1}) + (\text{row2}) \rightarrow (\text{row2})$$

$$-ik(\text{row2}) + (\text{row4}) \rightarrow (\text{row4})$$

yeilding

$$\begin{vmatrix} 1 & 1 & -1 & -1 \\ 0 & -2ik & (ik - \phi) & (ik + \phi) \\ e^{ika} & e^{-ika} & -e^{-\phi b + il(a+b)} & -e^{\phi b + il(a+b)} \\ 0 & -2ike^{-ika} & (ik - \phi)e^{-\phi b + il(a+b)} & (ik + \phi)e^{\phi b + il(a+b)} \end{vmatrix} = 0$$

$$-e^{ika}(\text{row1}) + (\text{row3}) \rightarrow (\text{row3})$$

yeilding

$$\begin{vmatrix} 1 & 1 & -1 & -1 \\ 0 & -2ik & (ik - \phi) & (ik + \phi) \\ 0 & e^{-ika} - e^{ika} & e^{ika} - e^{-\phi b + il(a+b)} & e^{ika} - e^{\phi b + il(a+b)} \\ 0 & -2ike^{-ika} & (ik - \phi)e^{-\phi b + il(a+b)} & (ik + \phi)e^{\phi b + il(a+b)} \end{vmatrix} = 0$$

then

$$-e^{-ika}(\text{row2}) + (\text{row4}) \rightarrow (\text{row4})$$

yeilding

$$\begin{vmatrix} 1 & 1 & -1 & -1 \\ 0 & -2ik & (ik - \phi) & (ik + \phi) \\ 0 & e^{-ika} - e^{ika} & e^{ika} - e^{-\phi b + il(a+b)} & e^{ika} - e^{\phi b + il(a+b)} \\ 0 & 0 & (ik - \phi)[-e^{-ika} + e^{-\phi b + il(a+b)}] & (ik + \phi)[-e^{-ika} + e^{\phi b + il(a+b)}] \end{vmatrix} = 0$$

At this point the finding the 4x4 determinant becomes the same as finding the sub 3x3 determinant. To simplify our notation let

$$(1) = (ik - \phi) [-e^{-ika} + e^{-\phi b + il(a+b)}]$$

$$(2) = (ik + \phi) [-e^{-ika} + e^{\phi b + il(a+b)}]$$

then the 3x3 becomes

$$\begin{vmatrix} -2ik & (ik - \phi) & (ik + \phi) \\ e^{-ika} - e^{ika} & e^{ika} - e^{-\phi b + il(a+b)} & e^{ika} - e^{\phi b + il(a+b)} \\ 0 & (ik - \phi)(1) & (ik + \phi)(2) \end{vmatrix} = 0$$

which is equivalent to

$$\begin{aligned} & - 2ik \begin{vmatrix} e^{ika} - e^{-\phi b + il(a+b)} & e^{ika} - e^{\phi b + il(a+b)} \\ (ik - \phi)(1) & (ik + \phi)(2) \end{vmatrix} \\ & + 2i \sin(ka) \begin{vmatrix} (ik - \phi) & (ik + \phi) \\ (ik - \phi)(1) & (ik + \phi)(2) \end{vmatrix} = 0 \\ & - 2ik \begin{vmatrix} e^{ika} - e^{-\phi b + il(a+b)} & e^{ika} - e^{\phi b + il(a+b)} \\ (ik - \phi)(1) & (ik + \phi)(2) \end{vmatrix} \\ & + 2i \sin(ka) [(-k^2 - \phi^2)(2) - (-k^2 - \phi^2)(1)] = 0 \\ & - 2ik \begin{vmatrix} e^{ika} - e^{-\phi b + il(a+b)} & e^{ika} - e^{\phi b + il(a+b)} \\ (ik - \phi)(1) & (ik + \phi)(2) \end{vmatrix} \\ & + 2i(-k^2 - \phi^2) \sin(ka) [(2) - (1)] = 0 \\ & - 2ik \begin{vmatrix} e^{ika} - e^{-\phi b + il(a+b)} & e^{ika} - e^{\phi b + il(a+b)} \\ (ik - \phi)(1) & (ik + \phi)(2) \end{vmatrix} \\ & + 4i(-k^2 - \phi^2) \sin(ka) \sinh(\phi b) e^{il(a+b)} = 0 \end{aligned}$$

The top determinant is a little tedious but can be evaluate to give

$$\begin{aligned} & - 4ik \left[-\phi - \phi e^{2il(a+b)} + 2e^{il(a+b)} (-k \sinh(\phi b) \sin(ka) + \phi \cosh(\phi b) \cos(ka)) \right] \\ & + 4i(-k^2 - \phi^2) \sin(ka) \sinh(\phi b) e^{il(a+b)} = 0 \end{aligned}$$

This expands to

$$4ik\phi + 4ik\phi e^{2il(a+b)} - 8ike^{il(a+b)} (-k \sinh(\phi b) \sin(ka) + \phi \cosh(\phi b) \cos(ka)) \\ + 4i(-k^2 - \phi^2)e^{il(a+b)} \sin(ka) \sinh(\phi b) = 0$$

Multiply by $e^{-il(a+b)}$ to get

$$4ik\phi e^{-il(a+b)} + 4ik\phi e^{il(a+b)} - 8ik(-k \sinh(\phi b) \sin(ka) + \phi \cosh(\phi b) \cos(ka)) \\ + 4i(-k^2 - \phi^2) \sinh(\phi b) \sin(ka) = 0$$

Divide by $4i$ to get

$$2k\phi \cos(l(a+b)) + 2k^2 \sinh(\phi b) \sin(ka) - 2k\phi \cosh(\phi b) \cos(ka) \\ + (-k^2 - \phi^2) \sinh(\phi b) \sin(ka) = 0$$

This leads to the following textbook expression after some simplification

$$\boxed{\left(\frac{\phi^2 - k^2}{2k\phi}\right) \sinh(\phi b) \sin(ka) + \cosh(\phi b) \cos(ka) = \cos(l(a+b))} \quad (15.8)$$

Kronig Penney $V=0$ Limit

Starting with our Kronig Penney expression derived above

$$\left(\frac{\phi^2 - k^2}{2k\phi}\right) \sinh(\phi b) \sin(ka) + \cosh(\phi b) \cos(ka) = \cos(l(a+b))$$

recall that

$$\phi^2 = \frac{2m(V - \varepsilon)}{\hbar^2} \rightarrow \phi = \sqrt{\frac{2m(V - \varepsilon)}{\hbar^2}} \\ k^2 = \frac{2m\varepsilon}{\hbar^2} \rightarrow k = \sqrt{\frac{2m\varepsilon}{\hbar^2}}$$

if $V=0$

$$\phi^2 = -\frac{2m\varepsilon}{\hbar^2} \rightarrow \phi = i\sqrt{\frac{2m\varepsilon}{\hbar^2}} = ik$$

the Kronig Penney expression becomes

$$\left(\frac{-k^2 - k^2}{2ik^2}\right) \sinh(ikb) \sin(ka) + \cosh(ikb) \cos(ka) = \cos(l(a+b))$$

where

$$\begin{aligned} \sinh(iu) &= i \sin(u) \\ \cosh(iu) &= \cos(u) \end{aligned}$$

leading to

$$-\sin(kb) \sin(ka) + \cos(kb) \cos(ka) = \cos(l(a+b))$$

For this to be true

$$\begin{aligned} l &= k \\ &\text{or} \\ l &= \sqrt{\frac{2m\varepsilon}{\hbar^2}} \end{aligned} \tag{15.9}$$

This gives the energies of a free particle

$$\boxed{\varepsilon = \frac{\hbar^2 l^2}{2m}}$$

Kronig Penney $V = \infty$ Limit

Starting with our original Kronig Penney expression

$$\left(\frac{\phi^2 - k^2}{2k\phi}\right) \sinh(\phi b) \sin(ka) + \cosh(\phi b) \cos(ka) = \cos(l(a+b))$$

where

$$\begin{aligned} k^2 &= \frac{2m\varepsilon}{\hbar^2} \\ \phi^2 &= \frac{2m(V - \varepsilon)}{\hbar^2} \end{aligned}$$

Let's rearrange everything in terms of unitless variables if possible. Let $\delta = \varepsilon_V$.

$$\begin{aligned}
 k^2 &= \frac{2m\varepsilon}{\hbar^2} \left(\frac{V}{V} \right) = \frac{2mV}{\hbar^2} \left(\frac{\varepsilon}{V} \right) = \frac{2mV}{\hbar^2} \delta \\
 \phi^2 &= \frac{2m}{\hbar^2} \left(V - \varepsilon \frac{V}{V} \right) = \frac{2mV}{\hbar^2} (1 - \delta)
 \end{aligned}$$

where in both cases $\delta = \varepsilon/V$. Therefore

$$\begin{aligned}
 \left(\frac{\phi^2 - k^2}{2k\phi} \right) &= \frac{\left[\frac{2mV}{\hbar^2} (1 - \delta) \right] - \left[\frac{2mV}{\hbar^2} \delta \right]}{2\sqrt{\frac{2mV}{\hbar^2} \delta} \sqrt{\frac{2mV}{\hbar^2} (1 - \delta)}} \\
 &= \frac{(1 - 2\delta)}{2\sqrt{\delta(1 - \delta)}}
 \end{aligned}$$

Consider also how ϕb and ka simplify.

$$\begin{aligned}
 \phi b &= \sqrt{\frac{2mV}{\hbar^2} (1 - \delta)} b = \sqrt{\frac{2mV}{\hbar^2} (1 - \delta)} b \left(\frac{a}{a} \right) \\
 &= \left[\sqrt{\frac{2mV}{\hbar^2}} a \right] \left(\frac{b}{a} \right) \sqrt{1 - \delta} \\
 &= Ar\sqrt{1 - \delta}
 \end{aligned}$$

where $A = \sqrt{\frac{2mV}{\hbar^2}} a$ and $r = b/a$ are constants. Likewise

$$\begin{aligned}
 ka &= \sqrt{\frac{2mV\delta}{\hbar^2}} a \\
 &= A\sqrt{\delta}
 \end{aligned}$$

Replace both into our previous equation to get

$$\begin{aligned}
 \frac{(1 - 2\delta)}{2\sqrt{\delta(1 - \delta)}} \sinh(Ar\sqrt{1 - \delta}) \sin(A\sqrt{\delta}) + \cosh(Ar\sqrt{1 - \delta}) \cos(A\sqrt{\delta}) \\
 = \cos(la(1 + r))
 \end{aligned}$$

Case ($\delta < 1$)

$$\begin{aligned} \frac{(1-2\delta)}{2\sqrt{\delta(1-\delta)}} \sinh(Ar\sqrt{1-\delta}) \sin(A\sqrt{\delta}) + \cosh(Ar\sqrt{1-\delta}) \cos(A\sqrt{\delta}) \\ = \cos(la(1+r)) \end{aligned} \quad (15.10)$$

where $\delta < 1$ implicitly.

Case ($\delta > 1$)

In this case

$$\begin{aligned} \sqrt{1-\delta} &= i\sqrt{\delta-1} \\ \sinh(iu) &= i\sin(u) \\ \cosh(iu) &= \cos(u) \end{aligned}$$

Replace into our previous equation to get

$$\begin{aligned} \frac{(1-2\delta)}{2i\sqrt{\delta(\delta-1)}} i \sin(Ar\sqrt{\delta-1}) \sin(A\sqrt{\delta}) + \cos(Ar\sqrt{\delta-1}) \cos(A\sqrt{\delta}) \\ = \cos(la(1+r)) \end{aligned}$$

which reduces to

$$\begin{aligned} \frac{(1-2\delta)}{2\sqrt{\delta(\delta-1)}} \sin(Ar\sqrt{\delta-1}) \sin(A\sqrt{\delta}) + \cos(Ar\sqrt{\delta-1}) \cos(A\sqrt{\delta}) \\ = \cos(la(1+r)) \end{aligned} \quad (15.11)$$

Case ($\delta = 1$)

Here note that

$$\begin{aligned} \sinh(0) &= 0 \\ \cosh(0) &= 1 \end{aligned}$$

and that

$$\lim_{t \rightarrow 0} \frac{\sin(t)}{t} = 1$$

Rearrange the original expression as follows

$$\begin{aligned} \frac{(1-2\delta)}{2\sqrt{\delta}} \left(\frac{Ar}{Ar}\right) \frac{\sin(Ar\sqrt{\delta-1})}{\sqrt{\delta-1}} \sin(A\sqrt{\delta}) + \cos(Ar\sqrt{\delta-1}) \cos(A\sqrt{\delta}) \\ = \cos(la(1+r)) \end{aligned}$$

Let $\delta = 1$ and recall the limit of $\sin(t)/t$ as $t \rightarrow 0$.

$$-\frac{1}{2}Ar \frac{\sin(Ar)}{Ar} \sin(A) + \cos(A) = \cos(la(1+r))$$

continue reducing the expression to get

$$-\frac{Ar}{2} \sin(A) + \cos(A) = \cos(la(1+r)) \quad (15.12)$$

All cases ($\delta < 1$), ($\delta > 1$) and ($\delta = 1$)

In all three cases above the resulting relationship can be expressed crudely as some function of δ equals the right hand side.

$$F(\delta) = \cos(la(1+r)) \quad (15.13)$$

This relation cannot exactly be solved analytically but is done numerically. However one notes at this point that the right hand side of the expression $\cos(la(1+r))$ will cycle periodically between +1 and -1.

Let $V \rightarrow \infty$

Continuing our simplification of the this expression, we let $V \rightarrow \infty$. As a consequence $\delta = \frac{\epsilon}{V} \rightarrow 0$ and $A \rightarrow \infty$.

$$\begin{aligned} \frac{(1-2\delta)}{2\sqrt{\delta(1-\delta)}} \sinh(Ar\sqrt{1-\delta}) \sin(A\sqrt{\delta}) + \cosh(Ar\sqrt{1-\delta}) \cos(A\sqrt{\delta}) \\ = \cos(la(1+r)) \end{aligned}$$

approximately becomes

$$\frac{1}{2\sqrt{\delta}} \sinh(Ar) \sin(A\sqrt{\delta}) + \cosh(Ar) = \cos(A\sqrt{\delta}) \cos(la(1+r))$$

Here in the approximation, the first term on the left hand side will dominate the second ($\cosh(Ar)$) because of the $\frac{1}{\sqrt{\delta}}$ term. We then get something like

$$\frac{1}{2\sqrt{\delta}} \left(\frac{e^{Ar} - e^{-Ar}}{2} \right) \sin(A\sqrt{\delta}) \approx \cos(la(1+r))$$

or

$$\frac{e^{Ar}}{4\sqrt{\delta}} \sin(A\sqrt{\delta}) \approx \cos(la(1+r))$$

Since $\frac{e^{Ar}}{4\sqrt{\delta}}$ diverges, $\sin(A\sqrt{\delta})$ must approach zero to make the left hand side fall between +1 and -1 which are the limits of the right hand side ($\cos(la(1+r))$).

In the limit that $\sin(A\sqrt{\delta}) = 0$,

$$A\sqrt{\delta} = n\pi$$

where recall that $A = \sqrt{\frac{2mV}{\hbar^2}}a$ and $\sqrt{\delta} = \sqrt{\frac{\varepsilon}{V}}$. This then leads to the familiar expression for the energies of a particle in a box.

$$\boxed{\varepsilon = \frac{n^2 \hbar^2}{8ma^2}} \quad (15.14)$$

Kronig Penney Approximation with Delta Function Barriers

Sometimes the Kronig Penney relation is derived assuming a Delta function potential rather than using rectangular barriers with finite width.

We run through this derivation for completeness. So starting with our original Kronig Penney relation

$$\frac{\phi^2 - k^2}{2k\phi} \sinh(\phi b) \sin(ka) + \cosh(\phi b) \cos(ka) = \cos(l(a+b))$$

One generally makes the following substitution. Let

$$P = \frac{\phi^2 ab}{2} \quad (15.15)$$

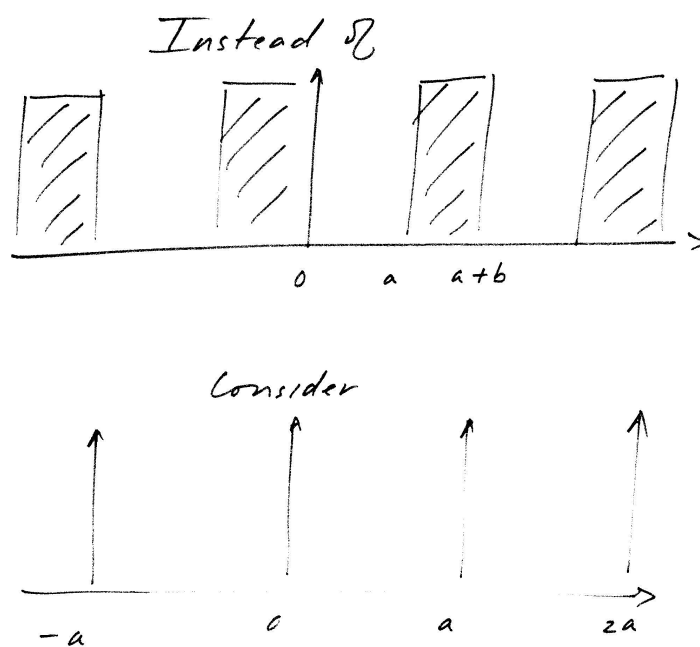


Figure 15.2: Above: original periodic potential of rectangular barriers. Below: approximation using a periodic array of delta functions.

where P is basically a dimensionless number. As a consequence $b = \frac{2P}{\phi^2 a}$. Replace into the original expression and make the following simplifications

$$\frac{\phi^2 - k^2}{2k\phi} \approx \frac{\phi^2}{2k\phi} = \frac{\phi}{2k} \quad (15.16)$$

Recall that $\phi = \sqrt{\frac{2m(V-\varepsilon)}{\hbar^2}}$ and $k = \sqrt{\frac{2m\varepsilon}{\hbar^2}}$ with $\phi \rightarrow \infty$ since $V \rightarrow \infty$. as well as

$$\begin{aligned} \sinh(\phi b) &= \lim_{\left(\frac{2P}{\phi a}\right) \rightarrow 0} \sinh\left(\frac{2P}{\phi a}\right) = \frac{2P}{\phi a} \\ \cosh(\phi b) &= \lim_{\left(\frac{2P}{\phi a}\right) \rightarrow 0} \cosh\left(\frac{2P}{\phi a}\right) = 1 \end{aligned}$$

where we have made use of a Taylor series as follows.

$$\begin{aligned} \cosh(x) &= \frac{e^x + e^{-x}}{2} \approx \frac{(1+x) + (1-x)}{2} = 1 \\ \sinh(x) &= \frac{e^x - e^{-x}}{2} \approx \frac{(1+x) - (1-x)}{2} = x \end{aligned}$$

Then our main equation becomes

$$P \frac{\sin(ka)}{ka} + \cos(ka) = \cos(la) \quad (15.17)$$

At this point we can take our limits of $V = 0 (P = 0)$ or $V = \infty (P = \infty)$ and see what happens.

Case ($P = 0$)

Here we get

$$\cos(ka) = \cos(la)$$

Therefore $k = l$ or that $l = \sqrt{\frac{2m\varepsilon}{\hbar^2}}$ leading to

$$\boxed{\varepsilon = \frac{\hbar^2 l^2}{2m}} \quad (15.18)$$

which again is the energy of a free electron.

Case ($P = \infty$)

Here to allow the left hand side of the expression to have finite values $\frac{\sin(ka)}{ka}$ must be very small. In the limit that $\frac{\sin(ka)}{ka} = 0$, $ka = n\pi$ This immediately leads to

$$\boxed{\varepsilon = \frac{n^2 h^2}{8ma^2}} \quad (15.19)$$

which are the energies of a particle in a box.

Summary

One can therefore see that the Kronig Penney model lies in between the two limiting cases for the behavior of an electron in a potential. At one end there is the free electron limit and at the other is the particle in a box limit. In between is the regime giving existence to the notion of bands separated by energy gaps. So as usually mentioned in textbooks, at this point we point out that metals are those materials that have their conduction band half full, semiconductors are those materials where the valence band is full but where the band gap is not humongous (let's say < 4 eV) and insulators are those materials where the valence band is full and where the band gap is huge (say > 4 eV).

Exercises

1. Consider the rectangular barrier Kronig Penney model discussed early on. Use Mathcad, Matlab, Mathematica or your favorite mathematical modeling program to visualize the actual bands. Choose a potential well width of your own choosing, a barrier width, again defined by you, and a nominally high barrier height. Plot the free electron energies on top of the bands you have drawn. Show the plot in both the periodic zone scheme and the reduced zone scheme.
2. Consider the delta function modification of the Kronig Penney model. Again use a mathematical modeling program to draw the bands in the periodic and reduced zone schemes. Choose whatever barrier height, well and barrier width you desire.
3. What temperature in K would you need to promote an electron from the valence band to the conduction band of a semiconductor with a

Kronig Penney Model

some definitions we will need

$\hbar = 1.0546 \cdot 10^{-34}$ self explanatory, units of Joule * seconds

$m_0 = 9.11 \cdot 10^{-31}$ free electron mass, units of kg

$eV = 1.602 \cdot 10^{-19}$ conversion factor from eV to joule, units of coulomb

initial user defined values

$a = 2 \cdot 10^{-10}$ potential well width, units of meters

$\zeta = 0.1$ user defined multiplicative constant for barrier width, see below

$V_0 = 100$ this is the barrier height, units of eV

$b(\zeta) = \zeta \cdot a$ barrier width as multiple of well width, units of meters

$P(\zeta) = a + b(\zeta)$ this is the period of the potential, made function of b since we change later

$x = 0, 0.01 \cdot P(\zeta) .. 3 \cdot P(\zeta)$ this is the range in real space to be evaluated

$n = 1$ first zone only

$k_{max} = n \cdot \left(\frac{\pi}{P(\zeta)} \right)$

$k = -k_{max}, -0.95 \cdot k_{max} .. k_{max}$ this is the range of k space to be evaluated

$\alpha(E) = \sqrt{\frac{2 \cdot m_0 \cdot (E \cdot eV)}{\hbar^2}}$ this is part of the exponent to wavefunction in zero potential well

$\beta(E, V_0) = \sqrt{\frac{2 \cdot m_0 \cdot (V_0 - E) \cdot eV}{\hbar^2}}$ this is part of the exponent to wavefunction in barrier

$C(E, k, \zeta, V_0) = \cosh(\beta(E, V_0) \cdot b(\zeta)) \cdot \cos(\alpha(E) \cdot a) - \cos(k \cdot P(\zeta))$ C is part of full expression which we are breaking up for printing purposes, otherwise overruns right margin

$F(E, k, \zeta, V_0) = \frac{\beta(E, V_0)^2 - \alpha(E)^2}{2 \cdot \alpha(E) \cdot \beta(E, V_0)} \cdot \sinh(\beta(E, V_0) \cdot b(\zeta)) \cdot \sin(\alpha(E) \cdot a) + C(E, k, \zeta, V_0)$
this is that large equation whose roots we desire

$V1(x, \zeta, V_0) = V_0 \cdot [\Phi(x - a) \cdot \Phi(P(\zeta) - x) + \Phi[x - (P(\zeta) + a)] \cdot \Phi(2 \cdot P(\zeta) - x)]$
 Breaking up the potential for printing purposes, otherwise overruns right margin

Figure 15.3: Mathcad numerical solutions to the general Kronig Penney model

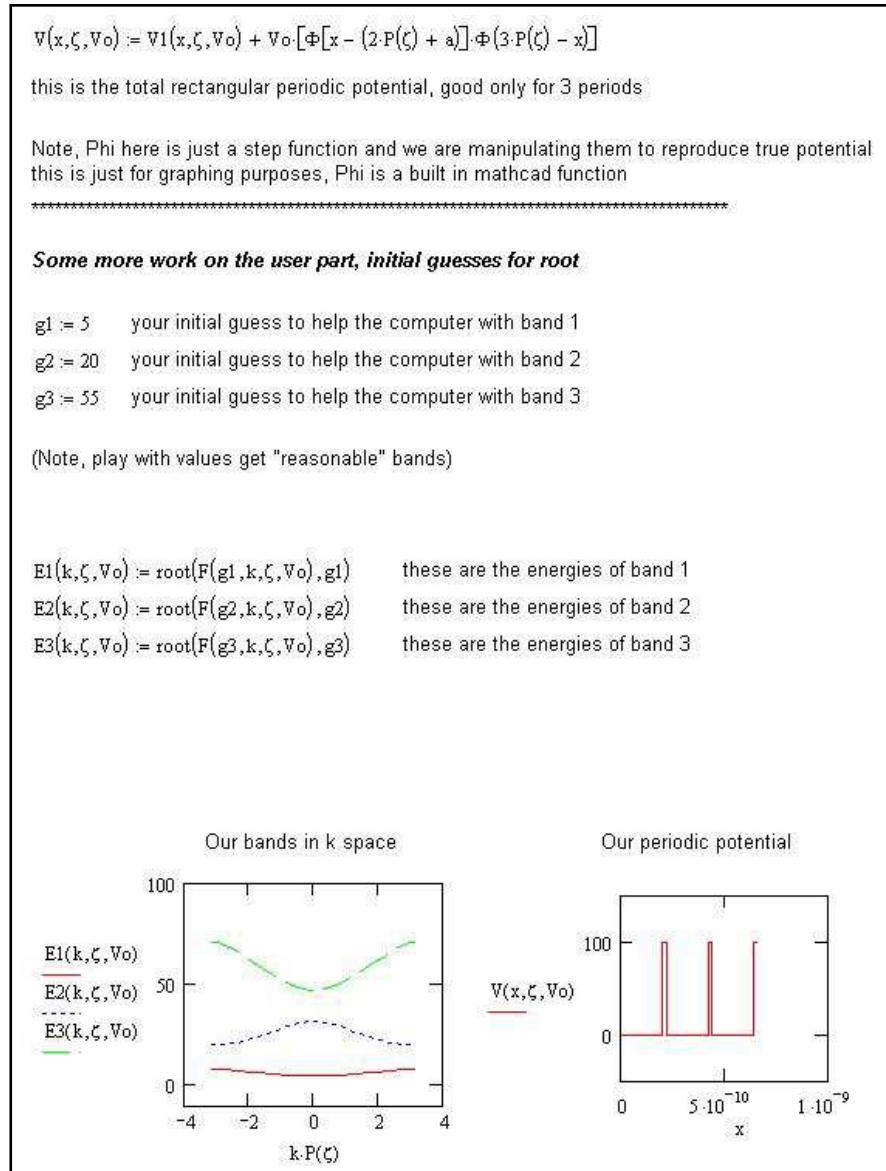


Figure 15.4: Mathcad numerical solutions to the general Kronig Penney model continued

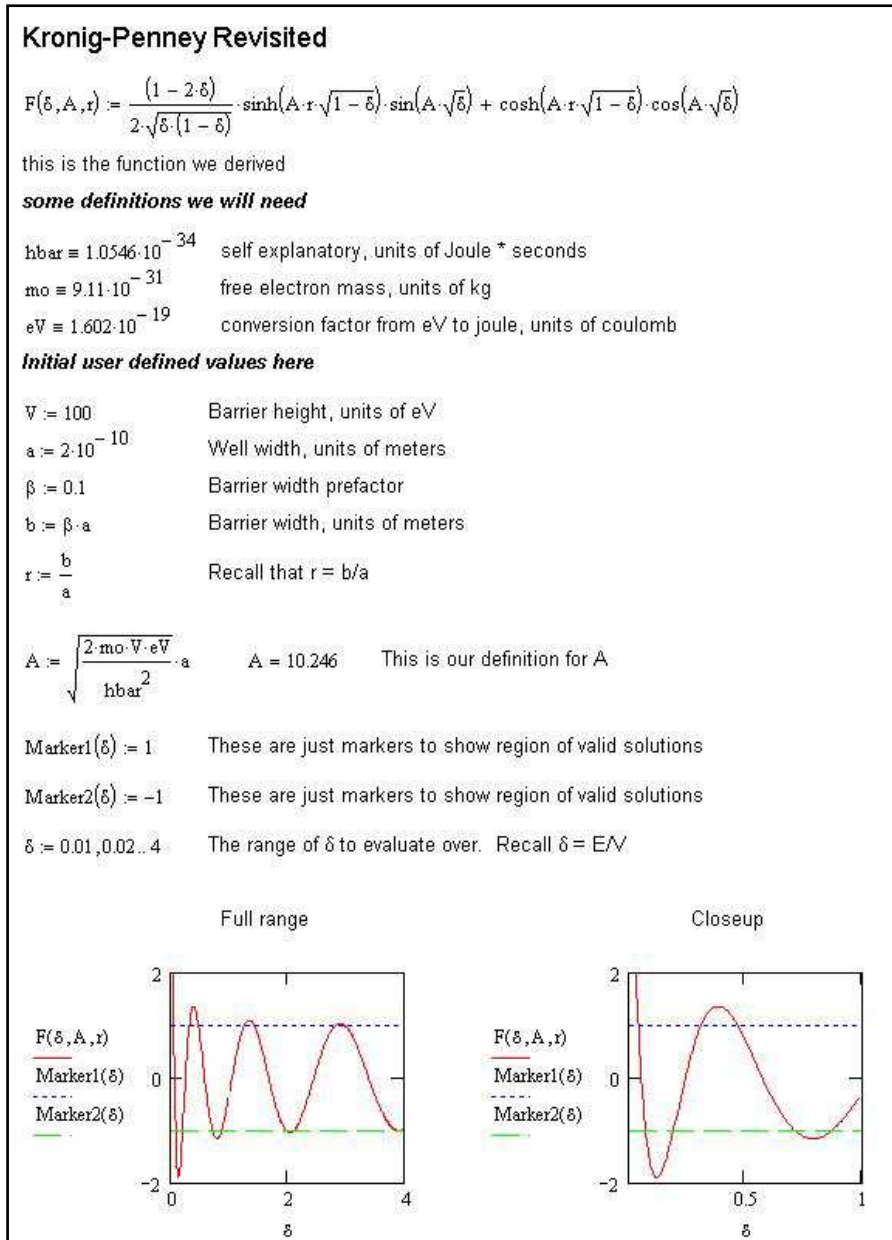


Figure 15.5: Mathcad numerical solutions to approximations of the Kronig Penney model

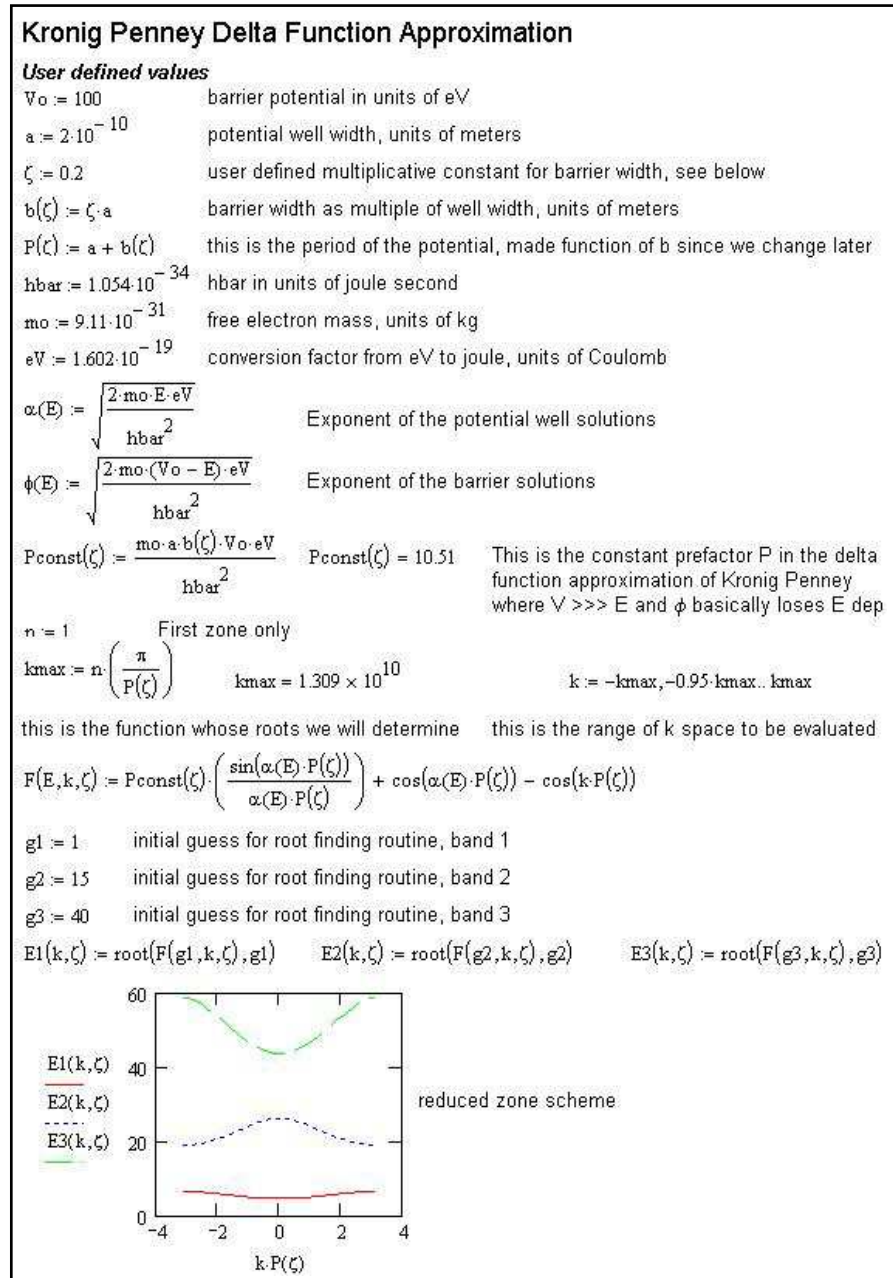


Figure 15.6: Mathcad numerical solutions to the delta function Kronig Penney model

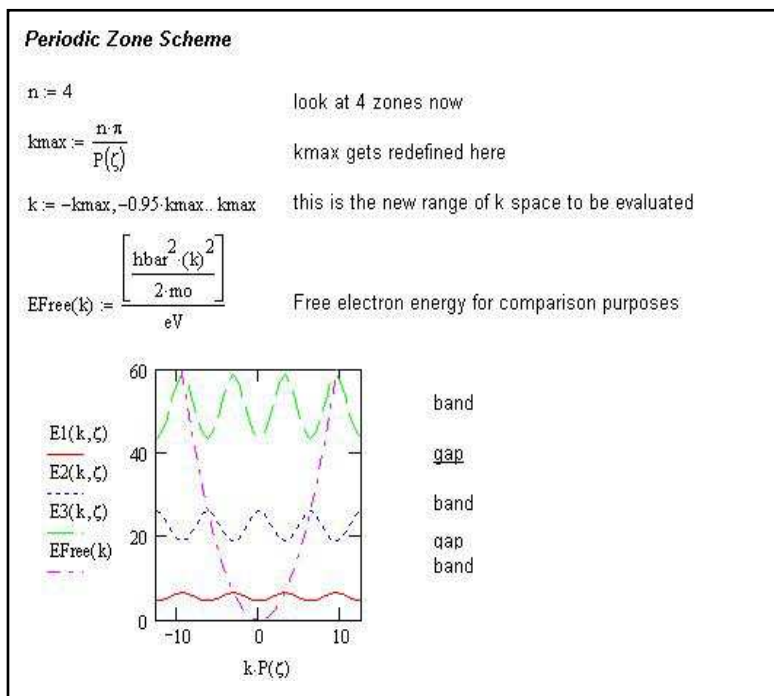


Figure 15.7: Mathcad numerical solutions to the delta function Kronig Penney model continued

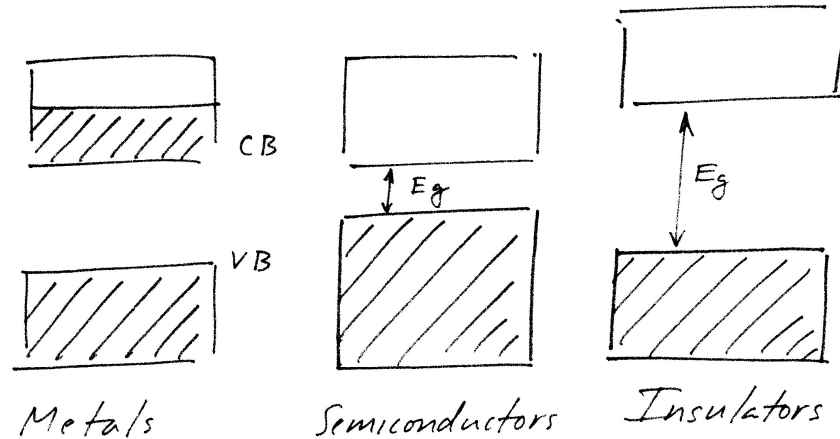


Figure 15.8: Cartoon of the differences between metals, semiconductors and insulators.

bandgap of $E_g = 1.17eV$? How about for a semiconductor with a bandgap of $E_g = 2.5eV$?

4. What temperature in K would you need to promote an electron from the valence band to the conduction band of an insulator with a gap of $E_g = 5eV$?

Examples from the literature

Here are selected papers from the literature dealing with one area of nanoscience or nanotechnology which is the creation of artificial solids with tailored properties. Crudely one version of this is to use quantum dots as artificial atoms. When arranged in a periodic manner (basically a crystal) the overlap of the electron wavefunctions will create user defined optical and electrical properties as well as the formation of artificial bands as we just described. These go by the name minibands and have already been achieved in the case of arrays of stacked quantum wells.

The papers below are listed in no particular order

1. “Architectonic quantum dot solids”
G. Markovich, C. P. Collier, S. E. Henrichs, F. Remacle, R. D. Levine,

- J. R. Heath
Accounts of Chemical Research, 32, 415 (1999).
2. "Nanocrystal superlattices"
C. P. Collier, T. Vossmeier, J. R. Heath
Annual Review of Physical Chemistry, 49, 371 (1998).
 3. "Highly oriented molecular Ag nanocrystal arrays"
S. A. Harfenist, Z. L. Wang, M. M. Alvarez, I. Vezmar, R. L. Whetten
Journal of Physical Chemistry, 100, 13904 (1996).
 4. "Synthesis of a quantum dot superlattice using molecularly linked metal clusters"
R. G. Osifchin, W. J. Mahoney, J. D. Bielefeld, R. P. Andres, J. I. Henderson, C. P. Kubiak
Superlattices and Microstructures, 18, 283 (1995).
 5. "Architecture with designer atoms: Simple theoretical consideration"
F. Remacle and R. D. Levine
Proceedings of the National Academy of Science, 97, 553 (2000).
 6. "Preparation and characterisation of silver quantum dot superlattice using self-assembled monolayers of pentanedithiol"
S. Pethkar, M. Aslam, I. S. Mulla, P. Ganeshan, K. Vijayamohan
Journal of Materials Chemistry, 11, 1710 (2001).
 7. "Miniband formation in a quantum dot crystal"
O. L. Lazarenkova and A. A. Balandin
Journal of Applied Physics, 89, 5509 (2001).
 8. "Reversible tuning of silver quantum dot monolayers through the metal-insulator transition"
C. P. Collier, R. J. Saykally, J. J. Shiang, S. E. Henrichs, J. R. Heath
Science, 277, 1978 (1997).

9. "Optical properties of thin films of Au@SiO₂ Particles"
T. Ung, L. M. Liz-Marzan, P. Mulvaney
Journal of Physical Chemistry B, 105, 3441 (2001).

Chapter 16

K · P Approximation (Pronounced k dot p)

Start with the Schrödinger equation

$$H\Psi = \epsilon\Psi \quad (16.1)$$

where

$$H = -\frac{\hbar^2\nabla^2}{2m} + V \quad (16.2)$$

Assume the Bloch form of the wavefunction Ψ

$$\Psi = u(r)e^{ikr} \quad (16.3)$$

Replace this into the above Schrödinger equation.

$$\begin{aligned} \left(-\frac{\hbar^2\nabla^2}{2m} + V\right) u(r)e^{ikr} &= \epsilon u(r)e^{ikr} \\ -\frac{\hbar^2\nabla^2}{2m} u(r)e^{ikr} + Vu(r)e^{ikr} &= \epsilon u(r)e^{ikr} \end{aligned}$$

where

$$\begin{aligned}
 \nabla^2(ue^{ikr}) &= \nabla \nabla u e^{ikr} \\
 &= \nabla \left(u i k e^{ikr} + e^{ikr} \nabla u \right) \\
 &= \nabla \left(i k u e^{ikr} + \nabla u e^{ikr} \right) \\
 &= \nabla (i k u + \nabla u) e^{ikr} \\
 &= (i k u + \nabla u) i k e^{ikr} + e^{ikr} (i k \nabla u + \nabla^2 u) \\
 &= -k^2 u e^{ikr} + i k \nabla u e^{ikr} + i k \nabla u e^{ikr} + \nabla^2 u e^{ikr} \\
 &= \nabla^2 u e^{ikr} - k^2 u e^{ikr} + 2 i k \nabla u e^{ikr} \\
 &= (\nabla^2 u - k^2 u + 2 i k \nabla u) e^{ikr}
 \end{aligned}$$

So

$$\begin{aligned}
 -\frac{\hbar^2 \nabla^2}{2m} u e^{ikr} + V u e^{ikr} &= \epsilon u e^{ikr} \\
 -\frac{\hbar^2}{2m} [\nabla^2 u - k^2 u + 2 i k \nabla u] e^{ikr} + V u e^{ikr} &= \epsilon u e^{ikr}
 \end{aligned}$$

Divide out e^{ikr}

$$\begin{aligned}
 -\frac{\hbar^2}{2m} [\nabla^2 u + 2 i k \nabla u - k^2 u] + V u &= \epsilon u \\
 -\frac{\hbar^2 \nabla^2 u}{2m} - \frac{i \hbar^2 k \nabla u}{m} + \frac{\hbar^2 k^2 u}{2m} + V u &= \epsilon u \\
 -\frac{\hbar^2 \nabla^2 u}{2m} - \frac{i \hbar^2 k \nabla u}{m} + V u &= \left(\epsilon - \frac{\hbar^2 k^2}{2m} \right) u \\
 \left(-\frac{\hbar^2 \nabla^2}{2m} + V \right) u - \frac{i \hbar^2 k \nabla u}{m} &= \left(\epsilon - \frac{\hbar^2 k^2}{2m} \right) u
 \end{aligned}$$

Recall now that $p = -i \hbar \nabla$ (Angular momentum operator) so that

$$\left(-\frac{\hbar^2 \nabla^2}{2m} + V \right) u + \left(\frac{\hbar}{m} k p \right) u = \left(\epsilon - \frac{\hbar^2 k^2}{2m} \right) u$$

Also recall that $p^2 = -\hbar^2 \nabla^2$ leading to

$$\left(\frac{p^2}{2m} + V \right) u + \frac{\hbar(k \cdot p)u}{m} = \left(\epsilon - \frac{\hbar^2 k^2}{2m} \right) u$$

where $H_0 = \left(\frac{p^2}{2m} + V\right)$ and $H_1 = \frac{\hbar(k \cdot p)}{m}$. This means that the Schrödinger equation becomes

$$(H_0 + H_1)u = \lambda u \quad (16.4)$$

where $\lambda = \left(\epsilon - \frac{\hbar^2 k^2}{2m}\right)$. So in general, to start solving this we create the Hamiltonian

$$\langle u_1 | (H_0 + H_1) | u_2 \rangle = \lambda_{12} \langle u_1 | u_2 \rangle \quad (16.5)$$

where the Hamiltonian H is $H = \langle u_1 | (H_0 + H_1) | u_2 \rangle$. Then we solve the secular determinant

$$|H - \lambda| = 0 \quad (16.6)$$

To summarize, the Schrodinger equation becomes

$$\left(\frac{p^2}{2m} + V\right)u + \left(\frac{\hbar}{m}k \cdot p\right)u = \lambda u \quad (16.7)$$

where we explicitly consider the $k \cdot p$ perturbation to the eigenenergies.

Simple 2 band model

Let's say we have a unit cell function u_s that is for the conduction band (CB) and has s-like symmetry. Also we have a unit cell function u_p that is for the valence band (VB) and has p-like symmetry. We need to solve

$$\langle u_1 | H_0 + H_1 | u_2 \rangle \quad (16.8)$$

where $H_0 = \frac{p^2}{2m} + V$ (We know the solutions of these) and $H_1 = \frac{\hbar}{m}k \cdot p$. We ignore spin orbit interactions here and don't even consider any spin for simplicity. So we basically have the following 2 by 2 Hamiltonian.

	$ u_s \rangle$	$ u_p \rangle$
$\langle u_s $	a	b
$\langle u_p $	c	d

Let's do the diagonals first.

Array element a

$$\begin{aligned}
a &= \langle u_s | H_0 + H_1 | u_s \rangle \\
&= \langle u_s | H_0 | u_s \rangle + \langle u_s | H_1 | u_s \rangle \\
&= \epsilon_s + \langle u_s | H_1 | u_s \rangle
\end{aligned}$$

and $\langle u_s | H_1 | u_s \rangle = \langle u_s | \frac{\hbar}{m} (k \cdot p) | u_s \rangle$ where $p = -i\hbar\nabla$ (momentum operator). Since ∇ flips the parity of the unit cell function $(k \cdot p) | u_s \rangle$ (even) = (odd). Therefore

$$\frac{\hbar}{m} \langle u_s | k \cdot p | u_s \rangle = 0$$

such that in whole

$$a = \epsilon_s \quad (16.9)$$

Array element d

$$\begin{aligned}
d &= \langle u_p | H_0 + H_1 | u_p \rangle \\
&= \langle u_p | H_0 | u_p \rangle + \langle u_p | H_1 | u_p \rangle \\
&= \epsilon_v + \langle u_p | \frac{\hbar}{m} k \cdot p | u_p \rangle
\end{aligned}$$

Recall that $\frac{\hbar}{m} k \cdot p | u_p \rangle = (\text{even})$ and that $u_p = (\text{odd})$. Hence

$$\langle u_p | \frac{\hbar}{m} k \cdot p | u_p \rangle = 0$$

In whole then

$$d = \epsilon_v \quad (16.10)$$

Next, let's do the off diagonals

Array element b

$$\begin{aligned}
b &= \langle u_s | H_0 + H_1 | u_p \rangle \\
&= \langle u_s | H_0 | u_p \rangle + \langle u_s | H_1 | u_p \rangle \\
&= 0 + \langle u_s | H_1 | u_p \rangle \\
&= \langle u_s | \frac{\hbar}{m} k \cdot p | u_p \rangle \neq 0 = \frac{\hbar}{m} KP
\end{aligned}$$

where we define $KP \equiv \langle u_s | k \cdot p | u_p \rangle$. Therefore

$$b = \frac{\hbar}{m} KP \quad (16.11)$$

Array element c

We can work out c but since the Hamiltonian is Hermitian c will be the same as b . Therefore

$$c = \frac{\hbar}{m}KP \quad (16.12)$$

To summarize,

$$H = \begin{pmatrix} \epsilon_s & \frac{\hbar}{m}KP \\ \frac{\hbar}{m}KP & \epsilon_v \end{pmatrix}$$

Now we will solve for the secular determinant $|H - \lambda| = 0$

$$\begin{vmatrix} \epsilon_s & \frac{\hbar}{m}KP \\ \frac{\hbar}{m}KP & \epsilon_v \end{vmatrix} = 0$$

This leads to solving

$$(\epsilon_s - \lambda)(\epsilon_v - \lambda) - \frac{\hbar^2 k^2 p^2}{m^2} = 0$$

Define

$$\epsilon_p \equiv \frac{2p^2}{m} \quad (16.13)$$

leading to

$$(\epsilon_s - \lambda)(\epsilon_v - \lambda) - \frac{\hbar^2 k^2}{2m} \left(\frac{2p^2}{m} \right) = 0$$

$$(\epsilon_s - \lambda)(\epsilon_v - \lambda) - \frac{\hbar^2 k^2}{2m} \epsilon_p = 0$$

$$\epsilon_s \epsilon_v - \lambda \epsilon_s - \lambda \epsilon_v + \lambda^2 - \frac{\hbar^2 k^2}{2m} \epsilon_p = 0$$

$$\lambda^2 - \lambda(\epsilon_s + \epsilon_v) - \frac{\hbar^2 k^2}{2m} \epsilon_p = 0$$

The last equation is everyone's favorite quadratic equation. The solutions are generally speaking

$$\begin{aligned} \lambda &= \frac{(\epsilon_s + \epsilon_v) \pm \sqrt{(\epsilon_s + \epsilon_v)^2 - 4(1)[\epsilon_s \epsilon_v - \frac{\hbar^2 k^2}{2m} \epsilon_p]}}{2} \\ &= \frac{(\epsilon_s + \epsilon_v)}{2} \pm \frac{1}{2} \sqrt{(\epsilon_s + \epsilon_v)^2 - 4\epsilon_s \epsilon_v + 4\frac{\hbar^2 k^2}{2m} \epsilon_p} \\ &= \frac{(\epsilon_s + \epsilon_v)}{2} \pm \frac{1}{2} \sqrt{(\epsilon_s - \epsilon_v)^2 + 4\frac{\hbar^2 k^2}{2m} \epsilon_p} \end{aligned}$$

Recall now that $\epsilon_g = (\epsilon_s - \epsilon_v)$ (gap energy) so that the solutions are:

$$\lambda = \frac{(\epsilon_s + \epsilon_v)}{2} \pm \frac{1}{2} \sqrt{\epsilon_g^2 + 4 \frac{\hbar^2 k^2}{2m} \epsilon_p} \quad (16.14)$$

and since $\lambda = \epsilon - \frac{\hbar^2 k^2}{2m}$

$$\lambda = \frac{(\epsilon_s + \epsilon_v)}{2} + \frac{\hbar^2 k^2}{2m} \pm \frac{1}{2} \sqrt{\epsilon_g^2 + 4 \frac{\hbar^2 k^2}{2m} \epsilon_p} \quad (16.15)$$

Conduction band solution

$$\epsilon_{cb} = \frac{(\epsilon_s + \epsilon_v)}{2} + \frac{\hbar^2 k^2}{2m} + \frac{1}{2} \sqrt{\epsilon_g^2 + 4 \frac{\hbar^2 k^2}{2m} \epsilon_p} \quad (16.16)$$

Valence band solution

$$\epsilon_{vb} = \frac{(\epsilon_s + \epsilon_v)}{2} + \frac{\hbar^2 k^2}{2m} - \frac{1}{2} \sqrt{\epsilon_g^2 + 4 \frac{\hbar^2 k^2}{2m} \epsilon_p} \quad (16.17)$$

You can already tell that the solutions are non-parabolic in k . We can make some further insightful simplifications using the binomial expansion and keeping only the first two terms $\sqrt{a+b} \simeq a^{\frac{1}{2}} + \frac{1}{2} a^{-\frac{1}{2}} b + \dots$ when $|b^2| < a^2$.

So $\sqrt{\epsilon_g^2 + 4 \frac{\hbar^2 k^2}{2m} \epsilon_p} \simeq \epsilon_g + \frac{1}{2} \frac{4 \frac{\hbar^2 k^2}{2m} \epsilon_p}{\epsilon_g}$ (assume k is small). Therefore

$$\sqrt{\epsilon_g^2 + 4 \frac{\hbar^2 k^2}{2m} \epsilon_p} \simeq \epsilon_g + \frac{2\epsilon_p}{\epsilon_g} \left(\frac{\hbar^2 k^2}{2m} \right) \quad (16.18)$$

Leading to

$$\begin{aligned} \epsilon_{cb} &= \frac{(\epsilon_s + \epsilon_v)}{2} + \frac{\hbar^2 k^2}{2m} + \frac{1}{2} \left[\epsilon_g + \frac{2\epsilon_p}{\epsilon_g} \left(\frac{\hbar^2 k^2}{2m} \right) \right] \\ &= \frac{(\epsilon_s + \epsilon_v)}{2} + \frac{(\epsilon_s - \epsilon_v)}{2} + \frac{\hbar^2 k^2}{2m} + \left(\frac{\epsilon_p}{\epsilon_g} \right) \left(\frac{\hbar^2 k^2}{2m} \right) \\ &= \epsilon_s + \frac{\hbar^2 k^2}{2m} \left(1 + \frac{\epsilon_p}{\epsilon_g} \right) \end{aligned} \quad (16.19)$$

Likewise for the valence band

$$\begin{aligned}
\epsilon_{vb} &= \frac{(\epsilon_s + \epsilon_v)}{2} + \frac{\hbar^2 k^2}{2m} - \frac{1}{2} \left[\epsilon_g + \frac{2\epsilon_p}{\epsilon_g} \left(\frac{\hbar^2 k^2}{2m} \right) \right] \\
&= \frac{(\epsilon_s + \epsilon_v)}{2} - \frac{(\epsilon_s - \epsilon_v)}{2} + \frac{\hbar^2 k^2}{2m} - \frac{\epsilon_p}{\epsilon_g} \left(\frac{\hbar^2 k^2}{2m} \right) \\
&= \epsilon_v + \frac{\hbar^2 k^2}{2m} \left(1 - \frac{\epsilon_p}{\epsilon_g} \right)
\end{aligned} \tag{16.20}$$

Now we can go after the effective masses. For the conduction band

$$\frac{\hbar^2 k^2}{2m_{eff}} = \frac{\hbar^2 k^2}{2m} \left(1 + \frac{\epsilon_p}{\epsilon_g} \right)$$

Therefore

$$\begin{aligned}
\frac{1}{m_{eff}} &= \frac{1}{m} \left(1 + \frac{\epsilon_p}{\epsilon_g} \right) \\
m_{eff} &= m \left(\frac{1}{1 + \frac{\epsilon_p}{\epsilon_g}} \right)
\end{aligned}$$

Leading to the conduction band effective mass

$$\left(\frac{m_{eff}}{m} \right) = \frac{1}{1 + \frac{\epsilon_p}{\epsilon_g}} \tag{16.21}$$

Likewise, for the valence band

$$\frac{\hbar^2 k^2}{2m_{eff}} = \frac{\hbar^2 k^2}{2m} \left(1 - \frac{\epsilon_p}{\epsilon_g} \right)$$

Therefore

$$\begin{aligned}
\frac{1}{m_{eff}} &= \frac{1}{m} \left(1 - \frac{\epsilon_p}{\epsilon_g} \right) \\
m_{eff} &= m \left(\frac{1}{1 - \frac{\epsilon_p}{\epsilon_g}} \right)
\end{aligned}$$

Leading to the valence band effective mass

$$\left(\frac{m_{eff}}{m} \right) = \frac{1}{1 - \frac{\epsilon_p}{\epsilon_g}} \tag{16.22}$$

Now since $\epsilon_p \simeq 21$ eV typically and also does not vary much between different semiconductors one can see that the conduction band effective mass is (a) small and (b) is positive. Likewise, one can see that the valence band effective mass is (a) small and (b) negative. Although the 2 band model is a large simplification of reality, generally speaking, $m_e < m_h$.

Alternative view of 2 band model

Sometimes you will see people conveniently redefine the energy origin to simplify the problem at hand. Doing this was not necessary for our two band model because this 2 by 2 was easy to solve. However, for 4 by 4 or higher band model it becomes a great convenience. We illustrate this with the 2 by 2.

Starting with the Hamiltonian

$$H = \begin{pmatrix} \epsilon_g & \frac{\hbar}{m} KP \\ \frac{\hbar}{m} KP & 0 \end{pmatrix}$$

we again need to solve the secular determinant

$$\begin{aligned} |H - \lambda| &= 0 \\ \begin{vmatrix} \epsilon_g - \lambda & \frac{\hbar}{m} KP \\ \frac{\hbar}{m} KP & -\lambda \end{vmatrix} &= 0 \end{aligned}$$

leading to

$$\begin{aligned} (\epsilon_g - \lambda)(-\lambda) - \frac{\hbar^2 k^2 p^2}{m^2} &= 0 \\ -\lambda\epsilon_g + \lambda^2 - \frac{\hbar^2 k^2 p^2}{m^2} &= 0 \\ \lambda^2 - \epsilon_g \lambda - \frac{\hbar^2 k^2 p^2}{m^2} &= 0 \end{aligned}$$

As before let $\epsilon_p = \frac{2p^2}{m}$. This leads to

$$\lambda^2 - \epsilon_g \lambda - \frac{\hbar^2 k^2}{2m} \epsilon_p = 0 \quad (16.23)$$

We could do this exactly by solving the quadratic equation. However, we can do some approximations at this point to solve the problem quicker. For $k \rightarrow 0$ (small) we see that the equation reduces to

$$\lambda^2 - \epsilon_g \lambda = 0$$

or

$$\lambda(\lambda - \epsilon_g) = 0$$

The two $k = 0$ limiting roots are ($\lambda = 0$) and ($\lambda = \epsilon_g$) so we will assume that the actual roots are of this order and will make approximate solutions to

$$\begin{aligned}\lambda^2 - \epsilon_g \lambda - \frac{\hbar^2 k^2}{2m} \epsilon_p &= 0 \\ \lambda(\lambda - \epsilon_g) - \frac{\hbar^2 k^2}{2m} \epsilon_p &= 0\end{aligned}$$

Solve for λ close to zero

$$\begin{aligned}\lambda(-\epsilon_g) - \frac{\hbar^2 k^2}{2m} \epsilon_p &= 0 \\ \lambda \epsilon_g + \frac{\hbar^2 k^2}{2m} \epsilon_p &= 0\end{aligned}$$

leading to

$$\lambda = -\frac{\hbar^2 k^2}{2m} \left(\frac{\epsilon_p}{\epsilon_g} \right) \quad (16.24)$$

where $\lambda = \epsilon - \frac{\hbar^2 k^2}{2m}$ leading to

$$\begin{aligned}\epsilon_{vb} &= \frac{\hbar^2 k^2}{2m} - \frac{\hbar^2 k^2}{2m} \left(\frac{\epsilon_p}{\epsilon_g} \right) \\ &= \frac{\hbar^2 k^2}{2m} \left(1 - \frac{\epsilon_p}{\epsilon_g} \right)\end{aligned} \quad (16.25)$$

This is the valence band energy of the 2 band model where one notes that the origin is now at zero.

Effective mass of the VB

$$\begin{aligned}\frac{\hbar^2 k^2}{2m_{eff}} &= \frac{\hbar^2 k^2}{2m} \left(1 - \frac{\epsilon_p}{\epsilon_g} \right) \\ m_{eff} &= m \frac{1}{1 - \frac{\epsilon_p}{\epsilon_g}}\end{aligned}$$

leading to

$$\frac{m_{eff}}{m} = \frac{1}{1 - \frac{\epsilon_p}{\epsilon_g}} \quad (16.26)$$

This is a negative effective mass since $\epsilon_p \simeq 21$ eV and $\epsilon_p \gg \epsilon_g$ typically. Now we will do the conduction band energy and effective mass in the next section.

Solve for λ close to ϵ_g

$$\lambda(\lambda - \epsilon_g) - \frac{\hbar^2 k^2}{2m} \epsilon_p = 0$$

becomes

$$\epsilon_g(\lambda - \epsilon_g) - \frac{\hbar^2 k^2}{2m} \epsilon_p = 0$$

leading to

$$\lambda = \epsilon_g + \frac{\hbar^2 k^2}{2m} \left(\frac{\epsilon_p}{\epsilon_g} \right) \quad (16.27)$$

Recall that $\lambda = \epsilon - \frac{\hbar^2 k^2}{2m}$ leading to

$$\begin{aligned} \epsilon_{cb} &= \epsilon_g + \frac{\hbar^2 k^2}{2m} + \frac{\hbar^2 k^2}{2m} \left(\frac{\epsilon_p}{\epsilon_g} \right) \\ &= \epsilon_g + \frac{\hbar^2 k^2}{2m} \left(1 + \frac{\epsilon_p}{\epsilon_g} \right) \end{aligned} \quad (16.28)$$

This is the conduction band energies of the 2 band model. Note that the origin is now at ϵ_g rather than at ϵ_f (Fermi level) of an intrinsic semiconductor. Now we find the effective mass.

Effective mass of the CB

$$\begin{aligned} \frac{\hbar^2 k^2}{2m_{eff}} &= \frac{\hbar^2 k^2}{2m} \left(1 + \frac{\epsilon_p}{\epsilon_g} \right) \\ m_{eff} &= m \left(\frac{1}{1 + \frac{\epsilon_p}{\epsilon_g}} \right) \end{aligned}$$

leading to

$$\frac{m_{eff}}{m} = \frac{1}{1 + \frac{\epsilon_p}{\epsilon_g}} \quad (16.29)$$

This is a positive effective mass since $\epsilon_p \simeq 21$ eV for most semiconductors and $\epsilon_p \gg \epsilon_g$.

Last note on effective masses

Since

$$\frac{m_{eff}}{m} \simeq \frac{1}{1 + \frac{\epsilon_p}{\epsilon_g}} \text{ CB}$$

$$\frac{m_{eff}}{m} \simeq \frac{1}{1 - \frac{\epsilon_p}{\epsilon_g}} \text{ VB}$$

You can see that the effective mass will scale with the band gap of the semiconductor. Small effective masses correlates with a small band gap. Large effective masses correlate with a large band gap.

A four band model excluding spin

Assume the conduction band is s-like. Assume that the valence band is p-like. Choose for convenience $k = (0, 0, k_z)$ so that $k \cdot p = k_z p_z$. Ok, now we know from before that we need to first generate the Hamiltonian

$$\langle u_1 | H_0 + H_1 | u_2 \rangle \quad (16.30)$$

where

$$H_0 = \frac{p^2}{2m} + V = -\frac{\hbar^2 \nabla^2}{2m} + V$$

$$H_1 = \frac{\hbar}{m} k \cdot p$$

and where $p = -i\hbar\nabla$. This leads to the following Hamiltonian

	$ u_s\rangle$	$ u_x\rangle$	$ u_y\rangle$	$ u_z\rangle$
$\langle u_s $	a	b	c	d
$\langle u_x $		e	f	g
$\langle u_y $			h	i
$\langle u_z $				j

where we have indicated the important array elements with letters. The others follow because the Hamiltonian is Hermitian. Let's do the diagonals first.

Array element a

$$\begin{aligned}
\langle u_s | H_0 + H_1 | u_s \rangle &= \langle u_s | H_0 | u_s \rangle + \langle u_s | H_1 | u_s \rangle \\
&= \epsilon_s + \frac{\hbar}{m} \langle u_s | k_z p_z | u_s \rangle \\
&= \epsilon_s
\end{aligned}$$

because u_s has even symmetry in x , y and z . p_z acting on u_s changes the overall parity to odd such that the term dies.

Array element b

$$\begin{aligned}
\langle u_x | H_0 + H_1 | u_x \rangle &= \langle u_x | H_0 | u_x \rangle + \langle u_x | H_1 | u_x \rangle \\
&= \epsilon_v + \frac{\hbar}{m} \langle u_x | k_z p_z | u_x \rangle \\
&= \epsilon_v
\end{aligned}$$

Here u_x is odd in x , even in y and even in z . p_z acting on u_x changes its parity to odd in x , even in y , odd in z . Since $\langle u_x$ is odd in x , even in y and even in z , the z components cancel.

Array element c

$$\begin{aligned}
\langle u_y | H_0 + H_1 | u_y \rangle &= \langle u_y | H_0 | u_y \rangle + \langle u_y | H_1 | u_y \rangle \\
&= \epsilon_v + \frac{\hbar}{m} \langle u_y | k_z p_z | u_y \rangle \\
&= \epsilon_v
\end{aligned}$$

Here u_y is even in x , odd in y , even in z . p_z acting on u_y changes its parity to even in x , odd in y and odd in z . Since $\langle u_y |$ is even in x , odd in y and even in z , the z components cancel.

Array element d

$$\begin{aligned}
\langle u_z | H_0 + H_1 | u_z \rangle &= \langle u_z | H_0 | u_z \rangle + \langle u_z | H_1 | u_z \rangle \\
&= \epsilon_v + \frac{\hbar}{m} \langle u_z | H_1 | u_z \rangle \\
&= \epsilon_v
\end{aligned}$$

Here u_z is even in x , even in y and odd in z . p_z acting on u_z changes its parity to even in x , even in y and even in z . Since $\langle u_z |$ is even in x , even in y , and odd in z , the z components cancel.

Summary

$$\begin{aligned}
 \langle u_s | H_0 + H_1 | u_s \rangle &= \epsilon_s \\
 \langle u_x | H_0 + H_1 | u_x \rangle &= \epsilon_v \\
 \langle u_y | H_0 + H_1 | u_y \rangle &= \epsilon_v \\
 \langle u_z | H_0 + H_1 | u_z \rangle &= \epsilon_v
 \end{aligned}$$

Now for the off diagonals

Array element e

$$\begin{aligned}
 \langle u_s | H_1 | u_x \rangle &= \frac{\hbar}{m} \langle u_s | k_z p_z | u_x \rangle \\
 &= 0
 \end{aligned}$$

Here u_x is odd in x, even in y, even in z. p_z acting on u_x changes its parity to odd in x, even in y, and odd in z. Since $\langle u_s |$ is even in x, even in y and even in z, the x and z components cancel.

Array element f

$$\begin{aligned}
 \langle u_s | H_1 | u_y \rangle &= \frac{\hbar}{m} \langle u_s | k_z p_z | u_y \rangle \\
 &= 0
 \end{aligned}$$

Here u_y is even in x, odd in y and even in z. p_z acting on u_y changes its parity to even in x, odd in y and odd in z. Since $\langle u_s$ is even in x, even in y and even in z the y and z components cancel.

Array element g

$$\begin{aligned}
 \langle u_s | H_1 | u_z \rangle &= \frac{\hbar}{m} \langle u_s | k_z p_z | u_z \rangle \\
 &= \frac{\hbar}{m} KP
 \end{aligned}$$

Here u_z is even in x, even in y and odd in z. p_z acting on u_z is even in x, even in y and even in z. Since $\langle u_s |$ is even in x, even in y and even in z none of the terms cancel.

Array element h

$$\begin{aligned}\langle u_x | H_1 | u_y \rangle &= \frac{\hbar}{m} \langle u_x | k_z p_z | u_y \rangle \\ &= 0\end{aligned}$$

Here u_y is even in x, odd in y, and even in z. p_z acting on u_y changes its parity to even in x, odd in y and odd in z. Since $\langle u_x |$ is odd in x, even in y and even in z everything cancels out.

Array element i

$$\begin{aligned}\langle u_x | H_1 | u_z \rangle &= \frac{\hbar}{m} \langle u_x | k_z p_z | u_z \rangle \\ &= 0\end{aligned}$$

Here u_z is even in x, even in y, and odd in z. p_z acting on u_z changes its parity to even in x, even in y and even in z. Since $\langle u_x |$ is odd in x, even in y, and even in z the x components cancel.

Array element j

$$\begin{aligned}\langle u_y | H_1 | u_z \rangle &= \frac{\hbar}{m} \langle u_y | k_z p_z | u_z \rangle \\ &= 0\end{aligned}$$

Here u_z is even in x, even in y, and odd in z. p_z acting on u_z changes its parity to even in x, even in y and even in z. Since $\langle u_y |$ is even in x, odd in y and even in z, the y components cancel.

Summary

$$\begin{aligned}\langle u_s | H_1 | u_x \rangle &= 0 \\ \langle u_s | H_1 | u_y \rangle &= 0 \\ \langle u_s | H_1 | u_z \rangle &= \frac{\hbar}{m} KP \\ \langle u_x | H_1 | u_y \rangle &= 0 \\ \langle u_x | H_1 | u_z \rangle &= 0 \\ \langle u_y | H_1 | u_z \rangle &= 0\end{aligned}$$

The desired Hamiltonian is then

$$H = \begin{pmatrix} \epsilon_s & 0 & 0 & \frac{\hbar}{m}KP \\ 0 & \epsilon_v & 0 & 0 \\ 0 & 0 & \epsilon_v & 0 \\ \frac{\hbar}{m}KP & 0 & 0 & \epsilon_v \end{pmatrix}$$

Solve the secular determinant now.

$$\begin{vmatrix} \epsilon_s - \lambda & 0 & 0 & \frac{\hbar}{m}KP \\ 0 & \epsilon_v - \lambda & 0 & 0 \\ 0 & 0 & \epsilon_v - \lambda & 0 \\ \frac{\hbar}{m}KP & 0 & 0 & \epsilon_v - \lambda \end{vmatrix} = 0$$

There are many ways to do this. One way looks like

$$(\epsilon_s - \lambda) \begin{vmatrix} \epsilon_v - \lambda & 0 & 0 \\ 0 & \epsilon_v - \lambda & 0 \\ 0 & 0 & \epsilon_v - \lambda \end{vmatrix} - \frac{\hbar}{m}KP \begin{vmatrix} 0 & 0 & \frac{\hbar}{m}KP \\ \epsilon_v - \lambda & 0 & 0 \\ 0 & \epsilon_v - \lambda & 0 \end{vmatrix} = 0$$

leading to

$$(\epsilon_s - \lambda)(\epsilon_v - \lambda)^3 + \frac{\hbar}{m}KP(\epsilon_v - \lambda) \begin{vmatrix} 0 & \frac{\hbar}{m}KP \\ \epsilon_v - \lambda & 0 \end{vmatrix} = 0$$

$$(\epsilon_s - \lambda)(\epsilon_v - \lambda)^3 - \frac{\hbar}{m}KP(\epsilon_v - \lambda)^2 \frac{\hbar}{m}KP = 0$$

$$(\epsilon_s - \lambda)(\epsilon_v - \lambda)^3 - \frac{\hbar^2 k^2 p^2}{m^2} (\epsilon_v - \lambda)^2 = 0$$

$$(\epsilon_v - \lambda)^2 \left[(\epsilon_s - \lambda)(\epsilon_v - \lambda) - \frac{\hbar^2 k^2 p^2}{m^2} \right] = 0 \text{ where } \epsilon_p = \frac{2p^2}{m}$$

$$(\epsilon_v - \lambda)^2 \left[(\epsilon_s - \lambda)(\epsilon_v - \lambda) - \frac{\hbar^2 k^2}{2m} \left(\frac{2p^2}{m} \right) \right] = 0$$

$$(\epsilon_v - \lambda)^2 \left[(\epsilon_s - \lambda)(\epsilon_v - \lambda) - \frac{\hbar^2 k^2}{2m} \epsilon_p \right] = 0$$

There are two immediate solutions that are apparent in the first term, $\lambda_{1,2} = \epsilon_v$.

$$\epsilon_1 = \frac{\hbar^2 k^2}{2m} + \epsilon_v \quad (16.31)$$

$$\epsilon_2 = \frac{\hbar^2 k^2}{2m} + \epsilon_v \quad (16.32)$$

To get the other solutions we must solve

$$(\epsilon_s - \lambda)(\epsilon_v - \lambda) - \frac{\hbar^2 k^2}{2m} \epsilon_p = 0$$

You can do this exactly but let's be smart. One solution is of order ϵ_s . The other solution is of order ϵ_v .

Solution 1 (of order ϵ_s)

$$\begin{aligned} (\epsilon_s - \lambda)(\epsilon_v - \epsilon_s) - \frac{\hbar^2 k^2}{2m} \epsilon_p &= 0 \\ (\epsilon_s - \lambda) &= \frac{\hbar^2 k^2}{2m} \frac{\epsilon_p}{(\epsilon_v - \epsilon_s)} \text{ where } \epsilon_v - \epsilon_s = -\epsilon_g \\ (\epsilon_s - \lambda) &= -\frac{\hbar^2 k^2}{2m} \left(\frac{\epsilon_p}{\epsilon_g} \right) \end{aligned}$$

This results in

$$\begin{aligned} \lambda_{cb} &= \epsilon_s + \frac{\hbar^2 k^2}{2m} \left(\frac{\epsilon_p}{\epsilon_g} \right) \\ \epsilon_{cb} &= \epsilon_s + \frac{\hbar^2 k^2}{2m} \left(1 + \frac{\epsilon_p}{\epsilon_g} \right) \end{aligned} \quad (16.33)$$

since $\lambda = \epsilon - \frac{\hbar^2 k^2}{2m}$. This is the desired conduction band energy.

Conduction band effective mass

$$\begin{aligned} \frac{\hbar^2 k^2}{2m_{eff}} &= \frac{\hbar^2 k^2}{2m} \left(1 + \frac{\epsilon_p}{\epsilon_g} \right) \\ \left(\frac{m_{eff}}{m} \right) &= \frac{1}{1 + \frac{\epsilon_p}{\epsilon_g}} \end{aligned} \quad (16.34)$$

Solution 2 (of order ϵ_v)

$$\begin{aligned} (\epsilon_s - \epsilon_v)(\epsilon_v - \lambda) - \frac{\hbar^2 k^2}{2m} \epsilon_p &= 0 \\ (\epsilon_v - \lambda) &= \frac{\hbar^2 k^2}{2m} \frac{\epsilon_p}{\epsilon_s - \epsilon_v} \\ (\epsilon_v - \lambda) &= \frac{\hbar^2 k^2}{2m} \left(\frac{\epsilon_p}{\epsilon_g} \right) \end{aligned}$$

This results in

$$\begin{aligned}\lambda &= \epsilon_v - \frac{\hbar^2 k^2}{2m} \left(\frac{\epsilon_p}{\epsilon_g} \right) \\ \epsilon_{vb} &= \epsilon_v + \frac{\hbar^2 k^2}{2m} \left(1 - \frac{\epsilon_p}{\epsilon_g} \right)\end{aligned}\quad (16.35)$$

since $\lambda = \epsilon - \frac{\hbar^2 k^2}{2m}$. This is the desired valence band energy.

Valence band effective mass

$$\begin{aligned}\frac{\hbar^2 k^2}{2m_{eff}} &= \frac{\hbar^2 k^2}{2m} \left(1 - \frac{\epsilon_p}{\epsilon_g} \right) \\ \left(\frac{m_{eff}}{m} \right) &= \frac{1}{1 - \frac{\epsilon_p}{\epsilon_g}}\end{aligned}\quad (16.36)$$

Kane model, (8 x 8)

Given the basis

$$\begin{aligned}&|is \downarrow\rangle \\ &\left| \frac{(x - iy)}{\sqrt{2}} \uparrow \right\rangle \\ &|z \downarrow\rangle \\ &\left| \frac{(x + iy)}{\sqrt{2}} \uparrow \right\rangle \\ &|is \uparrow\rangle \\ &\left| \frac{(x + iy)}{\sqrt{2}} \downarrow \right\rangle \\ &|z \uparrow\rangle \\ &\left| \frac{(x - iy)}{\sqrt{2}} \downarrow \right\rangle\end{aligned}$$

the 8 by 8 interaction matrix is written as

$$\begin{pmatrix} H & 0 \\ 0 & H \end{pmatrix}$$

where

$$H = \begin{pmatrix} \epsilon_s & 0 & KP & 0 \\ 0 & \epsilon_p - \frac{\Delta}{3} & \sqrt{\frac{2}{3}}\Delta & 0 \\ kP & \frac{\sqrt{2}\Delta}{3} & \epsilon_p & 0 \\ 0 & 0 & 0 & \epsilon_p + \frac{\Delta}{3} \end{pmatrix}$$

and where the positive constant Δ is the spin orbit splitting of the valence band.

$$P = -\frac{i\hbar}{m} \langle s | p_z | z \rangle$$

$$\Delta = \frac{3\hbar i}{4m^2 c^2} \left\langle x \left| \frac{\partial V}{\partial x} - \frac{\partial V}{\partial y} \right| y \right\rangle$$

Note that the vector k is taken along the z direction for convenience.

To solve this, choose $\epsilon_p = -\frac{\Delta}{3}$ to simplify things

$$H = \begin{pmatrix} \epsilon_s & 0 & kP & 0 \\ 0 & -\frac{2}{3}\Delta & \frac{\sqrt{2}}{3}\Delta & 0 \\ kP & \frac{\sqrt{2}}{3}\Delta & -\frac{\Delta}{3} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Now solve the secular determinant

$$\begin{vmatrix} \epsilon_s - \lambda & 0 & kP & 0 \\ 0 & -\frac{2}{3}\Delta - \lambda & \frac{\sqrt{2}}{3}\Delta & 0 \\ kP & \frac{\sqrt{2}}{3}\Delta & -\frac{\Delta}{3} - \lambda & 0 \\ 0 & 0 & 0 & -\lambda \end{vmatrix} = 0$$

So by inspection, the trivial solution is $\lambda = 0$. Since recall $\lambda = \epsilon - \frac{\hbar^2 k^2}{2m}$ the desired energy is

$$\epsilon = \frac{\hbar^2 k^2}{2m} \quad (16.37)$$

Now we go back for the other three energies. Basically this entails solving the following secular determinant

$$\begin{vmatrix} \epsilon_s - \lambda & 0 & kP \\ 0 & -\frac{2}{3}\Delta - \lambda & \frac{\sqrt{2}}{3}\Delta \\ kP & \frac{\sqrt{2}}{3}\Delta & -\frac{\Delta}{3} - \lambda \end{vmatrix} = 0$$

which becomes

$$(\epsilon_s - \lambda) \begin{vmatrix} -\frac{2}{3}\Delta - \lambda & \frac{\sqrt{2}}{3}\Delta \\ \frac{\sqrt{2}}{3}\Delta & -\frac{\Delta}{3} - \lambda \end{vmatrix} + kP \begin{vmatrix} 0 & -\frac{2}{3}\Delta - \lambda \\ kP & \frac{\sqrt{2}}{3}\Delta \end{vmatrix} = 0$$

This then becomes

$$\begin{aligned} (\epsilon_s - \lambda) \left[\left(\frac{2}{3} \Delta + \lambda \right) \left(\frac{\Delta}{3} + \lambda \right) - \frac{2}{9} \Delta^2 \right] + (kP)^2 \left(\frac{2}{3} \Delta + \lambda \right) &= 0 \\ (\epsilon_s - \lambda) [\lambda^2 + \Delta \lambda] + (kP)^2 \left(\frac{2}{3} \Delta + \lambda \right) &= 0 \\ (\lambda - \epsilon_s) \lambda (\lambda + \Delta) - (kP)^2 \left(\frac{2}{3} \Delta + \lambda \right) &= 0 \end{aligned}$$

We have now the following

$$\begin{aligned} \lambda &= 0 \\ \lambda (\lambda - \epsilon_s) (\lambda + \Delta) - k^2 p^2 \left(\frac{2}{3} \Delta + \lambda \right) &= 0 \end{aligned}$$

where remember that we chose $\epsilon_p = -\frac{\Delta}{3}$ to simplify the problem. These are also equations 9 and 10 from Kane's paper. J. Phys. Chem. Solids, 1, pg 249.

Now we want the actual energies. So first assume that k is small (alternatively that k^2 is really small)

$$\lambda (\lambda - \epsilon_s) (\lambda + \Delta) + \text{small terms} = 0$$

So we can see right away that the three desired energies will center around

$$\begin{aligned} \lambda &= 0 \\ \lambda &= \epsilon_s \\ \lambda &= -\Delta \end{aligned}$$

Choose $\lambda = 0$, $\lambda = \epsilon_s$ and $\lambda = -\Delta$ successively and solve the above equation.

Case where $\lambda = 0$

$$\begin{aligned} \lambda(-\epsilon_s)(\Delta) &= k^2 p^2 \left(\frac{2}{3} \Delta \right) \\ \lambda(-\epsilon_s \Delta) &= k^2 p^2 \frac{2}{3} \Delta \\ \lambda &= -\frac{2k^2 p^2}{3\epsilon_s} \end{aligned}$$

where recall that $\lambda = \epsilon_0 - \frac{\hbar^2 k^2}{2m}$ leading to

$$\epsilon = \frac{\hbar^2 k^2}{2m} - \frac{2k^2 p^2}{3\epsilon_s} \quad (16.38)$$

This is the first of our desired energies and is actually associated with the light hole.

Case where $\lambda = \epsilon_s$

$$\begin{aligned} \epsilon_s(\lambda - \epsilon_s)(\epsilon_s + \Delta) &= k^2 p^2 \left(\frac{2}{3} \Delta + \epsilon_s \right) \\ (\lambda - \epsilon_s) &= \frac{k^2 p^2}{\epsilon_s(\epsilon_s + \Delta)} \left(\frac{2\Delta + 3\epsilon_s}{3} \right) \\ (\lambda - \epsilon_s) &= \frac{k^2 p^2}{3} \left(\frac{2\Delta + 3\epsilon_s}{\epsilon_s(\epsilon_s + \Delta)} \right) \\ (\lambda - \epsilon_s) &= \frac{k^2 p^2}{3} \left(\frac{2(\epsilon_s + \Delta) + \epsilon_s}{\epsilon_s(\epsilon_s + \Delta)} \right) \\ (\lambda - \epsilon_s) &= \frac{k^2 p^2}{3} \left(\frac{2}{\epsilon_s} + \frac{1}{\epsilon_s + \Delta} \right) \\ \lambda &= \epsilon_s + \frac{k^2 p^2}{3} \left(\frac{2}{\epsilon_s} + \frac{1}{\epsilon_s + \Delta} \right) \text{ where recall that } \lambda = \epsilon - \frac{\hbar^2 k^2}{2m} \\ \epsilon - \frac{\hbar^2 k^2}{2m} &= \epsilon_s + \frac{k^2 p^2}{3} \left(\frac{2}{\epsilon_s} + \frac{1}{\epsilon_s + \Delta} \right) \end{aligned}$$

yielding

$$\epsilon = \epsilon_s + \frac{\hbar^2 k^2}{2m} + \frac{k^2 p^2}{3} \left(\frac{2}{\epsilon_s} + \frac{1}{\epsilon_s + \Delta} \right) \quad (16.39)$$

This is another one of our desired energies and is ultimately associated with the conduction band energy.

Case where $\lambda = -\Delta$

$$\begin{aligned} \Delta(\Delta + \epsilon_s)(\lambda + \Delta) &= k^2 p^2 \left(\frac{2}{3} \Delta - \Delta \right) \\ (\lambda + \Delta) &= \frac{k^2 p^2}{\Delta(\Delta + \epsilon_s)} \left(\frac{2}{3} \Delta - \frac{3}{3} \Delta \right) \\ (\lambda + \Delta) &= \frac{k^2 p^2}{\Delta(\epsilon_s + \Delta)} \left(-\frac{\Delta}{3} \right) \\ (\lambda + \Delta) &= -\frac{k^2 p^2}{3(\epsilon_s + \Delta)} \\ \lambda &= -\Delta - \frac{k^2 p^2}{3(\epsilon_s + \Delta)} \text{ where } \lambda = \epsilon - \frac{\hbar^2 k^2}{2m} \end{aligned}$$

This results in

$$\epsilon = -\Delta + \frac{\hbar^2 k^2}{2m} - \frac{k^2 p^2}{3(\epsilon_s + \Delta)} \quad (16.40)$$

This is the last of the three desired energies and is associated with the split off band. We can also toss in the trivial solution $\lambda = 0$ which gives

$$\epsilon = \frac{\hbar^2 k^2}{2m} \quad (16.41)$$

This is the fourth and last energy and is associated with the heavy hole band.

Alternative expressions

For

$$\lambda(\lambda - \epsilon_s)(\lambda + \Delta) - k^2 p^2 \left(\frac{2}{3} \delta + \lambda \right) = 0$$

where $\epsilon_p = \frac{2mp^2}{\hbar^2}$ one can also write and solve for the expression

$$\lambda(\lambda - \epsilon_s)(\lambda + \Delta) - \frac{\hbar^2 k^2}{2m} \epsilon_p \left(\frac{2}{3} \Delta + \lambda \right) = 0$$

Case where $\lambda = 0$

$$\begin{aligned} \lambda(-\epsilon_s)(\Delta) &= \frac{\hbar^2 k^2}{2m} \epsilon_p \left(\frac{2}{3} \Delta \right) \\ \lambda &= -\frac{\hbar^2 k^2}{2m} \frac{\epsilon_p}{\Delta \epsilon_s} \left(\frac{2}{3} \Delta \right) \end{aligned}$$

leads to

$$\lambda = -\frac{\hbar^2 k^2}{2m} \left(\frac{2\epsilon_p}{3\epsilon_s} \right) \quad (16.42)$$

This is our alternative expression for the light hole band energy. Rearranging this as follows gives the effective mass

$$\begin{aligned} \frac{\hbar^2 k^2}{2m_{eff}} &= -\frac{\hbar^2 k^2}{2m} \left(\frac{2\epsilon_p}{3\epsilon_s} \right) \\ \frac{1}{m_{eff}} &= -\frac{1}{m} \left(\frac{2\epsilon_p}{3\epsilon_s} \right) \\ m_{eff} &= -m \left(\frac{3\epsilon_s}{2\epsilon_p} \right) \end{aligned}$$

Resulting in

$$\frac{m_{eff}}{m} = -\frac{3\epsilon_s}{2\epsilon_p} \quad (16.43)$$

This is the effective mass of the light hole.

Case where $\lambda = \epsilon_s$

$$\begin{aligned} \lambda(\lambda - \epsilon_s)(\lambda + \Delta) - \frac{\hbar^2 k^2}{2m} \epsilon_p \left(\frac{2}{3} \Delta + \lambda \right) &= 0 \\ \epsilon_s(\lambda - \epsilon_s)(\epsilon_s + \Delta) &= \frac{\hbar^2 k^2}{2m} \epsilon_p \left(\frac{2}{3} \Delta + \epsilon_s \right) \\ (\lambda - \epsilon_s) &= \frac{\hbar^2 k^2}{2m} \frac{\epsilon_p}{\epsilon_s(\epsilon_s + \Delta)} \left(\frac{2}{3} \Delta + \epsilon_s \right) \\ (\lambda - \epsilon_s) &= \frac{\hbar^2 k^2}{2m} \frac{\epsilon_p (\epsilon_s + \frac{2}{3} \Delta)}{\epsilon_s(\epsilon_s + \Delta)} \end{aligned}$$

This leads to

$$\lambda = \epsilon_s + \frac{\hbar^2 k^2}{2m} \frac{\epsilon_p (\epsilon_s + \frac{2}{3} \Delta)}{\epsilon_s(\epsilon_s + \Delta)} \quad (16.44)$$

This is the alternative form of the conduction band energy. To get the effective mass we evaluate

$$\begin{aligned} \frac{\hbar^2 k^2}{2m_{eff}} &= \frac{\hbar^2 k^2}{2m} \frac{\epsilon_p (\epsilon_s + \frac{2}{3} \Delta)}{\epsilon_s(\epsilon_s + \Delta)} \\ \frac{1}{m_{eff}} &= \frac{1}{m} \frac{\epsilon_p (\epsilon_s + \frac{2}{3} \Delta)}{\epsilon_s(\epsilon_s + \Delta)} \\ m_{eff} &= m \frac{\epsilon_s(\epsilon_s + \Delta)}{\epsilon_p (\epsilon_s + \frac{2}{3} \Delta)} \end{aligned}$$

or

$$\left(\frac{m_{eff}}{m} \right) = \frac{\epsilon_s(\epsilon_s + \Delta)}{\epsilon_p (\epsilon_s + \frac{2}{3} \Delta)} \quad (16.45)$$

This is the effective mass of the conduction band.

Case where $\lambda = -\Delta$

$$\begin{aligned}\lambda(\lambda - \epsilon_s)(\lambda + \Delta) - \frac{\hbar^2 k^2}{2m} \epsilon_p \left(\frac{2}{3} \Delta + \lambda \right) &= 0 \\ \Delta(\Delta + \epsilon_s)(\lambda + \Delta) &= \frac{\hbar^2 k^2}{2m} \epsilon_p \left(\frac{2}{3} \Delta - \frac{3}{3} \Delta \right) \\ (\lambda + \Delta) &= \frac{\hbar^2 k^2}{2m} \frac{\epsilon_p}{\Delta(\Delta + \epsilon_s)} \left(-\frac{\Delta}{3} \right) \\ (\lambda + \Delta) &= -\frac{\hbar^2 k^2}{2m} \frac{\epsilon_p}{3(\Delta + \epsilon_s)}\end{aligned}$$

This results in

$$\lambda = -\Delta - \frac{\hbar^2 k^2}{2m} \frac{\epsilon_p}{3(\Delta) + \epsilon_s} \quad (16.46)$$

This is the alternative expression for the split off hole energy. Now for the effective mass

$$\begin{aligned}\frac{\hbar^2 k^2}{2m_{eff}} &= -\frac{\hbar^2 k^2}{2m} \frac{\epsilon_p}{3(\Delta + \epsilon_s)} \\ \frac{1}{m_{eff}} &= -\frac{1}{m} \frac{\epsilon_p}{3(\Delta + \epsilon_s)} \\ m_{eff} &= -m \frac{3(\Delta + \epsilon_s)}{\epsilon_p}\end{aligned}$$

or

$$\left(\frac{m_{eff}}{m} \right) = -\frac{3(\Delta + \epsilon_s)}{\epsilon_p} \quad (16.47)$$

This is the effective mass of the split off hole.

Alternative expressions deux

Assume here that $\Delta \gg kP$ and $\Delta \gg \epsilon_g$ in the original expression

$$\lambda(\lambda - \epsilon_g)(\lambda + \Delta) - k^2 p^2 \left(\lambda + \frac{2}{3} \Delta \right) = 0 \quad (16.48)$$

So basically this is kind of already forecasting the solutions of this problem. Namely, that there are two bands close to the valence band edge and one band way way removed from them by Δ . Also recall that we already have

the trivial solution $\lambda = 0$ which gives the energy of the heavy hole band. So there are four total solutions.

Solve here for the two valence bands (light hole and split off hole) by assuming $\Delta \gg \lambda$ in the above equation.

$$\begin{aligned}\lambda(\lambda - \epsilon_g)(\Delta) - k^2 p^2 \left(\frac{2}{3} \Delta \right) &= 0 \text{ divide out } \Delta \\ \lambda(\lambda - \epsilon_g) - \frac{2k^2 p^2}{3} &= 0 \\ \lambda^2 - \epsilon_g \lambda - \frac{2k^2 p^2}{3} &= 0\end{aligned}$$

This last expression is a quadratic equation whose general solutions we know.

$$\lambda = \frac{\epsilon_g \pm \sqrt{\epsilon_g^2 + 4(1) \left(\frac{2k^2 p^2}{3} \right)}}{2}$$

Recall that $\lambda = \epsilon - \frac{\hbar^2 k^2}{2m}$ leading to

$$\lambda = \frac{\epsilon_g}{2} \pm \frac{1}{2} \sqrt{\epsilon_g^2 + \frac{8k^2 p^2}{3}} \quad (16.49)$$

(this is equation 13 in Kane's paper) from which we get

$$\lambda_{cb} = \frac{\epsilon_g}{2} + \frac{1}{2} \sqrt{\epsilon_g^2 + \frac{8k^2 p^2}{3}} \text{ conduction band} \quad (16.50)$$

$$\lambda_{lh} = \frac{\epsilon_g}{2} - \frac{1}{2} \sqrt{\epsilon_g^2 + \frac{8k^2 p^2}{3}} \text{ light hole band} \quad (16.51)$$

Now we go back for the split off hole band energy

$$\lambda(\lambda - \epsilon_g)(\lambda + \Delta) - k^2 p^2 \left(\lambda + \frac{2}{3} \Delta \right) = 0$$

Assume that λ is of order $-\Delta$ to get

$$-\Delta(-\Delta - \epsilon_g)(\lambda + \Delta) - k^2 p^2 \left(-\Delta + \frac{2}{3} \Delta \right) = 0$$

Solve for λ .

$$\begin{aligned}
\Delta(\Delta + \epsilon_g)(\lambda + \Delta) &= k^2 p^2 \left(-\frac{\Delta}{3} \right) \\
(\lambda + \Delta) &= -\frac{k^2 p^2}{3(\Delta + \epsilon_g)} \\
\lambda &= -\Delta - \frac{k^2 p^2}{3(\Delta + \epsilon_g)}
\end{aligned} \tag{16.52}$$

where recall that $\lambda = \epsilon_{soh} - \frac{\hbar^2 k^2}{2m}$. This results in

$$\epsilon_{soh} = -\Delta + \frac{\hbar^2 k^2}{2m} - \frac{k^2 p^2}{3(\Delta + \epsilon_g)} \tag{16.53}$$

Summary (Δ large)

$$\begin{aligned}
\epsilon_{cb} &= \frac{\hbar^2 k^2}{2m} + \frac{\epsilon_g}{2} + \frac{1}{2} \sqrt{\epsilon_g^2 + \frac{8k^2 p^2}{3}} \\
\epsilon_{hh} &= \frac{\hbar^2 k^2}{2m} \\
\epsilon_{lf} &= \frac{\hbar^2 k^2}{2m} + \frac{\epsilon_g}{2} - \frac{1}{2} \sqrt{\epsilon_g^2 + \frac{8k^2 p^2}{3}} \\
\epsilon_{soh} &= -\Delta + \frac{\hbar^2 k^2}{2m} - \frac{k^2 p^2}{3(\Delta + \epsilon_g)}
\end{aligned}$$

Note that apart from the heavy hole, all bands are non-parabolic. Also recall that the heavy hole solution came from the trivial solution $\lambda = 0$.

Relevant reading

- “Band structure of indium antimonide”
E. O. Kane J. Phys. Chem. Solids. 1, 249, 1957.

Chapter 17

Tunneling

Potential step

This is the simplest example of tunneling. Picture the following potential step.

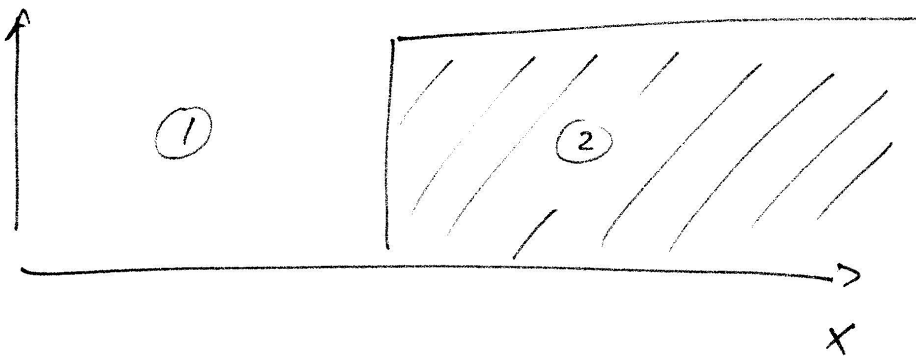


Figure 17.1: Cartoon of the potential step

The Schrodinger equation is

$$-\frac{\hbar^2}{2m} \frac{d^2\Psi}{dx^2} + V\Psi = \epsilon\Psi$$

This can be rearranged as follows.

In region 1 where $V = 0$

$$\begin{aligned} -\frac{\hbar^2}{2m} \frac{d^2\Psi}{dx^2} &= (\varepsilon - V)\Psi \\ \frac{d^2\Psi}{dx^2} &= -\frac{2m}{\hbar^2}\varepsilon\Psi \\ \frac{d^2\Psi}{dx^2} + k^2\Psi &= 0 \end{aligned}$$

where $k^2 = \frac{2m\varepsilon}{\hbar^2}$. Solutions to the wavefunction are

$$\Psi_1 = Ae^{ikx} + Be^{-ikx} \quad (17.1)$$

In region 2 where V is finite, there are two cases.

- Case 1: $\varepsilon > V$
- Case 2: $\varepsilon < V$

Case 1: ($\varepsilon > V$)

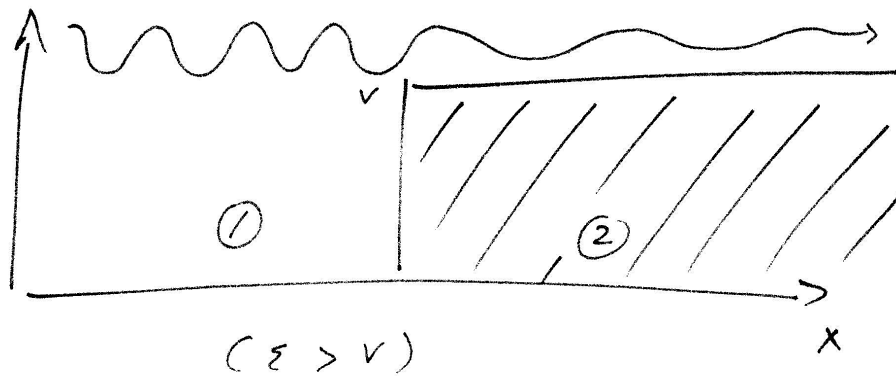


Figure 17.2: Behavior of the wavefunction in the potential step case where $\varepsilon > V$

$$\begin{aligned}
 -\frac{\hbar^2}{2m} \frac{d^2\Psi}{dx^2} + V\Psi &= \varepsilon\Psi \\
 -\frac{\hbar^2}{2m} \frac{d^2\Psi}{dx^2} &= (\varepsilon - V)\Psi \\
 \frac{d^2\Psi}{dx^2} &= -\frac{2m}{\hbar^2}(\varepsilon - V)\Psi \\
 \frac{d^2\Psi}{dx^2} + \phi^2\Psi &= 0
 \end{aligned}$$

where $\phi^2 = \frac{2m(\varepsilon - V)}{\hbar^2}$. Solutions to the wavefunction for this case, in this region, are

$$\Psi_2 = Ce^{i\phi x} + De^{-i\phi x} \quad (17.2)$$

Case 2: ($\varepsilon < V$)

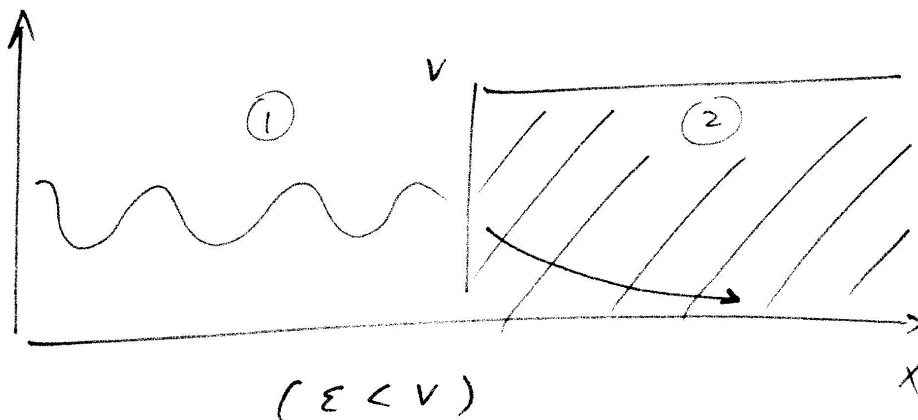


Figure 17.3: Behavior of the wavefunction in the potential step case where $\varepsilon < V$

$$\begin{aligned}
-\frac{\hbar^2}{2m} \frac{d^2\Psi}{dx^2} + V\Psi &= \varepsilon\Psi \\
-\frac{\hbar^2}{2m} \frac{d^2\Psi}{dx^2} &= (\varepsilon - V)\Psi \\
\frac{d^2\Psi}{dx^2} &= -\frac{2m}{\hbar^2}(\varepsilon - V)\Psi \\
\frac{d^2\Psi}{dx^2} &= \frac{2m}{\hbar^2}(V - \varepsilon)\Psi \\
\frac{d^2\Psi}{dx^2} - \beta^2\Psi &= 0
\end{aligned}$$

where $\beta^2 = \frac{2m(V-\varepsilon)}{\hbar^2}$. Solutions to the wavefunction for this case, in this region, are

$$\Psi_2 = Ce^{\beta x} + De^{-\beta x} \quad (17.3)$$

Now in order to proceed, we need to employ the matching conditions for the wavefunctions. The matching conditions are:

1. $\Psi_1(0) = \Psi_2(0)$
2. $\Psi_1'(0) = \Psi_2'(0)$

Furthermore, the wavefunction must be finite valued. That is, the wavefunction can't blow up or do strange things.

Transmission and reflection coefficients; $\varepsilon > V$

The wavefunctions in this case are

$$\begin{aligned}
\Psi_1 &= Ae^{ikx} + Be^{-ikx} \\
\Psi_2 &= Ce^{i\phi x} + De^{-i\phi x}
\end{aligned}$$

Since there are no sources of particles on the right to support a left going wave, $D = 0$. The wavefunctions become

$$\begin{aligned}
\Psi_1 &= Ae^{ikx} + Be^{-ikx} \\
\Psi_2 &= Ce^{i\phi x}
\end{aligned}$$

Apply the matching conditions. First one

$$\begin{aligned}
\Psi_1(0) &= \Psi_2(0) \\
A + B &= C
\end{aligned}$$

Second matching condition

$$\begin{aligned}\Psi_1'(0) &= \Psi_2'(0) \\ Aike^{ikx} - Bie^{-ikx} &= Ci\phi e^{i\phi x} \\ Aik - Bik &= Ci\phi \\ k(A - B) &= \phi C\end{aligned}$$

Now take the resulting equations and solve for A and B

$$\begin{aligned}A + B &= C \\ A - B &= \frac{\phi}{k}C \\ &\downarrow \\ 2A &= C\left(1 + \frac{\phi}{k}\right) \\ A &= \frac{C}{2}\left(1 + \frac{\phi}{k}\right)\end{aligned}\tag{17.4}$$

Next B

$$\begin{aligned}A + B &= C \\ -A + B &= -\frac{\phi}{k}C \\ &\downarrow \\ 2B &= C\left(1 - \frac{\phi}{k}\right) \\ B &= \frac{C}{2}\left(1 - \frac{\phi}{k}\right)\end{aligned}\tag{17.5}$$

Now the incident “flux” of particles from the left onto the potential step is

$$v_1|A|^2\tag{17.6}$$

where v_1 is the velocity in region 1 and $|A|^2$ is the relative probability of a right going wave in region 1. Similarly

$$v_1|B|^2\tag{17.7}$$

is the reflected flux at the potential step where again v_1 is the velocity in region 1 and $|B|^2$ is the relative probability of a left going wave. Finally

$$v_2|C|^2\tag{17.8}$$

is the transmitted flux where v_2 is the velocity in region 2 and $|C|^2$ is the relative probability of a right going wave in region 2.

The fraction of particles reflected is therefore

$$\begin{aligned}
 R &\equiv \frac{v_1|B|^2}{v_1|A|^2} = \left| \frac{B}{A} \right|^2 = \frac{B^*B}{A^*A} \\
 R &= \frac{\frac{C^2}{4} \left(1 - \frac{\phi}{k}\right)^2}{\frac{C^2}{4} \left(1 + \frac{\phi}{k}\right)^2} \\
 \boxed{R = \frac{(k-\phi)^2}{(k+\phi)^2} > 0} & \qquad (17.9)
 \end{aligned}$$

So oddly enough, this is non-classical. Some reflection occurs even though $\varepsilon > V$.

Next, the fraction of particles transmitted is

$$\begin{aligned}
 T &\equiv \frac{v_2|C|^2}{v_1|A|^2} = \frac{\phi|C|^2}{k|A|^2} = \frac{\phi C^2}{k \frac{C^2}{4} \left(1 + \frac{\phi}{k}\right)^2} \\
 &= \frac{4\phi}{k \left(1 + \frac{\phi}{k}\right)^2} = \frac{4k^2\phi}{k(k+\phi)^2}
 \end{aligned}$$

resulting in

$$\boxed{T = \frac{4k\phi}{(k+\phi)^2} \neq 0} \qquad (17.10)$$

Transmission and reflection coefficients; $\varepsilon < V$

Recall that the wavefunctions in this case are

$$\begin{aligned}
 \Psi_1 &= Ae^{ikx} + Be^{-ikx} \\
 \Psi_2 &= Ce^{\beta x} + De^{-\beta x}
 \end{aligned}$$

where $k = \sqrt{\frac{2m\varepsilon}{\hbar}}$ and $\beta = \sqrt{\frac{2m(V-\varepsilon)}{\hbar}}$. Now recall that the wavefunction in region 1 and 2 must be finite valued. This means that $C = 0$. Can't have something blow up here. Therefore

$$\begin{aligned}
 \Psi_1 &= Ae^{ikx} + Be^{-ikx} \\
 \Psi_2 &= De^{-\beta x}
 \end{aligned}$$

Apply the matching conditions now. First one

$$\begin{aligned}\Psi_1(0) &= \Psi_2(0) \\ A + B &= D\end{aligned}$$

Second one

$$\begin{aligned}\Psi_1'(0) &= \Psi_2'(0) \\ ikA - ikB &= -\beta D \\ ik(A - B) &= -\beta D\end{aligned}$$

Now take the resulting equations and solve for A and B . First let's solve for the A coefficient.

$$\begin{aligned}A + B &= D \\ A - B &= -\frac{\beta}{ik}D \\ &\downarrow \\ 2A &= D\left(1 - \frac{\beta}{ik}\right) \\ A &= \frac{D}{2}\left(1 - \frac{\beta}{ik}\right)\end{aligned}\tag{17.11}$$

Now let's solve for B

$$\begin{aligned}A + B &= D \\ -A + B &= \frac{\beta}{ik}D \\ 2B &= D\left(1 + \frac{\beta}{ik}\right) \\ B &= \frac{D}{2}\left(1 + \frac{\beta}{ik}\right)\end{aligned}\tag{17.12}$$

From before, the relevant fluxes are

$$\begin{aligned}v_1|A|^2 &\text{ (leftgoing)} \\ v_1|B|^2 &\text{ (rightgoing)} \\ v_2|D|^2 &\text{ (transmitted)}\end{aligned}$$

The fraction of particles reflected is

$$\begin{aligned}
 R &= \frac{v_1|B|^2}{v_1|A|^2} = \left| \frac{B}{A} \right|^2 = \frac{B^*B}{A^*A} \\
 &= \frac{\frac{D^2}{4} \left(1 + \frac{\beta}{ik}\right) \left(1 - \frac{\beta}{ik}\right)}{\frac{D^2}{4} \left(1 - \frac{\beta}{ik}\right) \left(1 + \frac{\beta}{ik}\right)} \\
 &= \frac{\frac{D^2}{4} \left(1 + \frac{\beta^2}{k^2}\right)}{\frac{D^2}{4} \left(1 + \frac{\beta^2}{k^2}\right)} \\
 &= 1
 \end{aligned}$$

However the fraction transmitted is

$$\begin{aligned}
 T &= \frac{v_2|D|^2}{v_1|A|^2} = \frac{\beta|D|^2}{k|A|^2} \\
 &= \frac{\beta D^2}{k \frac{D^2}{4} \left(1 + \frac{\beta^2}{k^2}\right)} \\
 &= \frac{4\beta}{k \left(1 + \frac{\beta^2}{k^2}\right)}
 \end{aligned}$$

giving

$$\boxed{T = \frac{4\beta k}{(k^2 + \beta^2)} \neq 0} \quad (17.13)$$

This gives a non-classical result. There is some probability of finding something in region 2.

Potential barrier

From the last section we know that the wavefunctions have the following form.

$$\begin{aligned}
 \Psi_1 &= Ae^{ikx} + Be^{-ikx} \\
 \Psi_2 &= Ce^{i\phi x} + De^{-i\phi x} \quad (\varepsilon > V) \\
 \Psi_2 &= Ce^{\beta x} + De^{-\beta x} \quad (\varepsilon < V) \\
 \Psi_3 &= Fe^{ikx} + Ge^{-ikx}
 \end{aligned}$$

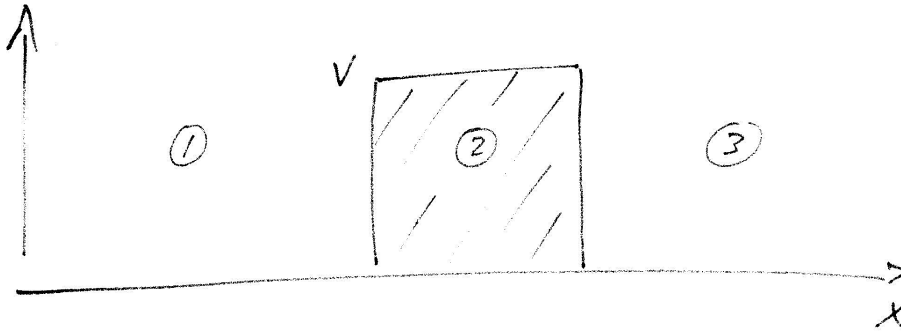
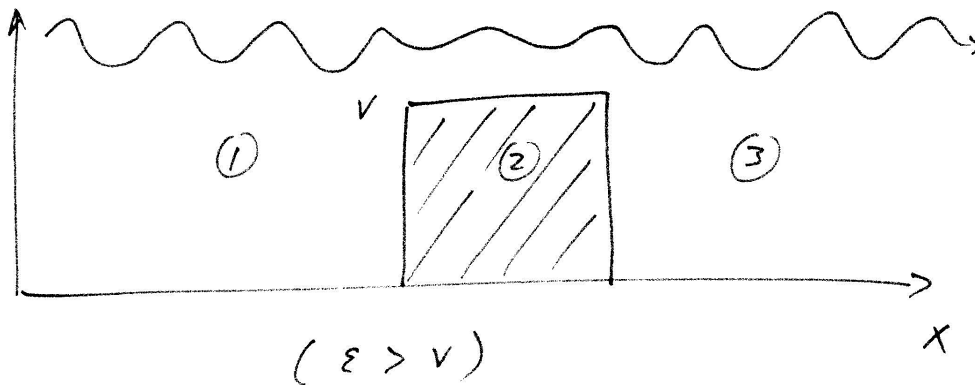


Figure 17.4: Cartoon of the potential barrier

where $k^2 = \frac{2m\varepsilon}{\hbar^2}$, $\phi^2 = \frac{2m(\varepsilon - V)}{\hbar^2}$, and $\beta^2 = \frac{2m(V - \varepsilon)}{\hbar^2}$. Now as before we apply the matching conditions to the wavefunctions to solve for the coefficients and in turn for the transmission (T) and reflection (R) coefficients.

Case 1: ($\varepsilon > V$)

Figure 17.5: Behavior of the wavefunction in the potential barrier case where $\varepsilon > V$

Assuming that there is no source of particles on the right then $G = 0$.

The wavefunctions in the three regions become

$$\begin{aligned}\Psi_1 &= Ae^{ikx} + Be^{-ikx} \\ \Psi_2 &= Ce^{i\phi x} + De^{-i\phi x} \\ \Psi_3 &= Fe^{ikx}\end{aligned}$$

Apply the matching conditions at $x = 0$. First one

$$\begin{aligned}\Psi_1(0) &= \Psi_2(0) \\ A + B &= C + D\end{aligned}\tag{17.14}$$

Second one

$$\begin{aligned}\Psi'_1(0) &= \Psi'_2(0) \\ Aik - Bik &= Ci\phi - Di\phi \\ A - B &= \frac{\phi}{k}(C - D)\end{aligned}\tag{17.15}$$

Apply the matching conditions at $x = a$. First one

$$\begin{aligned}\Psi_2(a) &= \Psi_3(a) \\ Ce^{i\phi a} + De^{-i\phi a} &= Fe^{ika}\end{aligned}\tag{17.16}$$

Second one

$$\begin{aligned}\Psi'_2(a) &= \Psi'_3(a) \\ Ci\phi e^{i\phi a} - Di\phi e^{-i\phi a} &= Fike^{ika} \\ C\phi e^{i\phi a} - D\phi e^{-i\phi a} &= Fk\epsilon^{ika}\end{aligned}\tag{17.17}$$

Solve for A.

$$\begin{aligned}A + B &= C + D \\ A - B &= \frac{\phi}{k}C - \frac{\phi}{k}D \\ &\downarrow \\ 2A &= C\left(1 + \frac{\phi}{k}\right) + D\left(1 - \frac{\phi}{k}\right) \\ A &= \frac{C}{2}\left(1 + \frac{\phi}{k}\right) + \frac{D}{2}\left(1 - \frac{\phi}{k}\right)\end{aligned}\tag{17.18}$$

Solve for B.

$$\begin{aligned}
 A + B &= C + D \\
 -A + B &= -\frac{\phi}{k}C + \frac{\phi}{k}D \\
 &\downarrow \\
 2B &= C\left(1 - \frac{\phi}{k}\right) + D\left(1 + \frac{\phi}{k}\right) \\
 B &= \frac{C}{2}\left(1 - \frac{\phi}{k}\right) + \frac{D}{2}\left(1 + \frac{\phi}{k}\right) \tag{17.19}
 \end{aligned}$$

Solve for C.

$$\begin{aligned}
 Ce^{i\phi a} + De^{-i\phi a} &= Fe^{ika} \\
 C\phi e^{i\phi a} - D\phi e^{-i\phi a} &= Fke^{ika} \\
 &\downarrow \\
 C\phi e^{i\phi a} + D\phi e^{-i\phi a} &= F\phi e^{ika} \\
 C\phi e^{i\phi a} - D\phi e^{-i\phi a} &= Fke^{ika} \\
 &\downarrow \\
 2C\phi e^{i\phi a} &= F(k + \phi)e^{ika} \\
 C &= F\frac{k + \phi}{2\phi}e^{ia(k - \phi)} \tag{17.20}
 \end{aligned}$$

Solve for D.

$$\begin{aligned}
 Ce^{i\phi a} + De^{-i\phi a} &= Fe^{ika} \\
 -C\phi e^{i\phi a} + D\phi e^{-i\phi a} &= -Fke^{ika} \\
 &\downarrow \\
 C\phi e^{i\phi a} + D\phi e^{-i\phi a} &= F\phi e^{ika} \\
 -C\phi e^{i\phi a} + D\phi e^{-i\phi a} &= -Fke^{ika} \\
 &\downarrow \\
 2D\phi e^{-i\phi a} &= F(\phi - k)e^{ika} \\
 D &= F\frac{\phi - k}{2\phi}e^{ia(k + \phi)} \tag{17.21}
 \end{aligned}$$

Now that we have expressions for A,B,C,D all in terms of F, let's re-express

everything in terms of F. From before

$$\begin{aligned}
A &= \frac{1}{2}C \left(1 + \frac{\phi}{k}\right) + \frac{1}{2}D \left(1 - \frac{\phi}{k}\right) \\
&= \frac{1}{2}F \frac{k + \phi}{2\phi} e^{ia(k-\phi)} \left(1 + \frac{\phi}{k}\right) + \frac{1}{2}F \frac{\phi - k}{2\phi} e^{ia(k+\phi)} \left(1 - \frac{\phi}{k}\right) \\
&= \frac{F}{4\phi} \left((k + \phi) \left(1 + \frac{\phi}{k}\right) e^{ia(k-\phi)} + (\phi - k) \left(1 - \frac{\phi}{k}\right) e^{ia(k+\phi)} \right) \\
&= \frac{F}{4\phi} \left(\frac{(k + \phi)(k + \phi)}{k} e^{ia(k-\phi)} - \frac{(k - \phi)(k - \phi)}{k} e^{ia(k+\phi)} \right) \\
&= \frac{F e^{ika}}{4k\phi} \left((k + \phi)^2 e^{-ia\phi} - (k - \phi)^2 e^{ia\phi} \right) \\
&= \frac{F e^{ika}}{4k\phi} \left(k^2 e^{-ia\phi} + 2k\phi e^{-ia\phi} + \phi^2 e^{-ia\phi} - k^2 e^{ia\phi} + 2k\phi e^{ia\phi} - \phi^2 e^{ia\phi} \right) \\
&= \frac{F e^{ika}}{4k\phi} \left(-k^2 (e^{ia\phi} - e^{-ia\phi}) - \phi^2 (e^{ia\phi} - e^{-ia\phi}) + 2k\phi (e^{ia\phi} + e^{-ia\phi}) \right) \\
&= \frac{F e^{ika}}{4k\phi} \left(4k\phi \frac{e^{ia\phi} + e^{-ia\phi}}{2} - 2i(k^2 + \phi^2) \frac{e^{ia\phi} - e^{-ia\phi}}{2i} \right)
\end{aligned}$$

This gives our desired expression for A

$$A = \frac{F e^{ika}}{4k\phi} (4k\phi \cos(\phi a) - 2i(k^2 + \phi^2) \sin(\phi a)) \quad (17.22)$$

Next we do B

$$\begin{aligned}
B &= \frac{1}{2}C \left(1 - \frac{\phi}{k}\right) + \frac{1}{2}D \left(1 + \frac{\phi}{k}\right) \\
&= \frac{1}{2}F \frac{k + \phi}{2\phi} e^{ia(k-\phi)} \left(1 - \frac{\phi}{k}\right) + \frac{1}{2}F \frac{\phi - k}{2\phi} e^{ia(k+\phi)} \left(1 + \frac{\phi}{k}\right) \\
&= \frac{F}{4\phi} \left((k + \phi) \left(1 - \frac{\phi}{k}\right) e^{ia(k-\phi)} + (\phi - k) \left(1 + \frac{\phi}{k}\right) e^{ia(k+\phi)} \right) \\
&= \frac{F}{4\phi} \left(\frac{(k + \phi)(k - \phi)}{k} e^{ia(k-\phi)} - \frac{(k - \phi)(k + \phi)}{k} e^{ia(k+\phi)} \right) \\
&= \frac{F}{4k\phi} \left((k^2 - \phi^2) e^{ia(k-\phi)} - (k^2 - \phi^2) e^{ia(k+\phi)} \right) \\
&= \frac{2iF(k^2 - \phi^2)}{4k\phi} e^{ika} \left(\frac{e^{-ia\phi} - e^{ia\phi}}{2i} \right) \\
&= \frac{2iF(-k^2 + \phi^2)}{4k\phi} e^{ika} \left(\frac{-e^{-ia\phi} + e^{ia\phi}}{2i} \right)
\end{aligned}$$

giving us our desired expression for B

$$B = \frac{iF(\phi^2 - k^2)}{2k\phi} e^{ika} \sin(\phi a) \quad (17.23)$$

Now (finally!) we can evaluate what the reflection and transmission coefficients are. From before the relevant fluxes are

$$\begin{aligned} v_1|A|^2 & \text{ (leftgoing)} \\ v_1|B|^2 & \text{ (rightgoing)} \\ v_3|F|^2 & \text{ (transmitted)} \end{aligned}$$

The fraction of particles reflected is

$$\begin{aligned} R &= \frac{v_1|B|^2}{v_1|A|^2} = \left| \frac{B}{A} \right|^2 = \frac{B^*B}{A^*A} \\ &= \frac{\frac{F^2(\phi^2 - k^2)^2}{4(k\phi)^2} \sin^2(\phi a)}{\frac{F^2}{16(k\phi)^2} (16(k\phi)^2 \cos^2(\phi a) + 4(k^2 + \phi^2) \sin^2(\phi a))} \\ & \boxed{R = \frac{(\phi^2 - k^2)^2 \sin^2(\phi a)}{4(k\phi)^2 \cos^2(\phi a) + (k^2 + \phi^2) \sin^2(\phi a)}} \end{aligned} \quad (17.24)$$

The fraction transmitted is

$$T = \frac{v_3|F|^2}{v_1|A|^2}$$

where note that $v_3 = v_1$. Therefore

$$\begin{aligned} T &= \frac{|F|^2}{|A|^2} = \frac{F^*F}{A^*A} \\ &= \frac{16(k\phi)^2}{16(k\phi)^2 \cos^2(\phi a) + 4(k^2 + \phi^2) \sin^2(\phi a)} \\ &= \frac{1}{\cos^2(\phi a) + \frac{(k^2 + \phi^2)^2}{4(k\phi)^2} \sin^2(\phi a)} \end{aligned} \quad (17.25)$$

Now we could stop here, but usually in many textbooks they expand this out a little further. Recall that

$$\begin{aligned} k^2 &= \frac{2m\varepsilon}{\hbar^2} \\ \phi^2 &= \frac{2m(\varepsilon - V)}{\hbar^2} \\ (k\phi)^2 &= \frac{4m^2}{\hbar^4} \varepsilon(\varepsilon - V) \end{aligned}$$

This leads to

$$\begin{aligned}
 T &= \frac{1}{\cos^2(\phi a) + \frac{(2m\varepsilon + 2m(\varepsilon - V))^2}{16m^2\varepsilon(\varepsilon - V)} \sin^2(\phi a)} \\
 &= \frac{1}{\cos^2(\phi a) + \frac{4m^2(2\varepsilon - V)^2}{16m^2\varepsilon(\varepsilon - V)} \sin^2(\phi a)} \\
 &= \frac{1}{\cos^2(\phi a) + \frac{(2\varepsilon - V)^2}{4\varepsilon(\varepsilon - V)} \sin^2(\phi a)} \\
 &= \frac{4\varepsilon(\varepsilon - V)}{4\varepsilon(\varepsilon - V)\cos^2(\phi a) + (4\varepsilon^2 - 4\varepsilon V + V^2)\sin^2(\phi a)} \\
 &= \frac{4\varepsilon(\varepsilon - V)}{4\varepsilon^2\cos^2(\phi a) - 4\varepsilon V\cos^2(\phi a) + 4\varepsilon^2\sin^2(\phi a) - 4\varepsilon V\sin^2(\phi a) + V^2\sin^2(\phi a)} \\
 &= \frac{4\varepsilon(\varepsilon - V)}{(4\varepsilon^2 - 4\varepsilon V) + V^2\sin^2(\phi a)} \\
 &= \frac{4\varepsilon(\varepsilon - V)}{4\varepsilon(\varepsilon - V) + V^2\sin^2(\phi a)}
 \end{aligned}$$

This leads to the standard textbook expression

$$\boxed{T = \frac{1}{1 + \frac{V^2}{4\varepsilon(\varepsilon - V)}\sin^2(\phi a)}} \quad (17.26)$$

Case 2: ($\varepsilon < V$)

Now rather than doing all this math again, one can be smart about this. Just note that $\beta = i\phi$. So take our previous results from case 1 and just replace ϕ with $i\phi$. (We just analyze the transmission coefficient since this is what we really care about). Another tortuous math workout. Starting with the assumption of no sources on the right ($G = 0$), the wavefunctions are

$$\begin{aligned}
 \Psi_1 &= Ae^{ikx} + Be^{-ikx} \\
 \Psi_2 &= Ce^{\beta x} + De^{-\beta x} \\
 \Psi_3 &= Fe^{ikx}
 \end{aligned}$$

Apply the matching conditions at $x = a$. First one

$$\begin{aligned}
 \Psi_1(0) &= \Psi_2(0) \\
 A + B &= C + D
 \end{aligned} \quad (17.27)$$

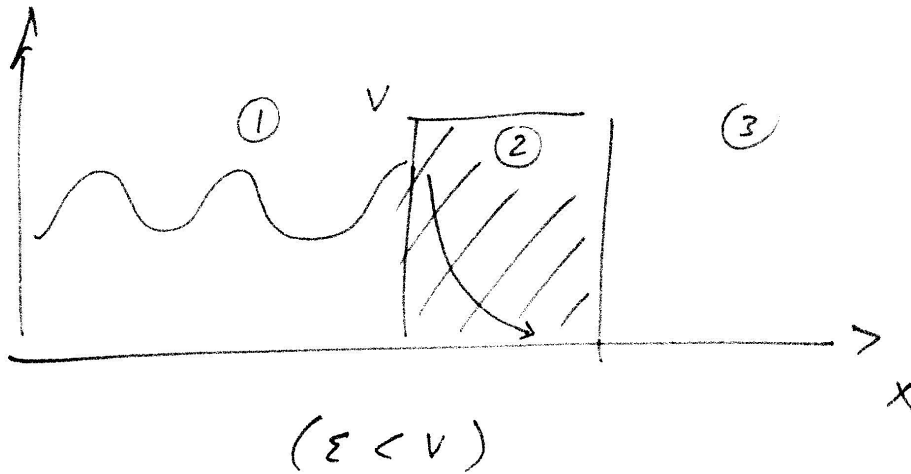


Figure 17.6: Behavior of the wavefunction in the potential barrier case where $\varepsilon < V$

Second one

$$\begin{aligned}\Psi_1'(0) &= \Psi_2'(0) \\ A - B &= \frac{\beta}{ik}(C - D)\end{aligned}\quad (17.28)$$

Apply the matching conditions at $x = a$. First one

$$\begin{aligned}\Psi_2(a) &= \Psi_3(a) \\ Ce^{\beta a} + De^{-\beta a} &= Fe^{ika}\end{aligned}\quad (17.29)$$

Second one

$$\begin{aligned}\Psi_2'(a) &= \Psi_3'(a) \\ Ce^{\beta a} - De^{-\beta a} &= \frac{ik}{\beta}Fe^{ika}\end{aligned}\quad (17.30)$$

Solve for A

$$\begin{aligned}
A + B &= C + D \\
A - B &= \frac{\beta}{ik}(C - D) \\
&\downarrow \\
2A &= C \left(1 + \frac{\beta}{ik}\right) + D \left(1 - \frac{\beta}{ik}\right) \\
A &= \frac{C}{2} \left(1 + \frac{\beta}{ik}\right) + \frac{D}{2} \left(1 - \frac{\beta}{ik}\right) \tag{17.31}
\end{aligned}$$

Solve for B

$$\begin{aligned}
A + B &= C + D \\
-A + B &= \frac{\beta}{ik}(-C + D) \\
&\downarrow \\
2B &= C \left(1 - \frac{\beta}{ik}\right) + D \left(1 + \frac{\beta}{ik}\right) \\
B &= \frac{C}{2} \left(1 - \frac{\beta}{ik}\right) + \frac{D}{2} \left(1 + \frac{\beta}{ik}\right) \tag{17.32}
\end{aligned}$$

Solve for C

$$\begin{aligned}
Ce^{\beta a} + De^{-\beta a} &= Fe^{ika} \\
Ce^{\beta a} - De^{-\beta a} &= \frac{ik}{\beta} Fe^{ika} \\
&\downarrow \\
2Ce^{\beta a} &= F \left(1 + \frac{ik}{\beta}\right) e^{ika} \\
C &= \frac{F}{2} \left(1 + \frac{ik}{\beta}\right) e^{ika - \beta a} \tag{17.33}
\end{aligned}$$

Solve for D

$$\begin{aligned}
Ce^{\beta a} + De^{-\beta a} &= Fe^{ika} \\
-Ce^{\beta a} + De^{-\beta a} &= -\frac{ik}{\beta} Fe^{ika} \\
&\downarrow \\
2De^{-\beta a} &= F \left(1 - \frac{ik}{\beta}\right) e^{ika} \\
D &= \frac{F}{2} \left(1 - \frac{ik}{\beta}\right) e^{ika + \beta a} \tag{17.34}
\end{aligned}$$

Now we have a relation for A in terms of C and D . Replace the above C and D expressions into A to obtain A as a function of F (we did B just for the sake of completeness).

$$\begin{aligned}
 A &= \frac{F}{4} \left(1 + \frac{ik}{\beta}\right) \left(1 + \frac{\beta}{ik}\right) e^{ika-\beta a} + \frac{F}{4} \left(1 - \frac{ik}{\beta}\right) \left(1 - \frac{\beta}{ik}\right) e^{ika+\beta a} \\
 &= \frac{F}{4} \left(2 + \frac{\beta}{ik} + \frac{ik}{\beta}\right) e^{ika-\beta a} + \frac{F}{4} \left(2 - \frac{\beta}{ik} - \frac{ik}{\beta}\right) e^{ika+\beta a} \\
 &= \frac{F}{4} e^{ika} \left(\left(2 + \frac{\beta}{ik} + \frac{ik}{\beta}\right) e^{-\beta a} + \left(2 - \frac{\beta}{ik} - \frac{ik}{\beta}\right) e^{\beta a} \right) \quad (17.35)
 \end{aligned}$$

Now from before we know that the transmission coefficient is

$$\begin{aligned}
 T &= \left| \frac{F}{A} \right|^2 = \frac{F^* F}{A^* A} \\
 &= \frac{16}{\left(2 + \frac{\beta}{ik} + \frac{ik}{\beta}\right)^2 + \left(2 - \frac{\beta}{ik} - \frac{ik}{\beta}\right)^2 + 2 \left(2 + \frac{\beta}{ik} + \frac{ik}{\beta}\right) \left(2 - \frac{\beta}{ik} - \frac{ik}{\beta}\right) \cosh(2\beta a)} \\
 &= \frac{16}{\left(2 + \frac{\beta^2 - k^2}{\beta ik}\right)^2 + \left(2 - \frac{\beta^2 - k^2}{\beta ik}\right)^2 + 2 \left(2 + \frac{\beta}{ik} + \frac{ik}{\beta}\right) \left(2 - \frac{\beta}{ik} - \frac{ik}{\beta}\right) \cosh(2\beta a)} \\
 &= \frac{16}{8 - 2 \frac{(\beta^2 - k^2)^2}{(\beta k)^2} + 2 \left(4 + \frac{\beta^2 - k^2}{(\beta k)^2}\right) \cosh(2\beta a)} \\
 &= \frac{16}{\left(8 - \frac{2(\beta^2 - k^2)^2}{(\beta k)^2}\right) + \left(8 + \frac{2(\beta^2 - k^2)^2}{(\beta k)^2}\right) \cosh(2\beta a)}
 \end{aligned}$$

it gets ugly here but can be simplified to give

$$= \frac{16(\beta k)^2}{8(\beta k)^2(1 + \cosh(2\beta a)) - 2(\beta^2 - k^2)^2(1 - \cosh(2\beta a))}$$

Now use the relation $\cosh(2x) = 1 + 2\sinh^2(x)$ to get

$$= \frac{1}{1 + \frac{(\beta^2 + k^2)^2}{4(\beta k)^2} \sinh^2(\beta a)}$$

where $\beta^2 = \frac{2m(V-\epsilon)}{\hbar^2}$, $k^2 = \frac{2m\epsilon}{\hbar^2}$, and $(\beta k)^2 = \frac{4m^2\epsilon(V-\epsilon)}{\hbar^4}$. This leads us to our final textbook expression

$$\boxed{T = \frac{1}{1 + \frac{V^2}{4\epsilon(V-\epsilon)} \sinh^2(\beta a)}} \quad (17.36)$$

Again $T \neq 0$ and we obtain a non-classical result.

Exercises

1. A conduction electron in a semiconductor can be described as a free particle in a constant potential with an effective mass m_e . Both the constant potential and the effective mass are material dependent. Consider the following idealized one dimensional system where semi-infinite slabs of semiconductors A and B are joined together. (See Figure 17.7) Assume that one of the matching conditions is $\frac{1}{m_A}\Psi_A(0) = \frac{1}{m_B}\Psi_B(0)$. Also assume $V_o > 0$. (A) Find the transmission coefficient T for $E > V_o$ for a right going wave. (B) Now assume that $m_B < m_A$. Show that an appropriate choice of incoming energy E yields complete transmission ($T = 1$). Is there such an energy if $m_B > m_A$. (C) Show that for $V_o = 0$ and $m_A \neq m_B$, T is always less than unity.

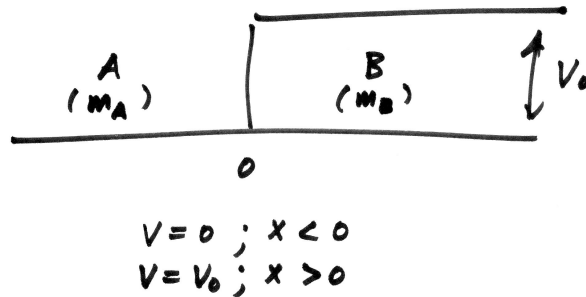


Figure 17.7: Semiconductor junction

Chapter 18

The WKB approximation

Preliminaries

Here we derive the WKB form of the wavefunction. Starting with the Schrodinger equation in the regime where $\varepsilon > V$

$$\begin{aligned} -\frac{\hbar^2}{2m} \frac{d^2\Psi}{dx^2} + V\Psi &= \varepsilon\Psi \\ -\frac{\hbar^2}{2m} \frac{d^2\Psi}{dx^2} &= (\varepsilon - V)\Psi \\ \frac{d^2\Psi}{dx^2} &= -\frac{2m}{\hbar^2}(\varepsilon - V)\Psi \\ \frac{d^2\Psi}{dx^2} + \frac{2m}{\hbar^2}(\varepsilon - V)\Psi &= 0 \end{aligned} \tag{18.1}$$

Let $k^2 = \frac{2m}{\hbar^2}(\varepsilon - V)$ so that we get

$$\frac{d^2\Psi}{dx^2} + k^2\Psi = 0$$

Now in general, if the potential is slowly varying, then both the amplitude and the phase of the wavefunction will change. The general form of the wavefunction is

$$\Psi(x) = A(x)e^{i\phi(x)} = Ae^{i\phi} \text{ (shorthand)} \tag{18.2}$$

So we insert this form of the wavefunction into Schrodinger's equation above. To begin we evaluate some derivatives that we will need. First derivative

$$\begin{aligned} \frac{d\Psi}{dx} &= Aie^{i\phi}\phi' + e^{i\phi}A' \\ &= (A' + iA\phi')e^{i\phi} \end{aligned} \tag{18.3}$$

Next we do the second derivative

$$\begin{aligned}
 \frac{d^2\Psi}{dx^2} &= i(A' + iA\phi')e^{i\phi}\phi' + e^{i\phi}(A'' + iA\phi'' + i\phi'A') \\
 &= (iA'\phi' - A\phi'^2)e^{i\phi} + (A'' + iA\phi'' + i\phi'A')e^{i\phi} \\
 &= (2iA'\phi' - A\phi'^2 + A'' + iA\phi'')e^{i\phi} \\
 &= (A'' + 2iA'\phi' + iA\phi'' - A\phi'^2)e^{i\phi}
 \end{aligned} \tag{18.4}$$

Replace this into the above Schrodinger equation; $\frac{d^2\Psi}{dx^2} + k^2\Psi = 0$

$$\begin{aligned}
 (A'' + 2iA'\phi' + iA\phi'' - A\phi'^2)e^{i\phi} + k^2Ae^{i\phi} &= 0 \\
 A'' + 2iA'\phi' + iA\phi'' - A\phi'^2 + k^2A &= 0
 \end{aligned} \tag{18.5}$$

Group the real and imaginary terms. First the real terms

$$\begin{aligned}
 A'' - A\phi'^2 + k^2A &= 0 \\
 A'' - A\phi'^2 &= -k^2A
 \end{aligned} \tag{18.6}$$

Next the imaginary terms

$$\begin{aligned}
 2iA'\phi' + iA\phi'' &= 0 \\
 2A'\phi' + A\phi'' &= 0 \\
 (A^2\phi')' &= 0
 \end{aligned} \tag{18.7}$$

Now we want to solve for A, ϕ so we can have an explicit wavefunction. Solve the imaginary equation first

$$\begin{aligned}
 (A^2\phi')' &= 0 \\
 A^2\phi' &= C^2 \text{ (a constant)} \\
 A^2 &= \frac{C^2}{\phi'} \\
 A &= \frac{C}{\sqrt{\phi'}}
 \end{aligned} \tag{18.8}$$

The real part cannot be solve exactly. However, if A varies very slowly then A'' is very small and can be ignored.

$$\begin{aligned}
 A'' - A\phi'^2 &= -k^2A \\
 -A\phi'^2 &= -k^2A \\
 -\phi'^2 &= -k^2 \\
 \phi' &= \pm k \\
 \phi &= \pm \int k dx
 \end{aligned} \tag{18.9}$$

Put these two expressions for A and ϕ back into our wavefunction

$$\begin{aligned}\Psi &= Ae^{i\phi} \\ &= \frac{C_{\pm}}{\sqrt{k}} e^{\pm i \int k dx}\end{aligned}\quad (18.10)$$

This is the desired form of the WKB wavefunction in the case where $\varepsilon > V$. In the case where $\varepsilon < V$ the WKB form of the wavefunction is

$$\Psi = \frac{C_{\pm}}{\sqrt{\beta}} e^{\pm \int \beta dx}\quad (18.11)$$

which is basically the same as the first case but where we just replace k with $i\beta$.

Arbitrary potential step

Using our just derived WKB wavefunctions we evaluate the following arbitrary potential step ($\varepsilon < V$).

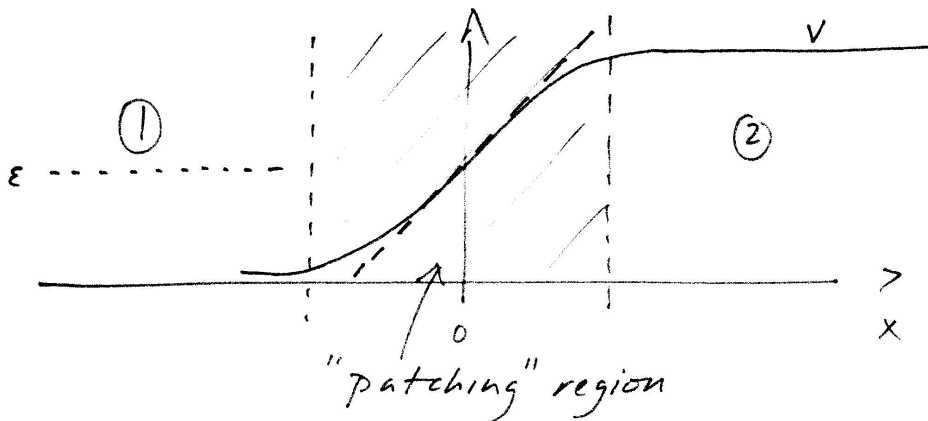


Figure 18.1: Cartoon of the arbitrary potential step

The wavefunctions in region 1 and region 2 are, in general

$$\begin{aligned}\Psi_1 &= \frac{A}{\sqrt{k}} e^{i \int_x^0 k dx} + \frac{B}{\sqrt{k}} e^{-i \int_x^0 k dx} \quad \text{where } (x < 0) \\ \Psi_2 &= \frac{C}{\sqrt{\beta}} e^{\int_0^x \beta dx} + \frac{D}{\sqrt{\beta}} e^{-\int_0^x \beta dx} \quad \text{where } (x > 0)\end{aligned}$$

Because we can't have the wavefunction blow up in region 2 ($C = 0$). The wavefunctions are therefore

$$\Psi_1 = \frac{A}{\sqrt{k}} e^{i \int_x^0 k dx} + \frac{B}{\sqrt{k}} e^{-i \int_x^0 k dx} \quad \text{where } (x < 0) \quad (18.12)$$

$$\Psi_2 = \frac{D}{\sqrt{\beta}} e^{-\int_0^x \beta dx} \quad \text{where } (x > 0) \quad (18.13)$$

Now consider a linear approximation to the potential between region 1 and region 2

$$V = \varepsilon + V' x \quad (18.14)$$

First we determine the patching wavefunction in the boundary, no man's land, between region 1 and region 2. Starting with the Schrodinger equation

$$-\frac{\hbar^2}{2m} \frac{d^2 \Psi_p}{dx^2} + V \Psi_p = \varepsilon \Psi_p$$

where $V = \varepsilon + V' x$. This reduces as follows

$$\begin{aligned} -\frac{\hbar^2}{2m} \frac{d^2 \Psi_p}{dx^2} + (\varepsilon + V' x) \Psi_p &= \varepsilon \Psi_p \\ -\frac{\hbar^2}{2m} \frac{d^2 \Psi_p}{dx^2} + V' x \Psi_p &= 0 \\ \frac{d^2 \Psi_p}{dx^2} - \frac{2mV'}{\hbar^2} x \Psi_p &= 0 \\ \frac{d^2 \Psi_p}{dx^2} &= \frac{2mV'}{\hbar^2} x \Psi_p \end{aligned}$$

Let $\alpha^3 = \frac{2mV'}{\hbar^2}$ leading to

$$\frac{d^2 \Psi_p}{dx^2} = \alpha^3 x \Psi_p$$

Now let $z = \alpha x$ leading to

$$\frac{d^2 \Psi_p}{dz^2} = z \Psi_p \quad (18.15)$$

This is Airy's equation whose solutions are called Airy functions. General solution is a linear combination of Airy functions, $A_i(z)$ and $B_i(z)$. The

asymptotic forms of Airy functions are for ($z < 0$)

$$\begin{aligned} A_i(z) &= \frac{1}{\sqrt{\pi}(-z)^{\frac{1}{4}}} \sin\left(\frac{2}{3}(-z)^{\frac{3}{2}} + \frac{\pi}{4}\right) \\ B_i(z) &= \frac{1}{\sqrt{\pi}(-z)^{\frac{1}{4}}} \cos\left(\frac{2}{3}(-z)^{\frac{3}{2}} + \frac{\pi}{4}\right) \end{aligned}$$

and for ($z > 0$)

$$\begin{aligned} A_i(z) &= \frac{1}{2\sqrt{\pi}z^{\frac{1}{4}}} e^{-\frac{2}{3}z^{\frac{3}{2}}} \\ B_i(z) &= \frac{1}{\sqrt{\pi}z^{\frac{1}{4}}} e^{\frac{2}{3}z^{\frac{3}{2}}} \end{aligned}$$

Now

$$\Psi_p = aA_i(z) + bB_i(z) \quad (18.16)$$

where a and b are constants. Now in region 1

$$k = \sqrt{\frac{2m(\varepsilon - V)}{\hbar^2}}$$

where $V = (\varepsilon + V'x)$

$$\begin{aligned} k &= \sqrt{\frac{2mV'}{\hbar^2}} \sqrt{-x} \\ &= \alpha^{\frac{3}{2}} \sqrt{-x} \end{aligned} \quad (18.17)$$

where recall that $\alpha = \left(\frac{2mV'}{\hbar^2}\right)^{\frac{1}{3}}$.

In region 2

$$\begin{aligned} \beta &= \sqrt{\frac{2m(V - \varepsilon)}{\hbar^2}} \\ &= \sqrt{\frac{2mV'}{\hbar^2}} \sqrt{x} \\ &= \alpha^{\frac{3}{2}} \sqrt{x} \end{aligned} \quad (18.18)$$

Now go back to the wavefunctions in region 1 and region 2 and make them more explicit.

Region 1

$$\Psi_1 = \frac{A}{\sqrt{k}} e^{i \int_x^0 k dx} + \frac{B}{\sqrt{k}} e^{-i \int_x^0 k dx}$$

and $k = \alpha^{\frac{3}{2}} \sqrt{-x}$

$$\Psi_1 = \frac{A}{\alpha^{\frac{3}{4}} (-x)^{\frac{1}{4}}} e^{i \int_x^0 \alpha^{\frac{3}{2}} \sqrt{-x} dx} + \frac{B}{\alpha^{\frac{3}{4}} (-x)^{\frac{1}{4}}} e^{-i \int_x^0 \alpha^{\frac{3}{2}} \sqrt{-x} dx}$$

evaluate integral by letting $y=-x$, doing the integral and then renaming y as x

$$\begin{aligned} &= \frac{A}{\alpha^{\frac{3}{4}} (-x)^{\frac{1}{4}}} e^{\frac{2i}{3} (-\alpha x)^{\frac{3}{2}}} + \frac{B}{\alpha^{\frac{3}{4}} (-x)^{\frac{1}{4}}} e^{-\frac{2i}{3} (-\alpha x)^{\frac{3}{2}}} \\ &= \frac{A}{\sqrt{\alpha} (-\alpha x)^{\frac{1}{4}}} e^{\frac{2i}{3} (-\alpha x)^{\frac{3}{2}}} + \frac{B}{\sqrt{\alpha} (-\alpha x)^{\frac{1}{4}}} e^{-\frac{2i}{3} (-\alpha x)^{\frac{3}{2}}} \end{aligned} \quad (18.19)$$

Region 2

$$\Psi_2 = \frac{D}{\sqrt{\beta}} e^{-\int_0^x \beta dx}$$

where $\beta = \alpha^{\frac{3}{2}} \sqrt{x}$

$$\begin{aligned} \Psi_2 &= \frac{D}{\alpha^{\frac{3}{4}} x^{\frac{1}{4}}} e^{-\int_0^x \alpha^{\frac{3}{2}} \sqrt{x} dx} \\ &= \frac{D}{\sqrt{\alpha} (\alpha x)^{\frac{1}{4}}} e^{-\frac{2}{3} (\alpha x)^{\frac{3}{2}}} \end{aligned} \quad (18.20)$$

Now at this point, compare these explicit wavefunctions to the patching wavefunctions in regions 1 and 2.

Region 2 ($x > 0$)($z > 0$)

$$\begin{aligned} \Psi_p &= aA_i(z) + bB_i(z) \\ &= \frac{a}{2\sqrt{\pi} z^{\frac{1}{4}}} e^{-\frac{2}{3} z^{\frac{3}{2}}} + \frac{b}{\sqrt{\pi} z^{\frac{1}{4}}} e^{\frac{2}{3} z^{\frac{3}{2}}} \end{aligned}$$

compared to

$$\Psi_2 = \frac{D}{\sqrt{\alpha} (\alpha x)^{\frac{1}{4}}} e^{-\frac{2}{3} (\alpha x)^{\frac{3}{2}}}$$

By inspection, one can tell immediately that

$$b = 0 \quad (18.21)$$

$$a = 2D \sqrt{\frac{\pi}{\alpha}} \quad (18.22)$$

Now go back and do the same comparison in region 1

Region 1 ($x < 0$)($z < 0$)

$$\begin{aligned}\Psi_p &= aA_i(z) \\ \Psi_p &= 2D\sqrt{\frac{\pi}{a}}A_i(z)\end{aligned}$$

where

$$\begin{aligned}A_i(z) &= \frac{1}{\sqrt{\pi}(-z)^{\frac{1}{4}}}\sin\left(\frac{2}{3}(-z)^{\frac{3}{2}} + \frac{\pi}{4}\right) \\ &= \frac{1}{\sqrt{\pi}(-z)^{\frac{1}{4}}}\left(\frac{e^{i(\frac{2}{3}(-z)^{\frac{3}{2}} + \frac{\pi}{4})} - e^{-i(\frac{2}{3}(-z)^{\frac{3}{2}} + \frac{\pi}{4})}}{2i}\right) \\ &= \frac{1}{(-z)^{\frac{1}{4}}}\left(\frac{e^{i\frac{\pi}{4}}}{2i\sqrt{\pi}}e^{i\frac{2}{3}(-z)^{\frac{3}{2}}} - \frac{e^{-i\frac{\pi}{4}}}{2i\sqrt{\pi}}e^{-i\frac{2}{3}(-z)^{\frac{3}{2}}}\right)\end{aligned}$$

yielding

$$\Psi_p = \frac{1}{(-z)^{\frac{1}{4}}}\left(\frac{De^{i\frac{\pi}{4}}}{i\sqrt{\alpha}}e^{\frac{2i}{3}(-z)^{\frac{3}{2}}} - \frac{De^{-i\frac{\pi}{4}}}{i\sqrt{\alpha}}e^{-\frac{2i}{3}(-z)^{\frac{3}{2}}}\right)$$

versus

$$\Psi_1 = \frac{1}{(-\alpha x)^{\frac{1}{4}}}\left(\frac{A}{\sqrt{\alpha}}e^{\frac{2i}{3}(-\alpha x)^{\frac{3}{2}}} + \frac{B}{\sqrt{\alpha}}e^{-\frac{2i}{3}(-\alpha x)^{\frac{3}{2}}}\right)$$

Now by inspection

$$A = -iDe^{i\frac{\pi}{4}} \quad (18.23)$$

$$B = iDe^{-i\frac{\pi}{4}} \quad (18.24)$$

Now put everything together for the WKB wavefunctions we showed at the beginning of the section

$$\begin{aligned}\Psi_1 &= \frac{A}{\sqrt{k}}e^{i\int_x^0 kdx} + \frac{B}{\sqrt{k}}e^{-i\int_x^0 kdx} \text{ where } (x < 0) \\ \Psi_2 &= \frac{D}{\sqrt{\beta}}e^{-\int_0^x \beta dx} \text{ where } (x > 0)\end{aligned}$$

where now $A = -iDe^{i\frac{\pi}{4}}$ and $B = iDe^{-i\frac{\pi}{4}}$ to get

$$\Psi_1 = \frac{2D}{\sqrt{k}}\sin\left(\frac{\pi}{4} + \int_x^0 kdx\right) \quad (18.25)$$

$$\Psi_2 = \frac{D}{\sqrt{\beta}}e^{-\int_0^x \beta dx} \quad (18.26)$$

Arbitrary potential drop

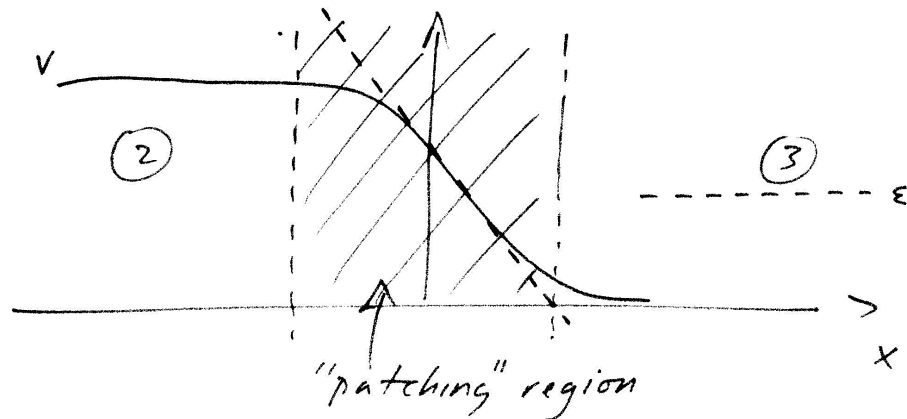


Figure 18.2: Cartoon of the arbitrary potential drop

Using our derived WKB wavefunctions we have the general form in regions 2 and 3

$$\Psi_2 = \frac{C}{\sqrt{\beta}} e^{\int_x^0 \beta dx} + \frac{D}{\sqrt{\beta}} e^{-\int_x^0 \beta dx} \quad \text{where } (x < 0)$$

$$\Psi_3 = \frac{F}{\sqrt{k}} e^{i \int_0^x k dx} + \frac{G}{\sqrt{k}} e^{-i \int_0^x k dx} \quad \text{where } (x > 0)$$

Since we can't have the wavefunctions behave badly, $C = 0$ to prevent it from blowing up in region 2. This leaves us with the wavefunction forms

$$\Psi_2 = \frac{D}{\sqrt{\beta}} e^{-\int_x^0 \beta dx} \quad (18.27)$$

$$\Psi_3 = \frac{F}{\sqrt{k}} e^{i \int_0^x k dx} + \frac{G}{\sqrt{k}} e^{-i \int_0^x k dx}$$

Now consider an approximation to the potential in the region between 2 and 3 of the linear form

$$V = (\varepsilon - V'x) \quad (18.28)$$

First we determine the patching wavefunctions in the region between 2 and 3. Starting with Schrodinger's equation (if you went through the last section

all of what follows will be familiar)

$$\begin{aligned} -\frac{\hbar^2}{2m} \frac{d^2\Psi_p}{dx^2} + V\Psi_p &= \varepsilon\Psi_p \\ -\frac{\hbar^2}{2m} \frac{d^2\Psi_p}{dx^2} + (\varepsilon - V'x)\Psi_p &= \varepsilon\Psi_p \\ -\frac{\hbar^2}{2m} \frac{d^2\Psi_p}{dx^2} - V'x\Psi_p &= 0 \\ \frac{d^2\Psi_p}{dx^2} + \frac{2mV'x}{\hbar^2}\Psi_p &= 0 \end{aligned}$$

Let $\alpha^3 = \frac{2mV'}{\hbar^2}$ leading to

$$\frac{d^2\Psi_p}{dx^2} + \alpha^3 x \Psi_p = 0$$

Now let $z = -\alpha x$

$$\frac{d^2\Psi_p}{dz^2} = z\Psi_p \quad (18.29)$$

Again, this is the Airy equation with accompanying Airy functions. The general solution is a linear combination of the form

$$\Psi_p = aA_i(z) + bB_i(z) \quad (18.30)$$

where a and b are constants.

In region 2

$$\beta = \sqrt{\frac{2m(V - \varepsilon)}{\hbar^2}}$$

where $V = \varepsilon - V'x$

$$\begin{aligned} &= \sqrt{\frac{2mV'}{\hbar^2}} \sqrt{-x} \\ &= \alpha^{\frac{3}{2}} \sqrt{-x} \end{aligned} \quad (18.31)$$

In region 3

$$\begin{aligned} k &= \sqrt{\frac{2m(\varepsilon - V)}{\hbar^2}} \\ &= \sqrt{\frac{2mV'}{\hbar^2}} \sqrt{x} \\ &= \alpha^{\frac{3}{2}} \sqrt{x} \end{aligned} \quad (18.32)$$

Now with β and k go back and make more explicit our wavefunctions in regions 2 and 3.

Region 2

$$\begin{aligned}\Psi_2 &= \frac{D}{\sqrt{\beta}} e^{-\int_x^0 \beta dx} \\ &= \frac{D}{\alpha^{\frac{3}{4}}(-x)^{\frac{1}{4}}} e^{-\alpha^{\frac{3}{2}} \int_x^0 (-x)^{\frac{1}{2}} dx}\end{aligned}$$

solve the integral by letting $y=-x$, integrating and then renaming $y=x$

$$= \frac{D}{\sqrt{\alpha}(-\alpha x)^{\frac{1}{4}}} e^{-\frac{2}{3}(-\alpha x)^{\frac{3}{2}}} \quad (18.33)$$

Region 3

$$\begin{aligned}\Psi_3 &= \frac{F}{\sqrt{k}} e^{i \int_0^x k dx} + \frac{G}{\sqrt{k}} e^{-i \int_0^x k dx} \\ &= \frac{F}{\sqrt{\alpha}(\alpha x)^{\frac{1}{4}}} e^{i \int_0^x \alpha^{\frac{3}{2}} x^{\frac{1}{2}} dx} + \frac{G}{\sqrt{\alpha}(\alpha x)^{\frac{1}{4}}} e^{-i \int_0^x \alpha^{\frac{3}{2}} x^{\frac{1}{2}} dx} \\ &= \frac{F}{\sqrt{\alpha}(\alpha x)^{\frac{1}{4}}} e^{\frac{2i}{3}(\alpha x)^{\frac{3}{2}}} + \frac{G}{\sqrt{\alpha}(\alpha x)^{\frac{1}{4}}} e^{-\frac{2i}{3}(\alpha x)^{\frac{3}{2}}}\end{aligned} \quad (18.34)$$

Now compare these explicit wavefunctions to the patching wavefunctions in regions 2 and 3

Region 2 ($x < 0$) ($z > 0$) since ($z = -\alpha x$)

$$\begin{aligned}\Psi_p &= aA_i(z) + bB_i(z) \\ &= \frac{a}{2\sqrt{\pi}z^{\frac{1}{4}}} e^{-\frac{2}{3}z^{\frac{3}{2}}} + \frac{b}{\sqrt{\pi}z^{\frac{1}{4}}} e^{\frac{2}{3}z^{\frac{3}{2}}}\end{aligned}$$

versus

$$\Psi_2 = \frac{D}{\sqrt{\alpha}(-\alpha x)^{\frac{1}{4}}} e^{-\frac{2}{3}(-\alpha x)^{\frac{3}{2}}}$$

By inspection

$$b = 0 \quad (18.35)$$

$$a = 2D\sqrt{\frac{\pi}{\alpha}} \quad (18.36)$$

Region 1 ($x > 0$) ($z < 0$) since ($z = -\alpha x$)

$$\begin{aligned}
 \Psi_p &= aA_i(z) \\
 &= 2D\sqrt{\frac{\pi}{\alpha}}A_i(z) \\
 &= \frac{a}{\sqrt{\pi}(-z)^{\frac{1}{4}}}\sin\left(\frac{2}{3}(-z)^{\frac{3}{2}} + \frac{\pi}{4}\right) \\
 &= \frac{a}{\sqrt{\pi}(-z)^{\frac{1}{4}}}\left(\frac{e^{i\frac{\pi}{4}}e^{\frac{2i}{3}(-z)^{\frac{3}{2}}} - e^{-i\frac{\pi}{4}}e^{-\frac{2i}{3}(-z)^{\frac{3}{2}}}}{2i}\right) \\
 &= \frac{1}{(-z)^{\frac{1}{4}}}\left(\frac{ae^{i\frac{\pi}{4}}}{\sqrt{\pi}2i}e^{\frac{2i}{3}(-z)^{\frac{3}{2}}} - \frac{ae^{-i\frac{\pi}{4}}}{\sqrt{\pi}2i}e^{-\frac{2i}{3}(-z)^{\frac{3}{2}}}\right)
 \end{aligned}$$

versus

$$\Psi_3 = \frac{1}{(\alpha x)^{\frac{1}{4}}}\left(\frac{F}{\sqrt{\alpha}}e^{\frac{2i}{3}(\alpha x)^{\frac{3}{2}}} + \frac{G}{\sqrt{\alpha}}e^{-\frac{2i}{3}(\alpha x)^{\frac{3}{2}}}\right)$$

By inspection

$$F = -iDe^{i\frac{\pi}{4}} \quad (18.37)$$

$$G = iDe^{-i\frac{\pi}{4}} \quad (18.38)$$

Now put everything together for the WKB wavefunctions we saw at the beginning of the section

$$\begin{aligned}
 \Psi_2 &= \frac{D}{\sqrt{\beta}}e^{-\int_x^0 \beta dx} \\
 \Psi_3 &= \frac{F}{\sqrt{k}}e^{i\int_0^x k dx} + \frac{G}{\sqrt{k}}e^{-i\int_0^x k dx}
 \end{aligned}$$

where now $F = -iDe^{i\frac{\pi}{4}}$ and $G = iDe^{-i\frac{\pi}{4}}$ leading to

$$\Psi_2 = \frac{D}{\sqrt{\beta}}e^{-\int_x^0 \beta dx} \quad (18.39)$$

$$\Psi_3 = \frac{2D}{\sqrt{k}}\sin\left(\frac{\pi}{4} + \int_0^x k dx\right) \quad (18.40)$$

Arbitrary potential barrier

This next section basically puts together all the work we did in the previous two sections for the arbitrary potential step and potential drop. The combination of both is the potential barrier.

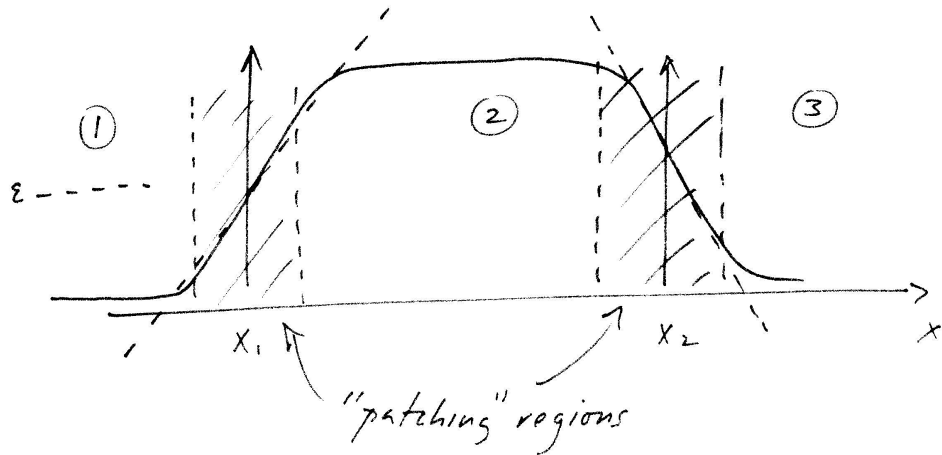


Figure 18.3: Cartoon of the arbitrary potential barrier

The solution for the transmission coefficient will lead to the general form of the tunneling probability that is ubiquitous in textbooks, in the literature and in experiments like scanning tunneling microscopy. The derivation involves a lot of tortuous math but follows the same train of thought as used in the earlier two sections.

Picture an arbitrary barrier of the form shown below. Now the general form of the WKB wavefunctions as derived earlier is

$$\Psi_1 = \frac{A}{\sqrt{k}} e^{i \int_x^{x_1} k dx} + \frac{B}{\sqrt{k}} e^{-i \int_x^{x_1} k dx} \text{ where } (x < x_1) \quad (18.41)$$

$$\Psi_2 = \frac{C}{\sqrt{\beta}} e^{\int_{x_1}^x \beta dx} + \frac{D}{\sqrt{\beta}} e^{-\int_{x_1}^x \beta dx} \text{ where } (x_1 < x < x_2) \quad (18.42)$$

$$\Psi_3 = \frac{F}{\sqrt{k}} e^{i \int_{x_2}^x k dx} \text{ where } (x > x_2) \quad (18.43)$$

Since there are no sources of particles on the right everything moves in a left to right direction and hence Ψ_3 has only one component above.

Remember in what follows that ultimately what we want to solve for is

$$T = \frac{v_3 |F|^2}{v_1 |A|^2} = \frac{k |F|^2}{k |A|^2} = \frac{|F|^2}{|A|^2}$$

You will find that our strategy is to express A in terms of F .

First we determine the patching wavefunction in the boundary region between 1 and 2. Starting with the Schrodinger equation

$$\begin{aligned} -\frac{\hbar^2}{2m} \frac{d^2\Psi_p}{dx^2} + V\Psi_p &= \varepsilon\Psi_p \\ -\frac{\hbar^2}{2m} \frac{d^2\Psi_p}{dx^2} + (\varepsilon + V'x)\Psi_p &= \varepsilon\Psi_p \\ -\frac{\hbar^2}{2m} \frac{d^2\Psi_p}{dx^2} + V'x\Psi_p &= 0 \\ \frac{d^2\Psi_p}{dx^2} - \frac{2mV'x}{\hbar^2}\Psi_p &= 0 \end{aligned}$$

Let $\alpha^3 = \frac{2mV'}{\hbar^2}$ resulting in

$$\frac{d^2\Psi_p}{dx^2} = \alpha^3 x \Psi_p$$

Now let $z = \alpha x$

$$\frac{d^2\Psi_p}{dz^2} = z\Psi_p$$

As before, this is the Airy equation whose general solution is a linear combination of Airy functions

$$\Psi_p = aA_i(z) + bB_i(z)$$

Now in the left part of region 2 (bordering region 1)

$$\begin{aligned} \beta &= \sqrt{\frac{2m(V - \varepsilon)}{\hbar^2}} \\ &= \sqrt{\frac{2mV'}{\hbar^2}}\sqrt{x} \\ &= \alpha^{\frac{3}{2}}\sqrt{x} \end{aligned} \tag{18.44}$$

In the right part of region 1 (bordering region 2)

$$\begin{aligned} k &= \sqrt{\frac{2m(\varepsilon - V)}{\hbar^2}} \\ &= \sqrt{\frac{2mV'}{\hbar^2}}\sqrt{-x} \\ &= \alpha^{\frac{3}{2}}\sqrt{-x} \end{aligned} \tag{18.45}$$

Now go back to the wavefunctions in regions 1 and 2 and make them more explicit using the above. Starting with

$$\begin{aligned}\Psi_1 &= \frac{A}{\sqrt{k}} e^{i \int_{x_1}^{x_1} k dx} + \frac{B}{\sqrt{k}} e^{-i \int_{x_1}^{x_1} k dx} \\ \Psi_2 &= \frac{C}{\sqrt{\beta}} e^{\int_{x_1}^x \beta dx} + \frac{D}{\sqrt{\beta}} e^{-\int_{x_1}^x \beta dx}\end{aligned}$$

temporarily shift the origin to x_1 (i.e. $x_1 = 0$). Then this basically becomes the same problem that we solved earlier. After we are done we can shift back the origin to where it used to be. First wavefunction

$$\Psi_1 = \frac{A}{\alpha^{\frac{3}{4}}(-x)^{\frac{1}{4}}} e^{i \int_x^0 \alpha^{\frac{3}{2}} \sqrt{-x} dx} + \frac{B}{\alpha^{\frac{3}{4}}(-x)^{\frac{1}{4}}} e^{-i \int_x^0 \alpha^{\frac{3}{2}} \sqrt{-x} dx}$$

solve integral by letting $y=-x$, integrate then rename $y=x$

$$= \frac{A}{\sqrt{\alpha}(-\alpha x)^{\frac{1}{4}}} e^{\frac{2i}{3}(-\alpha x)^{\frac{3}{2}}} + \frac{B}{\sqrt{\alpha}(-\alpha x)^{\frac{1}{4}}} e^{-\frac{2i}{3}(-\alpha x)^{\frac{3}{2}}} \quad (18.46)$$

Second wavefunction

$$\begin{aligned}\Psi_2 &= \frac{C}{\sqrt{\alpha}(\alpha x)^{\frac{1}{4}}} e^{\int_0^x \alpha^{\frac{3}{2}} \sqrt{x} dx} + \frac{D}{\sqrt{\alpha}(\alpha x)^{\frac{1}{4}}} e^{-\int_0^x \alpha^{\frac{3}{2}} \sqrt{x} dx} \\ &= \frac{C}{\sqrt{\alpha}(\alpha x)^{\frac{1}{4}}} e^{\frac{2}{3}(\alpha x)^{\frac{3}{2}}} + \frac{D}{\sqrt{\alpha}(\alpha x)^{\frac{1}{4}}} e^{-\frac{2}{3}(\alpha x)^{\frac{3}{2}}}\end{aligned} \quad (18.47)$$

Now compare these explicit wavefunctions with the patching wavefunction.

Left part of region 2 (bordering region 1) ($x > 0$) ($z > 0$)

$$\begin{aligned}\Psi_p &= aA_i(z) + bB_i(z) \\ &= \frac{a}{2\sqrt{\pi}z^{\frac{1}{4}}} e^{-\frac{2}{3}z^{\frac{3}{2}}} + \frac{b}{\sqrt{\pi}z^{\frac{1}{4}}} e^{\frac{2}{3}z^{\frac{3}{2}}}\end{aligned}$$

versus

$$\Psi_2 = \frac{C}{\sqrt{\alpha}(\alpha x)^{\frac{1}{4}}} e^{\frac{2}{3}(\alpha x)^{\frac{3}{2}}} + \frac{D}{\sqrt{\alpha}(\alpha x)^{\frac{1}{4}}} e^{-\frac{2}{3}(\alpha x)^{\frac{3}{2}}}$$

By inspection

$$b = C\sqrt{\frac{\pi}{\alpha}} \quad (18.48)$$

$$a = 2D\sqrt{\frac{\pi}{\alpha}} \quad (18.49)$$

Right part of region 1 (bordering region 2) ($x < 0$) ($z < 0$)

$$\begin{aligned}
 \Psi_p &= aA_i(z) + bB_i(z) \\
 &= 2D\sqrt{\frac{\pi}{\alpha}}A_i(z) + C\sqrt{\frac{\pi}{\alpha}}B_i(z) \\
 &= 2D\sqrt{\frac{\pi}{\alpha}}\frac{1}{\sqrt{\pi}(-z)^{\frac{1}{4}}}\sin\left(\frac{2}{3}(-z)^{\frac{3}{2}} + \frac{\pi}{4}\right) + C\sqrt{\frac{\pi}{\alpha}}\frac{1}{\sqrt{\pi}(-z)^{\frac{1}{4}}}\cos\left(\frac{2}{3}(-z)^{\frac{3}{2}} + \frac{\pi}{4}\right) \\
 &= \frac{2D}{\sqrt{\alpha}(-z)^{\frac{1}{4}}}\sin\left(\frac{2}{3}(-z)^{\frac{3}{2}} + \frac{\pi}{4}\right) + \frac{C}{\sqrt{\alpha}(-z)^{\frac{1}{4}}}\cos\left(\frac{2}{3}(-z)^{\frac{3}{2}} + \frac{\pi}{4}\right)
 \end{aligned}$$

skipping some steps

$$\begin{aligned}
 &= \frac{1}{\sqrt{\alpha}(-z)^{\frac{1}{4}}}\left(\frac{1}{2}\left(\frac{2D}{i} + C\right)\right)e^{i\left(\frac{\pi}{4} + \frac{2}{3}(-z)^{\frac{3}{2}}\right)} \\
 &+ \frac{1}{\sqrt{\alpha}(-z)^{\frac{1}{4}}}\left(\frac{1}{2}\left(-\frac{2D}{i} + C\right)\right)e^{-i\left(\frac{\pi}{4} + \frac{2}{3}(-z)^{\frac{3}{2}}\right)} \\
 &= \frac{1}{\sqrt{\alpha}(-z)^{\frac{1}{4}}}\left(\frac{e^{i\frac{\pi}{4}}}{2}\left(\frac{2D}{i} + C\right)\right)e^{\frac{2i}{3}(-z)^{\frac{3}{2}}} \\
 &+ \frac{1}{\sqrt{\alpha}(-z)^{\frac{1}{4}}}\left(\frac{e^{-i\frac{\pi}{4}}}{2}\left(-\frac{2D}{i} + C\right)\right)e^{-\frac{2i}{3}(-z)^{\frac{3}{2}}}
 \end{aligned}$$

compared to

$$\Psi_1 = \frac{A}{\sqrt{\alpha}(-\alpha x)^{\frac{1}{4}}}e^{\frac{2i}{3}(\alpha x)^{\frac{3}{2}}} + \frac{B}{\sqrt{\alpha}(-\alpha x)^{\frac{1}{4}}}e^{-\frac{2i}{3}(\alpha x)^{\frac{3}{2}}}$$

By inspection

$$A = \frac{e^{i\frac{\pi}{4}}}{2}\left(\frac{2D}{i} + C\right) \quad (18.50)$$

$$B = \frac{e^{-i\frac{\pi}{4}}}{2}\left(-\frac{2D}{i} + C\right) \quad (18.51)$$

Now consider the interface between regions 2 and 3. The potential here can be modeled in a linear fashion

$$V = (\varepsilon - V'x) \quad (18.52)$$

First determine the patching wavefunction in the boundary region. Start

with Schrodinger's equation again

$$\begin{aligned} -\frac{\hbar^2}{2m} \frac{d^2\Psi_p}{dx^2} + V\Psi_p &= \varepsilon\Psi_p \\ -\frac{\hbar^2}{2m} \frac{d^2\Psi_p}{dx^2} + (\varepsilon - V'x\Psi_p) &= \varepsilon\Psi_p \\ -\frac{\hbar^2}{2m} \frac{d^2\Psi_p}{dx^2} - V'x\Psi_p &= 0 \\ \frac{d^2\Psi_p}{dx^2} + \frac{2mV'x}{\hbar^2}\Psi_p &= 0 \end{aligned}$$

Let $\alpha^3 = \frac{2mV'}{\hbar^2}$ giving

$$\frac{d^2\Psi_p}{dx^2} + \alpha^3 x\Psi_p = 0$$

Now let $z = -\alpha x$ resulting in

$$\frac{d^2\Psi_p}{dz^2} = z\Psi_p$$

This is the Airy equation again with the general solution being a combination of Airy functions.

$$\Psi_p = aA_i(z) + bB_i(z)$$

Now in the right part of region 2 (bordering region 3)

$$\begin{aligned} \beta &= \sqrt{\frac{2m(V - \varepsilon)}{\hbar^2}} \\ &= \sqrt{\frac{2mV'}{\hbar^2}}\sqrt{-x} \\ &= \alpha^{\frac{3}{2}}\sqrt{-x} \end{aligned} \tag{18.53}$$

we recall $V = \varepsilon - V'x$

Now in the left part of region 3 (bordering region 2)

$$\begin{aligned} k &= \sqrt{\frac{2m(\varepsilon - V)}{\hbar^2}} \\ &= \sqrt{\frac{2mV'}{\hbar^2}}\sqrt{x} \\ &= \alpha^{\frac{3}{2}}\sqrt{x} \end{aligned} \tag{18.54}$$

Now go back to the wavefunctions in regions 2 and 3 and make them explicit.

$$\begin{aligned}\Psi_2 &= \frac{C}{\sqrt{\beta}} e^{\int_{x_1}^x \beta dx} + \frac{D}{\sqrt{\beta}} e^{-\int_{x_1}^x \beta dx} \\ \Psi_3 &= \frac{F}{\sqrt{k}} e^{i \int_{x_2}^x k dx}\end{aligned}$$

rearrange wavefunctions to get

$$\begin{aligned}\Psi_2 &= \frac{C}{\sqrt{\beta}} e^{\int_{x_1}^{x_2} \beta dx + \int_{x_2}^x \beta dx} + \frac{D}{\sqrt{\beta}} e^{-\int_{x_1}^{x_2} \beta dx - \int_{x_2}^x \beta dx} \\ \Psi_3 &= \frac{F}{\sqrt{k}} e^{i \int_{x_2}^x k dx}\end{aligned}$$

now let $\gamma = \int_{x_1}^{x_2} \beta dx$

$$\Psi_2 = \frac{C}{\sqrt{\beta}} e^{\gamma} e^{\int_{x_2}^x \beta dx} + \frac{D}{\sqrt{\beta}} e^{-\gamma} e^{\int_{x_2}^x \beta dx} \quad (18.55)$$

$$\Psi_3 = \frac{F}{\sqrt{k}} e^{i \int_{x_2}^x k dx} \quad (18.56)$$

Now temporarily translate the origin to x_2 (i.e. $x_2 = 0$). This leads to

$$\begin{aligned}\Psi_2 &= \frac{C}{\sqrt{\beta}} e^{\gamma} e^{\int_0^x \beta dx} + \frac{D}{\sqrt{\beta}} e^{-\gamma} e^{\int_0^x \beta dx} \\ \Psi_3 &= \frac{F}{\sqrt{k}} e^{i \int_0^x k dx}\end{aligned}$$

Now we use the derived explicit forms for β and k to get explicit forms for the wavefunctions. First wavefunction

$$\begin{aligned}\Psi_2 &= \frac{C}{\sqrt{\alpha}(-\alpha x)^{\frac{1}{4}}} e^{\gamma} e^{\int_0^x \alpha^{\frac{3}{2}} \sqrt{-x} dx} + \frac{D}{\sqrt{\alpha}(-\alpha x)^{\frac{1}{4}}} e^{-\gamma} e^{-\int_0^x \alpha^{\frac{3}{2}} \sqrt{-x} dx} \\ &\text{solve the integral by letting } y=-x, \text{ integrating, then rename } y=x \\ &= \frac{C e^{\gamma}}{\sqrt{\alpha}(-\alpha x)^{\frac{1}{4}}} e^{-\frac{2}{3}(-\alpha x)^{\frac{3}{2}}} + \frac{D e^{-\gamma}}{\sqrt{\alpha}(-\alpha x)^{\frac{1}{4}}} e^{\frac{2}{3}(-\alpha x)^{\frac{3}{2}}} \quad (18.57)\end{aligned}$$

Second wavefunction

$$\begin{aligned}\Psi_3 &= \frac{F}{\sqrt{\alpha}(\alpha x)^{\frac{1}{4}}} e^{i \int_0^x \alpha^{\frac{3}{2}} \sqrt{x} dx} \\ &= \frac{F}{\sqrt{\alpha}(\alpha x)^{\frac{1}{4}}} e^{\frac{2i}{3}(\alpha x)^{\frac{3}{2}}} \quad (18.58)\end{aligned}$$

Now compare these explicit wavefunctions to the patching wavefunction
Right side of region 2 (bordering region 3) ($x < 0$) ($z > 0$)

$$\begin{aligned}\Psi_p &= aA_i(z) + bB_i(z) \\ &= \frac{a}{2\sqrt{\pi}z^{\frac{1}{4}}}e^{-\frac{2}{3}z^{\frac{3}{2}}} + \frac{b}{\sqrt{\pi}z^{\frac{1}{4}}}e^{\frac{2}{3}z^{\frac{3}{2}}}\end{aligned}$$

compared to

$$\Psi_2 = \frac{Ce^\gamma}{\sqrt{\alpha}(-\alpha x)^{\frac{1}{4}}}e^{-\frac{2}{3}(-\alpha x)^{\frac{3}{2}}} + \frac{De^{-\gamma}}{\sqrt{\alpha}(-\alpha x)^{\frac{1}{4}}}e^{\frac{2}{3}(-\alpha x)^{\frac{3}{2}}}$$

By inspection

$$a = 2C\sqrt{\frac{\pi}{\alpha}}e^\gamma \quad (18.59)$$

$$b = D\sqrt{\frac{\pi}{\alpha}}e^{-\gamma} \quad (18.60)$$

Left side of region 3 (bordering region 2) ($x > 0$) ($z < 0$)

$$\begin{aligned}\Psi_p &= aA_i(z) + bB_i(z) \\ &= 2C\sqrt{\frac{\pi}{\alpha}}e^\gamma A_i(z) + D\sqrt{\frac{\pi}{\alpha}}e^{-\gamma} B_i(z) \\ &= \frac{2Ce^\gamma}{\sqrt{\alpha}(-z)^{\frac{1}{4}}}\sin\left(\frac{2}{3}(-z)^{\frac{3}{2}} + \frac{\pi}{4}\right) + \frac{De^{-\gamma}}{\sqrt{\alpha}(-z)^{\frac{1}{4}}}\cos\left(\frac{2}{3}(-z)^{\frac{3}{2}} + \frac{\pi}{4}\right) \\ &= \frac{2Ce^\gamma}{\sqrt{\alpha}(-z)^{\frac{1}{4}}}\left(\frac{e^{i\frac{\pi}{4}}e^{\frac{2i}{3}(-z)^{\frac{3}{2}}} - e^{-i\frac{\pi}{4}}e^{-\frac{2i}{3}(-z)^{\frac{3}{2}}}}{2i}\right) \\ &\quad + \frac{De^{-\gamma}}{\sqrt{\alpha}(-z)^{\frac{1}{4}}}\left(\frac{e^{i\frac{\pi}{4}}e^{\frac{2i}{3}(-z)^{\frac{3}{2}}} + e^{-i\frac{\pi}{4}}e^{-\frac{2i}{3}(-z)^{\frac{3}{2}}}}{2}\right) \\ &= \frac{1}{\sqrt{\alpha}(-z)^{\frac{1}{4}}}\left(\frac{1}{2}\left(\frac{2Ce^\gamma}{i} + De^{-\gamma}\right)e^{i\frac{\pi}{4}}\right)e^{\frac{2i}{3}(-z)^{\frac{3}{2}}} \\ &\quad + \frac{1}{\sqrt{\alpha}(-z)^{\frac{1}{4}}}\left(\frac{1}{2}\left(\frac{-2Ce^\gamma}{i} + De^{-\gamma}\right)e^{-i\frac{\pi}{4}}\right)e^{-\frac{2i}{3}(-z)^{\frac{3}{2}}}\end{aligned}$$

compare this to

$$\Psi_3 = \frac{F}{\sqrt{\alpha}(\alpha x)^{\frac{1}{4}}} e^{\frac{2i}{3}(\alpha x)^{\frac{3}{2}}}$$

By inspection

$$F = \frac{e^{i\frac{\pi}{4}}}{2} \left(\frac{2Ce^\gamma}{i} + De^{-\gamma} \right) \quad (18.61)$$

or more importantly

$$\frac{1}{2} \left(\frac{-2Ce^\gamma}{i} + De^{-\gamma} \right) e^{-i\frac{\pi}{4}} = 0$$

yielding

$$\boxed{C = \frac{iF}{2} e^{-i\frac{\pi}{4}} e^{-\gamma}} \quad (18.62)$$

$$\boxed{D = Fe^{-i\frac{\pi}{4}} e^\gamma} \quad (18.63)$$

Finally these expression can be inserted into our expression for A derived earlier.

$$\begin{aligned} A &= \frac{e^{i\frac{\pi}{4}}}{2} \left(\frac{2D}{i} + C \right) \\ &= iF \left(\frac{e^{-\gamma}}{4} - e^\gamma \right) \end{aligned} \quad (18.64)$$

Now the transmission coefficient is

$$\begin{aligned} T &= \frac{v_3|F|^2}{v_1|A|^2} = \frac{k|F|^2}{k|A|^2} = \frac{|F|^2}{|A|^2} = \frac{F^*F}{A^*A} \\ &= \frac{F^2}{F^2 \left(\frac{e^{-\gamma}}{4} - e^\gamma \right)^2} = \frac{1}{\left(\frac{e^{-\gamma}}{4} - e^\gamma \right)^2} \\ &= \frac{1}{e^{2\gamma} \left(\frac{e^{-2\gamma}}{4} - 1 \right)^2} = \frac{e^{-2\gamma}}{\left(\frac{e^{-2\gamma}}{4} - 1 \right)^2} \\ &= \frac{e^{-2\gamma}}{\left(1 - \frac{e^{-2\gamma}}{4} \right)^2} \end{aligned}$$

So after all this effort, the transmission coefficient has the form

$$T = \frac{e^{-2\gamma}}{\left(1 - \frac{e^{-2\gamma}}{4}\right)^2} \quad (18.65)$$

which in the case of large γ reduces to the ubiquitous textbook and literature expression

$$\boxed{T \approx e^{-2\gamma}} \quad (18.66)$$

where $\gamma = \int_{x_1}^{x_2} \beta dx$ and $\beta = \sqrt{\frac{2m(V-\varepsilon)}{\hbar^2}}$. No mas!

Example: Field emission

Also known as Fowler Nordheim tunneling. Assume a triangular form of the barrier where

$$(V - \varepsilon) = \phi \left(1 - \frac{x}{a}\right) \quad (18.67)$$

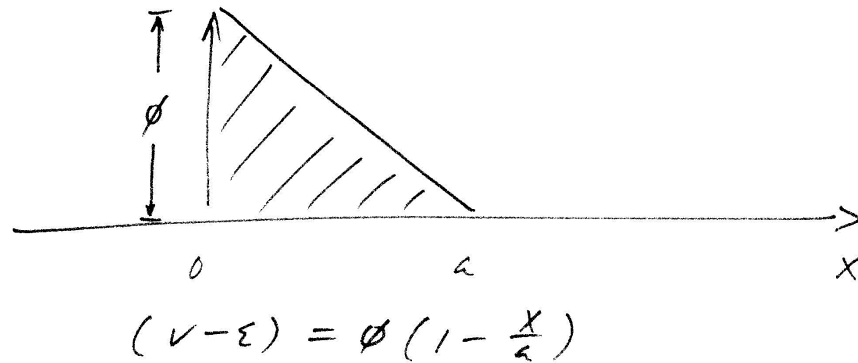


Figure 18.4: Sketch of the triangular barrier considered in field emission or Fowler Nordheim tunneling.

From the WKB approximation derived above

$$T \sim e^{-2\gamma}$$

where $\gamma = \int_0^a \sqrt{\frac{2m(V-\varepsilon)}{\hbar}} dx$. This leads to

$$\begin{aligned} T &= e^{-\frac{2}{\hbar} \sqrt{2m} \int_0^a \sqrt{V-\varepsilon} dx} \\ &= e^{-\frac{2}{\hbar} \sqrt{2m} \int_0^a \sqrt{\phi \left(1 - \frac{x}{a}\right)} dx} \\ &= e^{-\frac{2}{\hbar} \sqrt{2m} \int_0^a \sqrt{\phi - \frac{\phi x}{a}} dx} \end{aligned}$$

Integrate the exponent to get

$$\begin{aligned} -\frac{2}{\hbar} \sqrt{2m} \int_0^a \sqrt{\phi - \frac{\phi x}{a}} &= \frac{2}{\hbar} \sqrt{2m} \left(\frac{a}{\phi}\right) \left(\phi - \frac{\phi x}{a}\right)^{\frac{3}{2}} \frac{2}{3} \Big|_0^a \\ &= \frac{4}{3} \frac{\sqrt{2m}}{\hbar} \frac{a}{\phi} \phi^{\frac{3}{2}} \\ &= \frac{4}{3} \frac{\sqrt{2m}}{\hbar} \frac{1}{\vec{E}} \phi^{\frac{3}{2}} \end{aligned}$$

where $\vec{E} = \frac{\phi}{a}$ is the electric field. This leads to our final expression

$$\boxed{T = e^{-\frac{4}{3} \frac{\sqrt{2m}}{\hbar} \frac{\phi^{\frac{3}{2}}}{\vec{E}}}} \quad (18.68)$$

You can see that the tunneling probability depends to the 3/2 power of the barrier height. This is the underlying relation behind the Fowler Nordheim tunneling model.

Example: Schottky barrier

This problem arises when one has a metal-semiconductor junction. Here

$$(V - \varepsilon) \approx \phi \left(1 - \left(\frac{x}{a}\right)^2\right) \quad (18.69)$$

From the WKB approximation

$$T = e^{-2\gamma}$$

where $\gamma = \int_0^a \sqrt{\frac{2m(V-\varepsilon)}{\hbar}} dx$. This leads to

$$\begin{aligned} T &= e^{-\frac{2}{\hbar} \int_0^a \sqrt{2m\phi \left(1 - \left(\frac{x}{a}\right)^2\right)} dx} \\ &= e^{-\frac{2}{\hbar} \sqrt{2m\phi} \int_0^a \sqrt{1 - \left(\frac{x}{a}\right)^2} dx} \end{aligned}$$

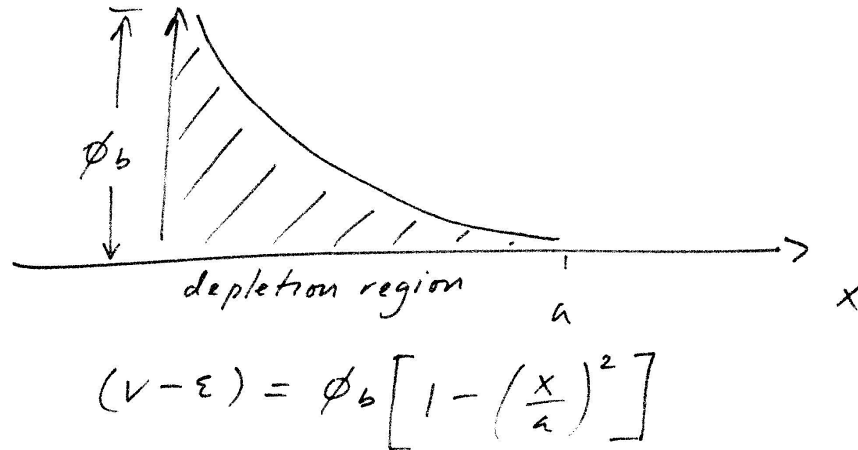


Figure 18.5: Sketch of the Shottky barrier

Let $y = \frac{x}{a}$ where also $dx = a dy$. Don't forget the limits of integration change as well.

$$\begin{aligned} T &= e^{-\frac{2}{\hbar} \sqrt{2m\phi} \int_0^1 \sqrt{1-y^2} a dy} \\ &= e^{-\frac{2}{\hbar} a \sqrt{2m\phi} \int_0^1 \sqrt{1-y^2} dy} \end{aligned}$$

To integrate, make another change of variables. Let $y = \sin(\theta)$; $dy = \cos(\theta) d\theta$

$$T = e^{-\frac{2}{\hbar} a \sqrt{2m\phi} \int_0^{\frac{\pi}{2}} \cos^2(\theta) d\theta}$$

where $\cos^2(\theta) = \frac{1}{2}(1 + \cos(2\theta))$

$$T = e^{-\frac{2}{\hbar} a \sqrt{2m\phi} \frac{1}{2} \int_0^{\frac{\pi}{2}} (1 + \cos(2\theta)) d\theta}$$

giving our final expression

$$\boxed{T = e^{-a \frac{\pi}{2} \sqrt{\frac{2m\phi}{\hbar^2}}}} \quad (18.70)$$

Here one notices that the tunneling probability varies as the square root of the barrier height, $\phi^{\frac{1}{2}}$.

Example: Symmetric parabolic barrier

Here

$$(V - \varepsilon) = C \left(\frac{a^2}{4} - x^2 \right) \quad (18.71)$$

where C is a constant with appropriate units.

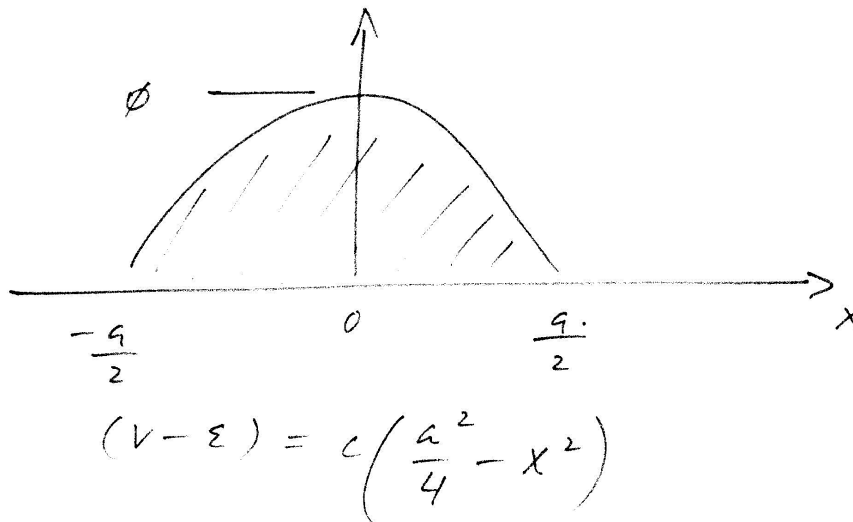


Figure 18.6: Sketch of an arbitrary symmetric parabolic barrier

From the WKB approximation, the tunneling probability through this barrier is

$$T = e^{-2\gamma}$$

where $\gamma = \int_0^a \sqrt{\frac{2m(V-\varepsilon)}{\hbar}} dx$. This leads to

$$T = e^{-\frac{2}{\hbar} \int_{-\frac{a}{2}}^{\frac{a}{2}} \sqrt{2mC \left(\frac{a^2}{4} - x^2 \right)} dx}$$

Evaluate the exponent

$$-\frac{2\sqrt{2mC}}{\hbar} \int_{-\frac{a}{2}}^{\frac{a}{2}} \sqrt{\frac{a^2}{4} - x^2} dx$$

Let $x = \frac{a}{2}\sin\theta$, $dx = \frac{a}{2}\cos\theta d\theta$. Also change the limits. The exponent becomes

$$\begin{aligned} & -\frac{2\sqrt{2mC}}{\hbar} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sqrt{\frac{a^2}{4} - \frac{a^2}{4}\sin^2\theta} \left(\frac{a}{2}\cos\theta\right) d\theta \\ &= -\frac{2\sqrt{2mC}}{\hbar} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \left(\frac{a}{2}\right)^2 \sqrt{1 - \sin^2\theta}(\cos\theta) d\theta \\ &= -\frac{2\sqrt{2mC}}{\hbar} \frac{a^2}{4} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^2\theta d\theta \\ &= -\frac{\sqrt{2mC}a^2}{2\hbar} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^2\theta d\theta \end{aligned}$$

where $\cos^2\theta = \frac{1}{2}(1 + \cos 2\theta)$

$$\begin{aligned} &= -\frac{\sqrt{2mC}a^2}{4\hbar} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} 1 + \cos 2\theta d\theta \\ &= -\frac{\sqrt{2mC}a^2\pi}{4\hbar} \\ &= -\frac{\sqrt{2m}\pi}{\hbar\sqrt{C}}\phi \end{aligned}$$

where ϕ the barrier height is $\phi = C \left(\frac{a^2}{4}\right)$. This leads to our final expression

$$\boxed{T = e^{-\sqrt{\frac{2m}{\hbar^2}} \frac{\pi}{\hbar} \phi}} \quad (18.72)$$

where the tunneling probability depends upon ϕ alone rather than to some power of it.

Exercises

1. Derive the transmission probability of a particle through a rectangular barrier of height V_o and width l .
2. Assume a rectangular barrier of height 4 eV and width 2 nm. For a free electron what is the tunneling probability through this barrier using a WKB formalism.
3. For the same 4 eV barrier and electron mass, what is the tunneling probability when the barrier width is increased by 50%. What is the probability when the barrier width is decreased by 50%.

4. For the same 2 nm wide barrier and electron mass, what is the tunneling probability when the barrier height is increased by 50%. What is the probability when the barrier height is decreased by 50%.
5. Use the WKB approximation to find the transmission coefficient for the potential $V(x) = 0$ if $x < 0$ and $V(x) = V_0 - kx$ when $x > 0$. Here V_0 and k are constants. Refer to figure 18.7. Hint: Find the turning points of the system where $V(x) = \epsilon$ to find the limits of integration.
6. Calculate the transmission coefficient of the following potential barrier using the WKB approximation. The potential is: $V(x) = V_0 \left(1 - \frac{x^2}{a^2}\right)$ when $-a \leq x \leq a$. Also $V(x) = 0$ elsewhere. See figure 18.8. Use the same hint about the turning points and the integration limits as described in the previous problem.
7. Consider the classic alpha decay problem using the just derived WKB approximation. The potential is $V(r) = \frac{2Ze^2}{4\pi\epsilon_0 r}$ for $r > r_1$ where r_1 is the radius of the nuclei. Carry the WKB integral (Equation 12.66) from r_1 to r_2 where $r_2 = \frac{2Ze^2}{4\pi\epsilon_0 E}$ and determine the tunneling probability. Make suitable approximations to simplify things as much as possible.

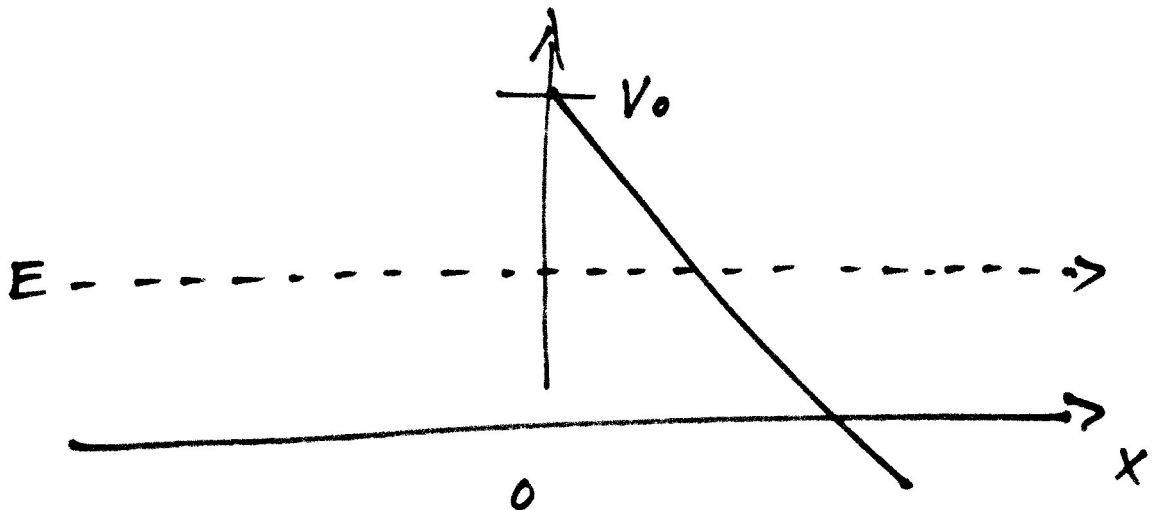


Figure 18.7: Sketch of a linearly decaying barrier

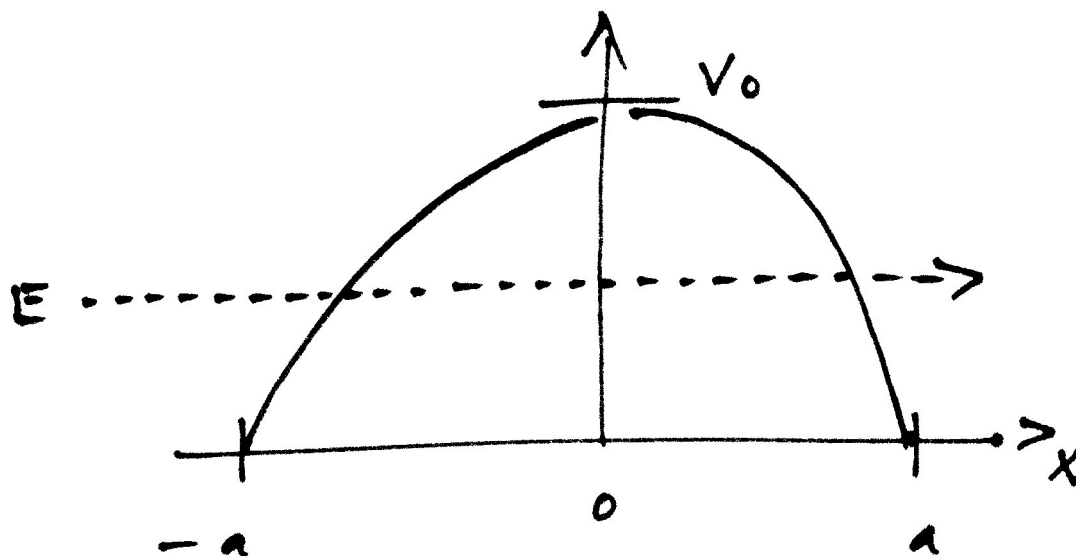


Figure 18.8: Sketch of a symmetric parabolic barrier

Relevant literature

These references are listed in no particular order

- “A carbon nanotube field-emission electron source”
W. A deHeer, A. Chatelain, D. Ugarte
Science, 270, 1179 (1995).
- “Field emission from carbon nanotubes: the first five years”
J-M Bonard, H. Kind, T. Stockli, L-O Nilsson
Solid-State Electronics, 45, 893 (2001).
- “Stripping voltammetry of Cu overlayers deposited on self-assembled monolayers: field emission of electrons through a phenylene ethylene oligomer”
M. S. Doescher, J. M. Tour, A. M. Rawlett, M. L. Myrick
J. Phys. Chem. B, 105, 105 (2001).
- “Time resolved photoelectrochemistry with size-quantized PbS adsorbed on gold”

E.P.A.M. Bakkers, J.J. Kelly, D. Vanmaekelbergh
Journal of Electroanalytical Chemistry, 482, 48 (2000).

- “Fluorescence intermittency in single InP quantum dots”
M. Kuno, D. P. Fromm, A. Gallagher, D. J. Nesbitt, O. I. Micic, A. J. Nozik
Nano Lett. 1, 557 (2001).
- “On/off fluorescence intermittency of single semiconductor quantum dots”
M. Kuno, D. P. Fromm, H. F. Hamann, A. Gallagher, D. J. Nesbitt
J. Chem. Phys. 115, 1028 (2001).

Chapter 19

Synthesis

Molecular Beam Epitaxy (MBE)

The basic idea behind this technique is fairly simple. In practice, however, its realization is more involved (and more expensive). MBE essentially consists of an ultrahigh vacuum chamber into which a substrate is loaded onto a heated sample holder. Precursors of desired elements (Ga, As, Al, P, In etc...) are then loaded into heated crucibles or furnaces (called Knudsen cells) outfitted with computer controlled shutters on their exits. The precursors are then heated such that when the shutters are opened one obtains a beam of atoms directed towards the substrate. Under such low pressures, the atomic species have very long mean free paths allowing them to reach the substrate without collisions with other gas phase species in the chamber. By controlling the temperature as well as the sequence/timing of opening and closing the shutters one can deposit very uniform films of semiconductor materials. In this fashion one can obtain precise nanometer lengthscale quantum well structures.

Under certain conditions where one deposits distinct semiconductor layers with different lattice constants, its also possible to grow islands of a semiconductor on top of another effectively allowing the synthesis of semiconductor quantum dots. The basic idea is that after a few layers of the new lattice mismatched semiconductor has been deposited, the strain at the interface changes the mode of growth from within the plane to out of the plane. Small islands are therefore formed which are the desired quantum dots. This technique is referred to as Stranski Krastanow growth.

A crude sketch of a MBE apparatus is shown below.

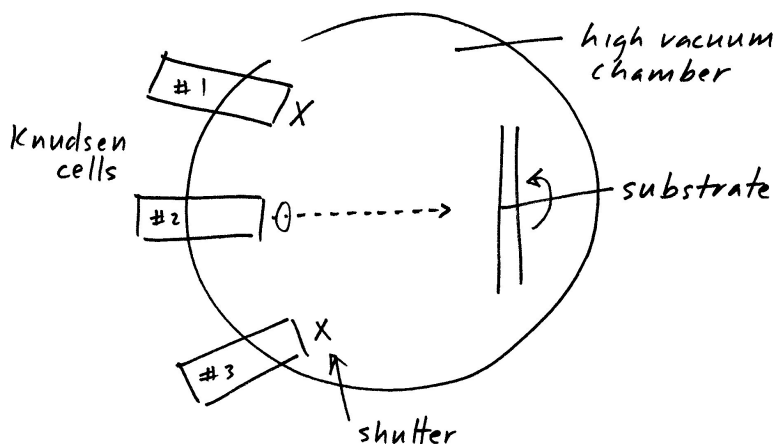


Figure 19.1: Cartoon of a MBE apparatus

Metal Organic Chemical Vapor Deposition (MOCVD)

MOCVD operates on a similar principle to MBE. However the apparatus differs greatly. MOCVD is generally conducted using a quartz tube furnace with a heated substrate. Clearly this is a much less expensive setup. One doesn't need turbo pumps, vacuum chambers, load locks, EELS guns and other MBE accoutrements. Organometallic or metal-organic compounds such as trimethylaluminum, trimethylgallium, trimethylindium as well as gases such as phosphine or arsine can be introduced into the heated reactor and allowed to decompose giving the desired elemental species. In the case of liquid organometallic precursors such as trimethylaluminum (above), the compound is brought into the gas phase by bubbling a carrier gas such as helium through the compound and allowing the gas to carry small amounts of the precursor into the reaction chamber. Like MBE, MOCVD can be used to grow thin films of materials. This technique has also been used, more recently, in the synthesis of semiconductor nanowires in the presence of gold nanoparticle catalysts. One potential disadvantage of the technique is the uniformity of the resulting films or deposition of materials. This is because the flow of gasses above the substrate may not be completely uniform. As a consequence one should consider ways to better control the flow and subsequent distribution of precursors over the substrate.

A sketch of the apparatus is shown below.

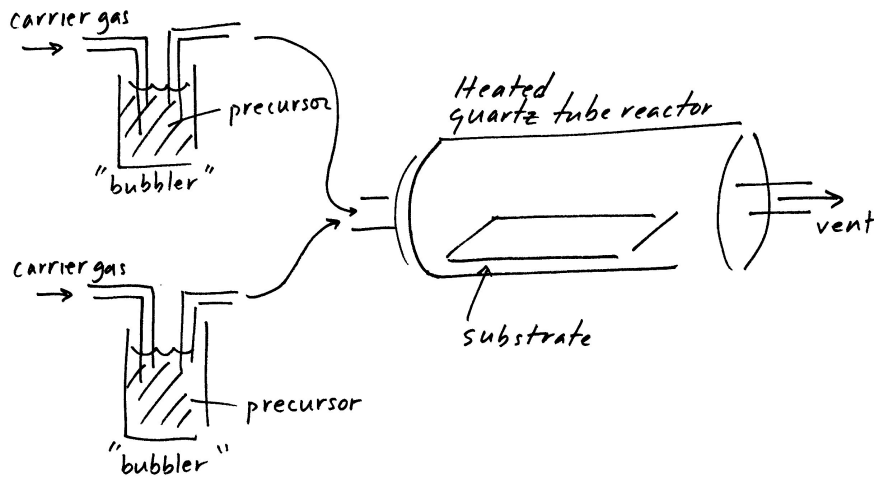


Figure 19.2: Cartoon of a MOCVD apparatus

Colloidal growth

Description

Colloidal growth is a chemist's approach to making nanostructures. There are different variations of this approach so no general description can fit all techniques. However a few are highlighted here. Some of the original quantum dots were created in supersaturated solutions within a glass matrix. Basically, molten silicate glasses were doped to the point of supersaturation with metal salts of the desired semiconducting material. The melt is then rapidly quenched, resulting in a precipitation of the desired semiconducting material into tiny seed nuclei (or alternatively the rapid quench can be thought of as a discrete temporal nucleation of seed particles). The glass solution then undergoes a secondary heat treatment with temperatures ranging from 400 to 1000 degrees Celcius. By varying the temperature and duration of the secondary heat treatment one can vary the average size of the nanocrystals.

Alternatively, others have employed inverse micelles as a means of quantum dot or even nanowire synthesis. In this preparation, surfactants such as bis(2-ethylhexyl)phosphate (also called AOT) are used to create small inverse micelles consisting of a hydrophilic interior and a hydrophobic exterior. Aqueous solutions of metal salts are then introduced into these compart-

mentalized water pools. Subsequent reactions are conducted in the aqueous phase (whether it be reduction of the metal precursor using sodium borohydride or reaction with a second chalcogen source such as S, Se or Te) to create metal or semiconductor nanoparticles. After the reaction, the nanocrystallites are sometimes extracted by adding a surface passivating agent which drops them out of solution. The recovered powder is then redissolved in a suitable solvent. To a first approximation, the average size of the nanocrystallites is determined by the initial ratio of water to surfactant, often called the “ W ” or “ Ω ” value.

One of the most successful approaches to the colloidal synthesis of nanocrystal and nanorods involves using coordinating as well as non coordinating solvents. In this approach organometallic precursors such as dimethylcadmium and trioctylphosphine selenide are injected into hot (temperatures on the order of 300 degrees Celcius) trioctylphosphine oxide. Upon injection, the precursors decompose to give desired elements of the final semiconductor. The rapid injection is analogous to the rapid quench in glasses and results in a discrete temporal nucleation of seed particles. The temperature of the solution is then slowly raised to allow the controlled growth of particles in the coordinating (or non-coordinating) solvent. This, again, is analogous to the secondary heat treatment with glasses with the main difference being lower overall temperatures (300 degrees versus 400 to 1000 degrees C). The average size of the nanomaterial is determined by the temperature and the duration of the heating and can be monitored spectroscopically.

A figure of such an apparatus is shown below.

LaMer and Dinegar growth model

Start with Fick’s first law.

$$\boxed{Q(t) = -4\pi r^2 D \frac{dc}{dr}} \quad (19.1)$$

where $Q(t)$ is the “flux” of stuff going to make the particle, D is the diffusion coefficient and r is a radial length from the center of the growing particle. We should point out that this “flux” has units of stuff/time as opposed to the more usual definition of flux. Now with this let’s manipulate the first expression a little bit to get.

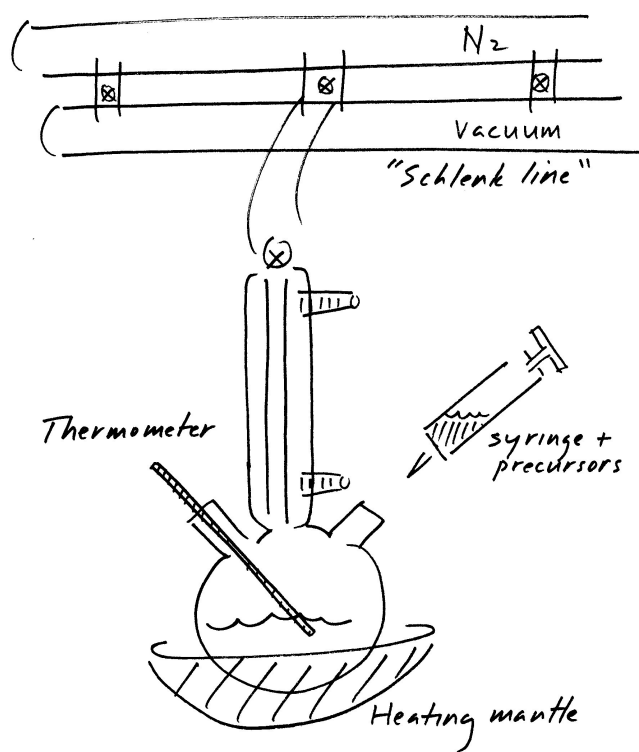


Figure 19.3: Cartoon of an apparatus commonly used for colloidal synthesis

$$\frac{Q(t)}{-4\pi r^2 D} = \frac{dc}{dr}$$

$$\frac{Q(t)dr}{-4\pi r^2 D} = dc$$

Integrate the last expression

$$\frac{Q(t)}{-4\pi D} \int \frac{dr}{r^2} = \int dc$$

to get

$$\frac{Q(t)}{4\pi D r} + Const = C(r, t)$$

This constant is a constant with respect to r . However it could, in general, depend on time. So more generally we get

$$\frac{Q(t)}{4\pi D r} + f(t) = C(r, t) \quad (19.2)$$

Apply boundary conditions now to make more explicit the expression. There are four to consider.

- When $r = a$, where a is the radius of the growing particle, $C(a, t) = C_s$. C_s is the saturation concentration of the stuff being deposited, called “monomer”.
- $C(r, t = 0) = C_{ss}$. Here C_{ss} is the supersaturation concentration of the solution at $t = 0$.
- $\frac{\partial C}{\partial t} \Big|_{r=h} = \frac{Q(t)}{\frac{4}{3}\pi h^3}$
- $Q(0) = 0$ Flux is zero at zero time as expected.

Apply boundary condition 2

$$C(r, t = 0) = C_{ss}$$

$$C(r, 0) = \frac{Q(0)}{4\pi D r} + f(0)$$

Since $Q(0) = 0$ (boundary condition 4)

$$C(r, 0) = f(0) = C_{ss} \quad (19.3)$$

Apply boundary condition 3

Start with the original expression

$$C(r, t) = \frac{Q(t)}{4\pi r D} + f(t)$$

take its time derivative

$$\left. \frac{\partial C(r, t)}{\partial t} \right|_{r=h} = \frac{1}{4\pi r D} \left. \frac{dQ(t)}{dt} \right|_{r=h} + \frac{df(t)}{dt} = \frac{Q(t)}{\frac{4}{3}\pi h^3}$$

or on rearranging

$$\frac{df(t)}{dt} = \frac{Q(t)}{\frac{4}{3}\pi h^3} - \frac{1}{4\pi h D} \frac{dQ(t)}{dt}$$

Now integrate this to get

$$f(t)|_0^t = \frac{1}{\frac{4}{3}\pi h^3} \int_0^t Q(t) dt - \frac{1}{4\pi h D} Q(t) \Big|_0^t$$

This becomes

$$f(t) - f(0) = \frac{1}{\frac{4}{3}\pi h^3} \int_0^t Q(t) dt - \frac{1}{4\pi h D} (Q(t) - Q(0))$$

where $f(0) = C_{ss}$ and $Q(0) = 0$. This reduces to

$$f(t) = C_{ss} + \frac{1}{\frac{4}{3}\pi h^3} \int_0^t Q(t) dt - \frac{Q(t)}{4\pi h D}$$

Since $C(r, t) = \frac{Q(t)}{4\pi r D} + f(t)$ we can use the above expression for $f(t)$ to get

$$\begin{aligned} C(r, t) &= \frac{Q(t)}{4\pi r D} + C_{ss} + \frac{1}{\frac{4}{3}\pi h^3} \int_0^t Q(t) dt - \frac{Q(t)}{4\pi h D} \\ &= \frac{Q(t)}{4\pi D} \left(\frac{1}{r} - \frac{1}{h} \right) + C_{ss} + \frac{1}{\frac{4}{3}\pi h^3} \int_0^t Q(t) dt \end{aligned}$$

Since in general $h \gg r$ we get

$$C(r, t) \approx \frac{Q(t)}{4\pi D} \left(\frac{1}{r} \right) + C_{ss} + \frac{1}{\frac{4}{3}\pi h^3} \int_0^t Q(t) dt$$

Now if

$$-Q(t) = 4\pi\rho a^2 \frac{da}{dt} \quad (19.4)$$

(a reasonable thought given that the flux in the opposite direction is reflected by the growth of the particle) one gets after replacing this in the previous expression, where ρ is the density of the material being deposited

$$C(r = a, t) \approx -\frac{4\pi\rho a^2 \frac{da}{dt}}{4\pi D a} + C_{ss} - \frac{4\pi\rho}{\frac{4}{3}\pi h^3} \int_0^t a^2 \frac{da}{dt} dt$$

or

$$C(r = a, t) \approx -\frac{\rho a \frac{da}{dt}}{D} + C_{ss} - \frac{3\rho}{h^3} \int_0^t a^2 \frac{da}{dt} dt$$

Integrate the last integral by parts. Let $u = a^2$, $du = 2ada$, $dv = \frac{da}{dt}$ and $v = a$. We get

$$\begin{aligned} C(r = a, t) &\approx -\frac{\rho a \frac{da}{dt}}{D} + C_{ss} - \frac{3\rho}{h^3} \left(a^3 \Big|_0^t - \int_0^t 2a^2 da \right) \\ &\approx -\frac{\rho a}{D} \left(\frac{da}{dt} \right) + C_{ss} - \frac{3\rho}{h^3} \left(a^3(t) - a^3(0) - \frac{2}{3} a^3 \Big|_0^t \right) \end{aligned}$$

Recall that at $t = 0$, $a(0) = 0$ Particle has not grown yet. This allows us to simplify the above expression to

$$\begin{aligned} &\approx -\frac{\rho a}{D} \left(\frac{da}{dt} \right) + C_{ss} - \frac{3\rho}{h^3} \left(a^3(t) - \frac{2}{3} (a^3(t) - a^3(0)) \right) \\ &\approx -\frac{\rho a}{D} \left(\frac{da}{dt} \right) + C_{ss} - \frac{3\rho}{h^3} \left(\frac{a^3(t)}{3} \right) \\ &\approx -\frac{\rho a}{D} \left(\frac{da}{dt} \right) + C_{ss} - \frac{\rho a^3}{h^3} \end{aligned}$$

Leading to our final expression

$$C(r = a, t) \approx C_{ss} - \frac{\rho a}{D} \left(\frac{da}{dt} \right) - \frac{\rho a^3}{h^3} \quad (19.5)$$

Apply boundary condition 1

Here apply it to the final expression we just derived above.

$$C(r = a, t) \approx C_{ss} - \frac{\rho a}{D} \left(\frac{da}{dt} \right) - \frac{\rho a^3}{h^3} = C_s$$

Rearrange to give

$$\frac{\rho a}{D} \left(\frac{da}{dt} \right) = (C_{ss} - C_s) - \frac{\rho a^3}{h^3}$$

or

$$a \frac{da}{dt} = \frac{D}{\rho} (C_{ss} - C_s) - \frac{Da^3}{h^3}$$

At this point note that $\frac{d(a^2)}{dt} = 2a \frac{da}{dt}$ so that $a \frac{da}{dt} = \frac{1}{2} \frac{d(a^2)}{dt}$. Insert this into the above expression to get

$$\frac{d(a^2)}{dt} = \frac{2D}{\rho} (C_{ss} - C_s) - \frac{2Da^3}{h^3} \quad (19.6)$$

This last last expression gives you the behavior for the size of the particle as a function of time, which is what we were ultimately after. We can proceed to solve this equation by either looking this up in a table of integrals, which is what LaMer and Dinegar ultimately did, or we could solve it numerically. Another option is to note that in general $\frac{a}{h} \ll 1$ so basically the last term in the expression drops out. We get the approximation

$$\frac{d(a^2)}{dt} \approx \frac{2D}{\rho} (C_{ss} - C_s) \quad (19.7)$$

You will notice that the radius of the particle with essentially grow as the square root of time ($a \propto \sqrt{t}$). Now following Sugimoto we invoke the Gibbs Thomson equation as follows

$$\begin{aligned} C_{ss} &= C_{\infty} e^{\frac{2\sigma V_m}{r^* RT}} \\ C_s &= C_{\infty} e^{\frac{2\sigma V_m}{a RT}} \end{aligned}$$

Here a is the radius of the particle at a given time. r^* is the radius of the particle that would make its ensemble concentration equivalent to the initial “monomer” concentration (think about this point, its will be important in a bit). To simplify the notation, let $r_o = \frac{2\sigma V_m}{RT}$

$$C_{ss} = C_\infty e^{\frac{r_o}{r^*}} \quad (19.8)$$

$$C_s = C_\infty e^{\frac{r_o}{a}} \quad (19.9)$$

If the exponent is small (and it doesn't have to be) then we can do a Taylor series expansion of the above to get (keeping only 1st 2 terms)

$$C_{ss} \approx C_\infty \left(1 + \frac{r_o}{r} + \dots\right)$$

$$C_s \approx C_\infty \left(1 + \frac{r_o}{a} + \dots\right)$$

Replace into our main equation to get

$$\frac{d(a^2)}{dt} \approx \frac{2D}{\rho} \left(C_\infty r_o \left(\frac{1}{r^*} - \frac{1}{a} \right) \right) \quad (19.10)$$

or alternatively

$$2a \frac{da}{dt} = \frac{2D}{\rho} C_\infty r_o \left(\frac{1}{r^*} - \frac{1}{a} \right)$$

$$\frac{da}{dt} = \frac{D}{\rho} C_\infty \left(\frac{r_o}{a} \right) \left(\frac{1}{r^*} - \frac{1}{a} \right)$$

Let

$$K = \frac{DC_\infty r_o}{\rho}$$

leading to

$$\boxed{\frac{da}{dt} = \frac{K}{a} \left(\frac{1}{r^*} - \frac{1}{a} \right)} \quad (19.11)$$

Now qualitatively speaking, just look at the sign of the right hand side, basically determined but the stuff in the parenthesis. You see if $a = r^*$ no growth occurs ($\frac{da}{dt} = 0$). If $a < r^*$ then $\frac{da}{dt}$ is negative and basically your

particles dissolve (negative growth). Now if $a > r^*$ then $\frac{da}{dt}$ is positive and your particles grow.

A graph of this equation is shown below. You will see that small particles have a steeper slope or faster rate of growth. Larger particles have a flatter slope so they grow slower.

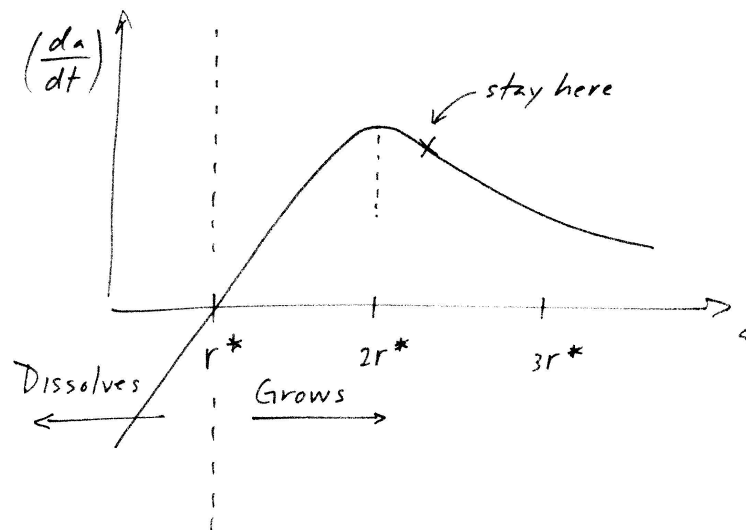


Figure 19.4: Sketch of the LaMer growth rate as a function of the critical radius r^*

Now one underlying point of this whole picture is that C_∞ is constant (i.e. r^* is constant). In reality though, since this is a closed system, the monomer concentration decreases as it gets consumed. So r^* is actually changing in time. If we think in dynamic terms as $t \rightarrow \infty$, $r^* \rightarrow \infty$ and one actually falls at different places on the above curve as a function of time. Basically your particle size will correspond to a steadily decreasing factor of r^* and you move to the left on the curve. If you have a situation where $\frac{da}{dt}$ becomes negative then your particles will start to dissolve. The worst case is that you have sizes that straddle r^* at any given time so that some particles grow, others dissolve and you end up with a very poor size distribution of your sample.

Size distribution

This section follows the work of Sugimoto who argued that the size distribution of the ensemble is proportional to

$$\text{size dist} \propto \frac{d\left(\frac{da}{dt}\right)}{da} \quad (19.12)$$

Using the relation derived earlier

$$\begin{aligned} \frac{d\left(\frac{da}{dt}\right)}{da} &= K \left(-\frac{1}{r^*a^2} + \frac{2}{a^3} \right) \left(\frac{da}{dt} \right) \\ &= \frac{K}{a^2} \left(\frac{2}{a} - \frac{1}{r^*} \right) \left(\frac{da}{dt} \right) \end{aligned}$$

where

$$\left(\frac{da}{dt} \right) = \left(\frac{da}{dt} \right) \left(\frac{dt}{da} \right)$$

Divide both sides by $\frac{dt}{da}$ to get

$$\begin{aligned} \left(\frac{da}{dt} \right) \frac{d\left(\frac{da}{dt}\right)}{da} &= \frac{K}{a^2} \left(\frac{2}{a} - \frac{1}{r^*} \right) \left(\frac{da}{dt} \right) \\ \frac{d\left(\frac{da}{dt}\right)}{dt} &= \frac{K}{a^2} \left(\frac{2}{a} - \frac{1}{r^*} \right) \left(\frac{da}{dt} \right) \end{aligned}$$

Let $\sigma = \frac{da}{dt}$ giving

$$\frac{d\sigma}{dt} = \frac{K}{a^2} \left(\frac{2}{a} - \frac{1}{r^*} \right) \sigma \quad (19.13)$$

Alternatively

$$\boxed{\frac{d\sigma}{dt} = \frac{K}{a^2} \left(\frac{2r^* - a}{r^*a} \right) \sigma} \quad (19.14)$$

Now its clear to see that if $a > 2r^*$, $\left(\frac{d\sigma}{dt}\right)$ is negative valued. The size distribution of the sample will narrow. This is called the “focusing” regime.

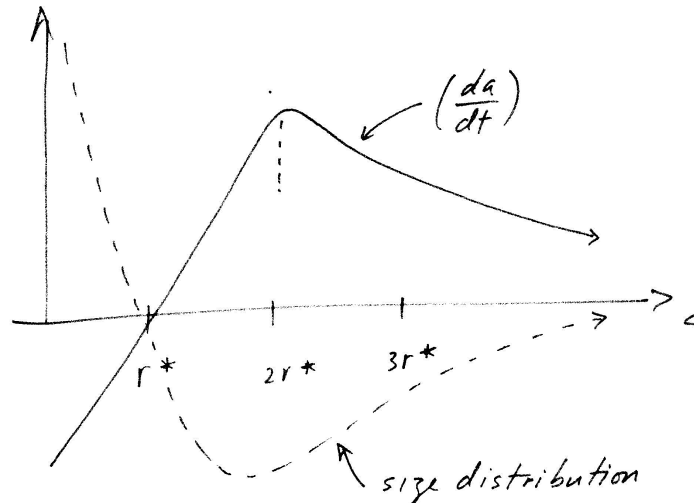


Figure 19.5: Sketch of both the LaMer growth rate and size distribution as a function of the critical radius r^*

However, if $a < 2r^* \left(\frac{da}{dt}\right)$ is positive valued and the size distribution will increase. This is called the “defocusing” regime.

Recall our discussion earlier about the dynamic nature of r^* . It goes without saying that in a closed system as time increases r^* will also increase. The actual size of the particle will in absolute terms become a smaller fraction of r^* causing you to progressively move left on our diagram. As you move left the distribution will broaden. To keep the size distribution or even narrow it you need to fight the depletion of monomers causing r^* to increase. This means as the reaction progresses you add more precursor to the reaction. The extra addition can be accomplished a number of ways. Peng for example just adds subsequent injection of precursors into the reaction mixture. Sugimoto and others however build in a reservoir of precursors that slowly get released to the reaction as time increases.

Reaction controlled growth

The previous discussion has assumed diffusion limited growth. However it's possible to have the reaction controlled situation. LaMer, Sugimoto and others have shown that in this case the relevant growth equation can be expressed as

$$\frac{da}{dt} = K_r \left(\frac{1}{r^*} - \frac{1}{a} \right) \quad (19.15)$$

where K_r is a constant. To see how the size distribution behaves in this situation repeat the analysis done earlier for the diffusion controlled growth case.

$$\text{size dist} \propto \frac{d\left(\frac{da}{dt}\right)}{da}$$

When explicitly evaluated this leads to

$$\frac{d\left(\frac{da}{dt}\right)}{da} = \frac{K_r}{a^2} \left(\frac{da}{da}\right)$$

where recall that

$$\left(\frac{da}{da}\right) = \left(\frac{da}{dt}\right) \left(\frac{dt}{da}\right)$$

Replace into our expression and divide by $\left(\frac{dt}{da}\right)$ on both sides. This gives

$$\begin{aligned} \frac{da}{dt} \frac{d\left(\frac{da}{dt}\right)}{da} &= \frac{K_r}{a^2} \left(\frac{da}{dt}\right) \\ \frac{d\left(\frac{da}{dt}\right)}{dt} &= \frac{K_r}{a^2} \left(\frac{da}{dt}\right) \end{aligned}$$

Let $\sigma = \frac{da}{dt}$ giving

$$\frac{d\sigma}{dt} = \frac{K_r}{a^2} \sigma \quad (19.16)$$

You will notice here that this expression is always positive. In effect it tells us that there will always be some broadening of the size distribution during the particle growth.

Exercises

1. Read any of the papers listed below or select a paper from the current literature and discuss it.

Relevant reading

1. "Theory, production and mechanism of formation of monodispersed hydrosols" V. K. LaMer, R. H. Dinegar *J. Am. Chem. Soc.* 72, 4847 (1950).
2. "Preparation of monodispersed colloidal particles" T. Sugimoto *Advances in colloid and interface science*, 28, 65 (1987).
3. "The kinetics of precipitation from supersaturated solid solutions" I. M. Lifshitz, V. V. Slyozov *J. Phys. Chem. Solids*, 19, 35 (1961).

References directly related to LaMer and Sugimoto models

These are some papers from the literature that deal with the LaMer/Dinegar and Sugimoto models discussed above. The papers are in no particular order.

1. "Formation of high quality CdS and other II-IV semiconductor nanocrystals in noncoordinating solvents: Tunable reactivity of monomers" W. W. Yu and X. Peng *Angew. Chem. Int. Ed.* 41, 2368 (2002).
2. "Kinetics of II-VI and III-V colloidal semiconductor nanocrystal growth: Focusing of size distributions" X. Peng, J. Wickham and A. P. Alivisatos *J. Am. Chem. Soc.* 120, 5343 (1998).
3. "Nearly monodisperse and shape-controlled CdSe nanocrystals via alternative routes: Nucleation and growth" Z. A. Peng and X. Peng *J. Am. Chem. Soc.* 124, 3343 (2002).
4. "Formation of high quality CdTe, CdSe, and CdS nanocrystals using CdO as precursor" Z. A. Peng and X. Peng *J. Am. Chem. Soc.* 123, 183 (2001).
5. "Formation of high quality InP and InAs nanocrystals in a noncoordinating solvent"

D. Battaglia and X. Peng
Nano Letters, 2, 1027 (2002).

6. “The kinetics of growth of semiconductor nanocrystals in a hot amphiphile matrix”
C. D. Dushkin, S. Saita, Y. Yoshie, Y. Yamaguchi
Advances in Colloid and Interface Science, 88, 37 (2000).
7. “Evolution of an ensemble of nanoparticles in a colloidal solution: Theoretical study” D. V. Talapin, A. L. Rogach, M. Haase, H. Weller J. Phys. Chem. B 105, 12278 (2001).

Relevant literature

The following papers describe some of the syntheses for nanoscale materials. They are listed in no particular order.

- “Large-scale production of single-walled carbon nanotubes by the electric-arc technique”
C. Journet et al.
Nature 388, 756 (1997).
- “Large scale CVD synthesis of single-walled carbon nanotubes”
A. M. Cassell, J. A. Raymakers, J. Kong, H. Dai
J. Phys. Chem. B 103, 6484 (1999).
- “Chemical vapor deposition of methane for single-walled carbon nanotubes”
J. Kong, A. M. Cassell, H. Dai
Chem. Phys. Lett. 292, 567 (1998).
- “General synthesis of compound semiconductor nanowires”
X. Duan, C. M. Lieber
Advanced Materials, 12, 298 (2000).
- “A laser ablation metode for the synthesis of crystalline semiconductor nanowires”
A. M. Morales, C. M. Lieber
Science, 279, 208 (1998).

- “Epitaxial core-shell and core-multishell nanowire heterostructures”
L. J. Lauhon, M. S. Gudiksen, D. Wang, C. M. Lieber
Nature, 420, 57 (2002).
- “Inorganic semiconductor nanowires”
Y. Wu, H. Yan, M. Huang, B. Messer, J. H. Song, P. Yang
Chem. Eur. J. 8, 1261 (2002).
- “Direct observation of vapor-liquid-solid nanowire growth”
Y. Wu, P. Yang
J. Am. Chem. Soc. 123, 3165 (2001).
- “High quality GaN nanowires synthesized using a CVD approach”
J. C. Wang, S. Q. Feng, D. P. Yu
Appl. Phys. A 75, 691 (2002).
- “Antimony nanowire arrays fabricated by pulsed electrodeposition in anodic alumina membranes”
Y. Zhang, G. Li, Y. Wu, B. Zhang, W. Song, L. Zhang
Adv. Mater. 14, 1227 (2002).
- “Silicon nanotubes”
J. Sha, J. Niu, X. Ma, J. Xu, X. Zhang, Q. Yang, D. Yang
Adv. Mater. 14, 1219 (2002).
- “Silicon nanowires: preparation, device fabrication, and transport properties”
J-Y Yu, S-W Chung, J. R. Heath
J. Phys. Chem. B 104, 11864 (2000).
- “Diameter-controlled synthesis of carbon nanotubes”
C. L. Cheung, A. Kurtz, H. Park, C. M. Lieber
J. Phys. Chem. B 106, 2429 (2002).
- “Superlattices of platinum and palladium nanoparticles”
J. E. Martin, J. P. Wilcoxon, J. Odinek, P. Provencio
J. Phys. Chem. B 106, 971 (2002).
- “Some recent advances in nanostructure preparation from gold and silver particles: a short topical review”
M. Brust, C. J. Kiely
Colloids and Surfaces A: Physicochemical and Engineering Aspects, 202, 175 (2002).

- “X-ray photoelectron spectroscopy of CdSe nanocrystals with applications to studies of the nanocrystal surface”
J. E. Bowen Katari, V. L. Colvin, A. P. Alivisatos
J. Phys. Chem. B 98, 4109 (1994).
- “Colloidal nanocrystal shape and size control: the case of cobalt”
V. F. Puentes, K. M. Krishnan, A. P. Alivisatos
Science, 291, 2115 (2001).
- “Synthesis, self-assembly and magnetic behavior of a two dimensional superlattice of single-crystal ϵ -Co nanoparticles”
V. F. Puentes, K. M. Krishnan, A. P. Alivisatos
Appl. Phys. Lett. 78, 2187 (2001).
- “Monodisperse FePt nanoparticles and ferromagnetic FePt nanocrystal superlattices”
S. Sun, C. B. Murray, D. Weller, L. Folks, A. Moser
Science, 287, 1989 (2000).
- “Water-in-oil microemulsion synthesis of platinum-ruthenium nanoparticles, their characterization and electrocatalytic properties”
X. Zhang, K-Y Chan
Chem. Mater. 15, 451 (2003).
- “Synthesis and characterization of monodisperse nanocrystals and close-packed nanocrystal assemblies”
C. B. Murray, C. R. Kagan, M. G. Bawendi
Annu. Rev. Mater. Sci. 30, 545 (2000).

Chapter 20

Tools

Electron microscopies

Transmission electron microscopy

In the TEM experiment, a thin or diluted sample is bombarded under high vacuum with a focused beam of electrons. Electrons that are transmitted through the material form contrast patterns that reproduce the image of the sample. This pattern arises from the scattering of electrons off of atoms composing the sample. In addition, diffracted electrons give information about the lattice structure of the material. In the case of nanocrystallites, analysis of TEM images is partially responsible for the sizing curves of colloidal quantum dots. The shape of the sample can also be determined from the image.

Secondary electron microscopy

In the SEM experiment, an electron beam is focused and raster scanned over the sample. When the incident electrons interact with the sample a number of effects take place, such as the emission of secondary electrons. These effects are highly localized to the region directly under the electron beam and can be used to create an image of the sample. In addition, elemental analysis through energy dispersive or wavelength dispersive techniques can be done using other detectors.

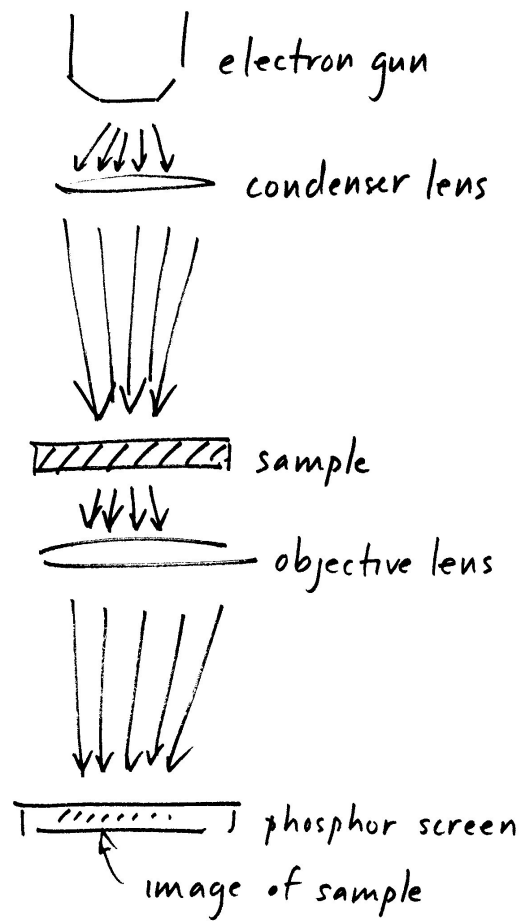


Figure 20.1: Cartoon showing the TEM technique

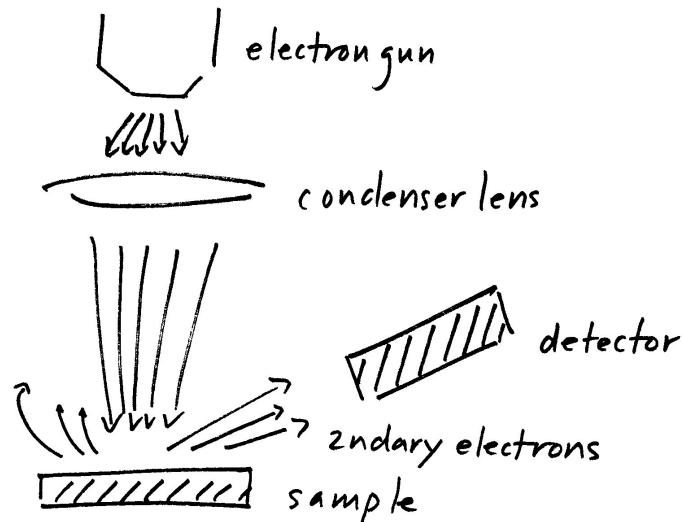


Figure 20.2: Cartoon showing the SEM technique

Probe microscopies

Atomic force microscopy

The atomic force experiment works by using a sharp silicon or silicon nitride cantilever. This tip is brought close to the surface of the sample in question, close enough that atomic forces occur between tip and sample. These forces can be either repulsive or attractive. In the repulsive mode or contact/intermittent contact (tapping mode) mode of operation, the tip position over the surface of the sample is kept constant through a feedback mechanism. In the attractive regime or non contact mode, attractive forces bend the tip bringing it closer to the sample which in turn is detected through a number of means. In any of the modes, the attractive or repulsive forces plus response of the system to counteract them are used as a means of generating a topographic image of the sample. The AFM has been used to move individual nanostructures. It has also been used as a tool to “scratch” surfaces and make nanoscale patterns. There are other variations of this probe technique such as magnetic force microscopy (MFM) and electrostatic force microscopy (EFM).

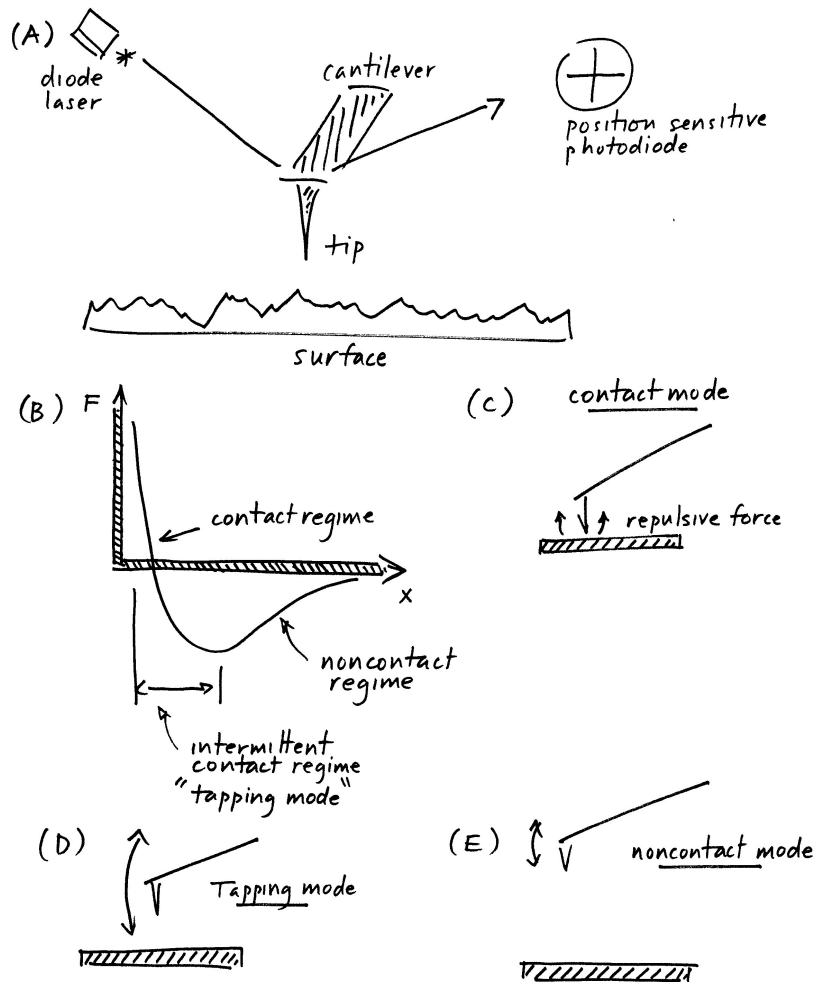


Figure 20.3: Cartoon showing the AFM technique. (A) General concept behind the technique. (B) Force distance curve. Parts of the curve below zero are in the attractive realm. Parts of the curve above zero are in the repulsive regime. (C) Contact mode AFM operating with tip very close to the sample, repulsive regime. (D) Tapping mode AFM. Tip is oscillated with large peak to peak amplitude resulting in brief forrays into the repulsive regime. Hence the name intermittent contact mode. Tapping mode is a Digital Instruments trademarked name. (E) Non-contact mode AFM. The tip is oscillated with small peak to peak amplitude. Operated exclusively in the attractive part of the potential.

Scanning tunneling microscopy

Scanning tunneling microscopy was the original probe microscopy. It was developed by Gerd Binnig and Heinrich Rohrer at IBM Zurich, ultimately leading to the Nobel Prize. The principle of operation is the tunneling of electron from a conductive tip to a conductive substrate or sample through a barrier. The tunneling current is found to be exponentially dependent upon the tip to sample separation allowing for very high sensitivity of sample height. STM has been used to investigate a number of nano related effects such as the discrete atomic like states of colloidal quantum dots and other systems. It has also been used to investigate the coulomb blockade and coulomb staircase phenomena. The STM has also been used to manipulate individual atoms as described in recent work on quantum corrals. One disadvantage of the STM is that it requires conductive samples or relatively conductive samples on a conductive substrate. To circumvent this limitation, the AFM was subsequently developed.

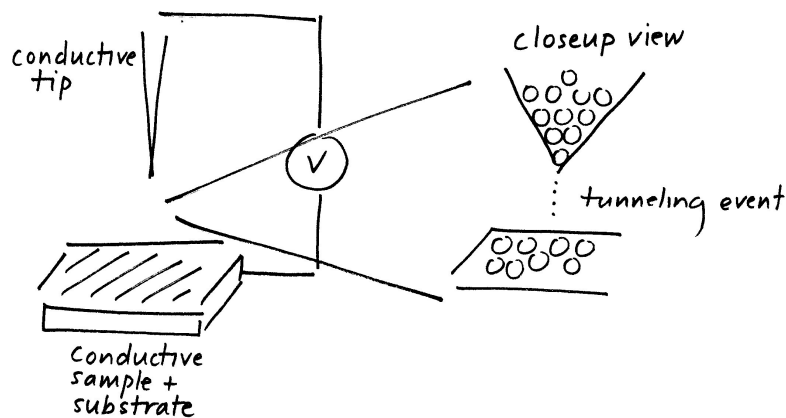


Figure 20.4: Cartoon showing the STM technique

Dip pen nanolithography

Dip-pen nanolithography is a recent atomic force microscopy based technique developed by Chad Mirkin at Northwestern university. The essential idea of the technique is to use the AFM tip as a quill pen. Dip or coat it with a molecular substance. Upon close approach to a substrate the mole-

cular “ink” rolls off the tip and comes into contact with the substrate. By scanning the tip one can pattern the substrate with a layer of molecules. Advantages of dip-pen over other patterning techniques is that it potentially has a very high resolution limited only by the AFM tip radius of curvature. The main disadvantage of the process is that it is serial in nature and hence patterning large areas may prove time consuming. A cartoon describing the technique is shown below.

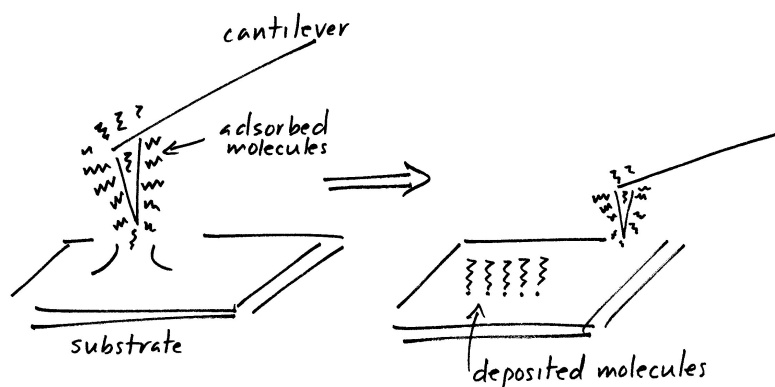


Figure 20.5: Cartoon showing the dip pen technique

Microcontact printing

Microcontact printing is a stamping technique developed by George Whitesides at Harvard University. The basis of the technique is to use a precursor of a polymer called polydimethylsiloxane (PDMS) which can be poured over a master pattern. This master pattern is created through standard photolithography and basically consists of raised features or islands in the desired pattern. The precursor can then be converted to PDMS and hardened to make a negative image of the original master. Then the PDMS “stamp” can be inked with molecular compounds and applied to a surface such as a thin film of gold. The molecular ink, thiols for example, are left behind on the substrate and reproduce the original master. The idea is much the same as with dip pen nanolithography, however, the microcontact printing is a serial process whereas dip-pen is a serial technique and is much slower. One of the disadvantages of microcontact printing, however,

is that it lacks the resolution of dip-pen which is ultimately limited only by the tip radius of curvature 1-10 nm. However, recent reports show that microcontact printing can be pushed to a resolution around 50 nm.

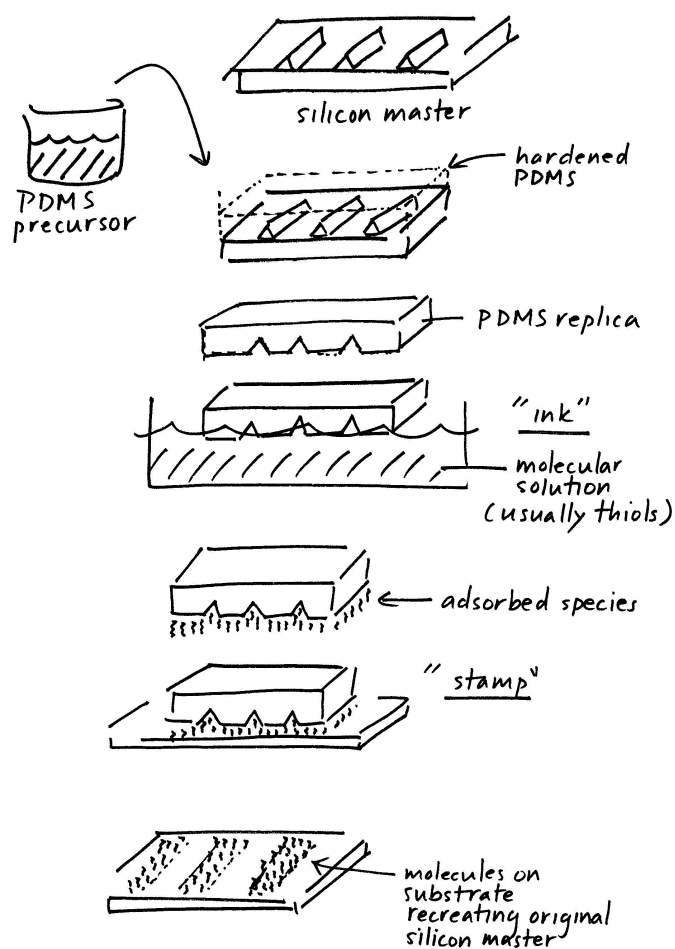


Figure 20.6: Cartoon showing the microcontact printing technique and general sequence of steps.

Exercises

1. Read any of the papers listed below or select a paper from the current literature and discuss it.

Relevant literature

The following papers are in no particular order

- “Directed assembly of one-dimensional nanostructures into functional networks”
Y. Huang, X. Duan, Q. Wei, C. M. Lieber
Science, 291, 630 (2001).
- “Submicrometer patterning of charge in thin-film electrets”
H. O. Jacobs, G. M. Whitesides
Science, 291, 1763 (2001).
- “Patterning self assembled monolayers: applications in materials science”
A. Kumar, H. A. Biebuyck, G. M. Whitesides
Langmuir, 10, 1498 (1994).
- “Generation of 30-50 nm structures using easily fabricated, composite PDMS masks”
T. W. Odom, V. R. Thalladi, J. C. Love, G. M. Whitesides
J. Am. Chem. Soc. 124, 12112 (2002).
- “Moving beyond molecules: patterning solid-state features via dip-pen nanolithography with sol-based inks”
M. Su, X. Liu, S-Y Li, V. P. Dravid, C. A. Mirkin
J. Am. Chem. Soc. 124, 1560 (2002).
- “Direct patterning of modified oligonucleotides on metals and insulators by dip-pen nanolithography”
L. M. Demers, D. S. Ginger, S-J. Park, Z. Li, S-W. Chung, C. A. Mirkin
Science, 296, 1836 (2002).
- “The art of building small”
G. M. Whitesides, C. J. Love
Scientific American 285, 38 (2001).

Chapter 21

Applications

As described in the introduction, nanostructures, whether it be quantum dots, wires or wells, have interesting size dependent optical and electrical properties. The study of these intrinsic properties is the realm of nanoscience. However, at the end of the day, we expect that some of this acquired knowledge (funded largely through our tax dollars) will be put to good use for developing next generation consumer products. So how exactly are today's nanotechnologists trying to harness the potential of nano?

Since there are almost too many applications of nano to catalog here, this section is not meant to be comprehensive. However, we briefly touch upon some applications of quantum wells, quantum wires and quantum dots that are seen in the current literature.

Nanowires

We begin with a short discussion about applications of nanowires. Devices using these low dimensional materials have not been made to any great extent. This is because the historical development of nanostructures seems to have skipped nanowires, moving from wells to dots first. More recently, though, researchers have learned how to make asymmetric nanowires using a number of approaches including vapor-liquid-solid (VLS) and solution-liquid-solid (SLS) growth. The move to applications has occurred quickly with the key selling point being that, in addition to exhibiting quantum confinement effects, nanowires are at the same time (as their name implies) wires. This means that making electrical connections to the outside world and assembling actual devices may be a lot easier than with other nanostructures such as quantum dots.

Crossed nanowire junctions have been made, using p-type and n-type wires. These junctions, in turn, serve as diodes in one case, memory elements in another and even electroluminescent devices. A schematic of such a nanowire device is provided below. Ultimately, though, the trick is to learn how to assemble such nanowires into useful structures in a convenient and reproducible fashion.

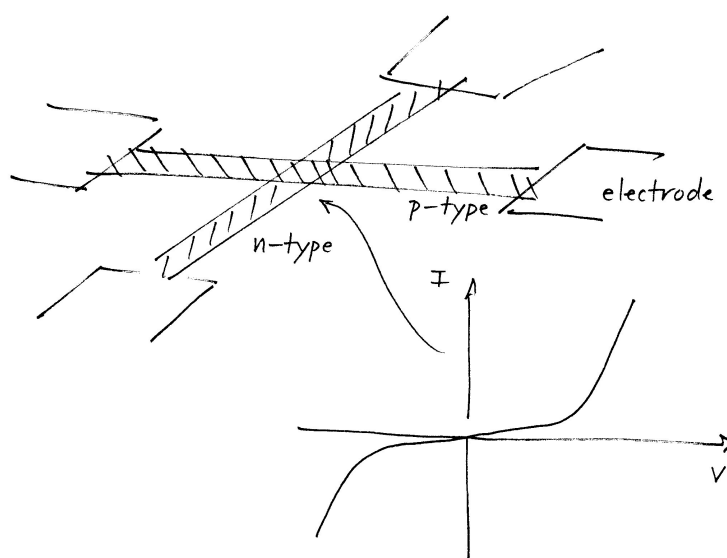


Figure 21.1: Cartoon of a crossed nanowire junction that has been used for proof of principle applications such as memory storage and electroluminescence.

Nanowires have also been used as sensors by monitoring changes in the conductance experienced when different compounds or gases are adsorbed to the wire's surface. In this respect, nanowires may one day be packaged as efficient sensors for minute amounts of toxic gases, chemical weapons, and explosives.

Quantum dots

In the realm of colloidal quantum dots the following applications have been proposed:

- Quantum dots for biological labeling

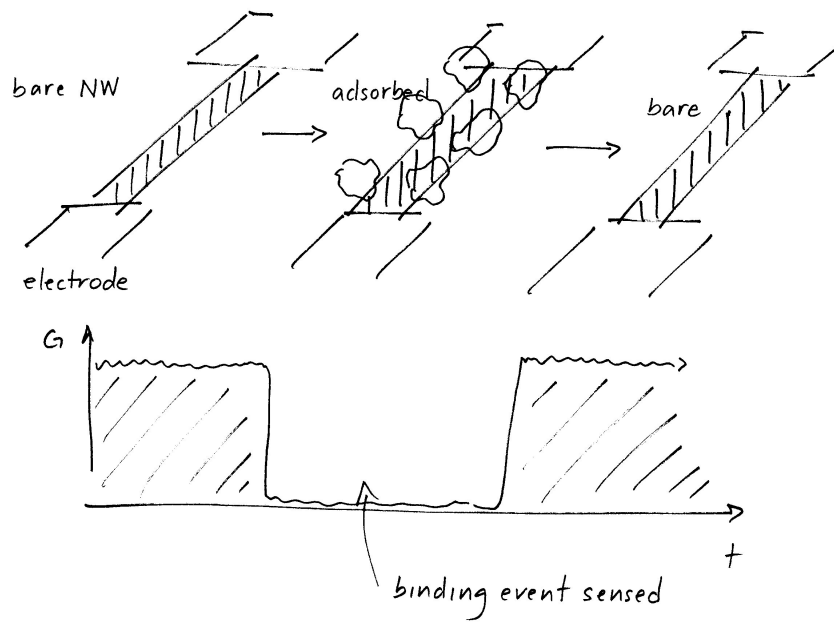


Figure 21.2: Cartoon of nanowire sensor based on changes in conductance.

- Quantum dots as lasing elements
- Quantum dots as sensitizers in photovoltaic applications
- Quantum dots for active layers in light emitting diodes
- Quantum dots as memory elements; single electron transistors

Brief descriptions of each application and reasons why quantum dots have distinct advantages over conventional solutions are presented below.

Medicine; Biological labeling

Conventional biological labeling is currently carried out using organic fluorescent molecules or in some cases radioactive sources. In the case of organic fluorophores such as tetramethylrhodamine (TMR), these molecules are covalently attached to a biological specimen of interest through specific linking chemistry. Organic fluorophores exhibit several disadvantages. Namely, organic dyes suffer from an effect called photobleaching where after exposure to incident light for a modest amount of time, they undergo some sort of photochemistry which ultimately renders them non-fluorescent. Basically the dyes “fade”. This makes labeling and tracking experiments difficult because of the finite observation window one has before the fluorescent signal disappears. As a general rule of thumb, organic dyes will absorb and/or emit approximately 10^6 photons before photobleaching. In addition, organic dyes typically have fairly discrete absorption spectra. So from dye to dye their absorption wavelength or energy will change dramatically. This makes multicolor experiments difficult because exciting each dye requires a different excitation color. Proper filtering of the desired emission signal becomes increasingly difficult in this environment of multiple excitation frequencies. Finally, achieving different colors for these multicolor experiments may mean synthesizing different compounds, which, in itself, can be fairly involved.

Quantum dots, especially CdSe have narrow emission spectra (~ 30 nm FWHM). Furthermore, because of quantum confinement effects, different sized dots emit different colors (one material, many discrete colors). This eliminates the need for synthesizing many different organic fluorophores. As one progresses to higher energies in the dot absorption spectra, there are increasingly larger numbers of excited states present. This is analogous to solutions of the particle in a 3D box with progressively larger quantum numbers, n . So all dots whether they be “small” or “large” will absorb excitation wavelengths in the “blue”. This makes multicolor experiments

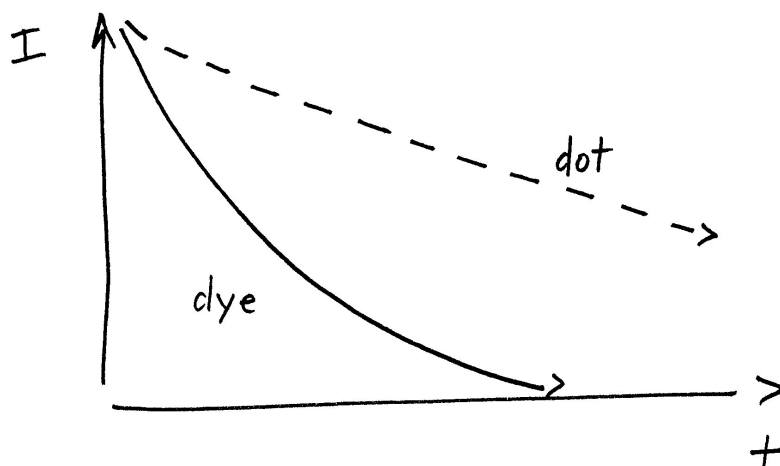


Figure 21.3: Comparison of a quantum dot to organic dye photobleaching rate.

easier since it eliminates the need for multiple excitation wavelengths. One laser, say the 488 nm line from an argon ion, can be used to excite all dots, giving emission anywhere in the visible. Filtering the 488 nm line is also much simpler than trying to simultaneously filter the 473 nm, 488 nm, 514 nm, 532 nm, and 543 nm lines of several lasers (argon ion lines plus YAG doubled line plus green HeNe line). Finally, semiconductor quantum dots are inorganic compounds. As such they are somewhat more robust than organic dyes when it comes to photobleaching. Dots have been seen to absorb and emit over 10^8 photons before experiencing irreversible photobleaching (two orders of magnitude more photons). Therefore, dots are much more resistant to fading. The accompanying figure is a depiction of this.

Ok, so what's the catch? Well, the surface chemistry of quantum dots is still in its infancy. There is still much that needs to be understood before we can begin to do specific chemistry, attaching dots to specific sites on proteins or cells or other biological specimens. This is an area where organic dyes still prevail. Furthermore, semiconductor quantum dots, although nanometer sized, may also be a little too big for some labeling experiments. There might be certain membranes or cellular regions that a dot cannot penetrate because of natural size restrictions (another area where organic dyes are better). Finally, labeling proteins or other specimens with relatively large

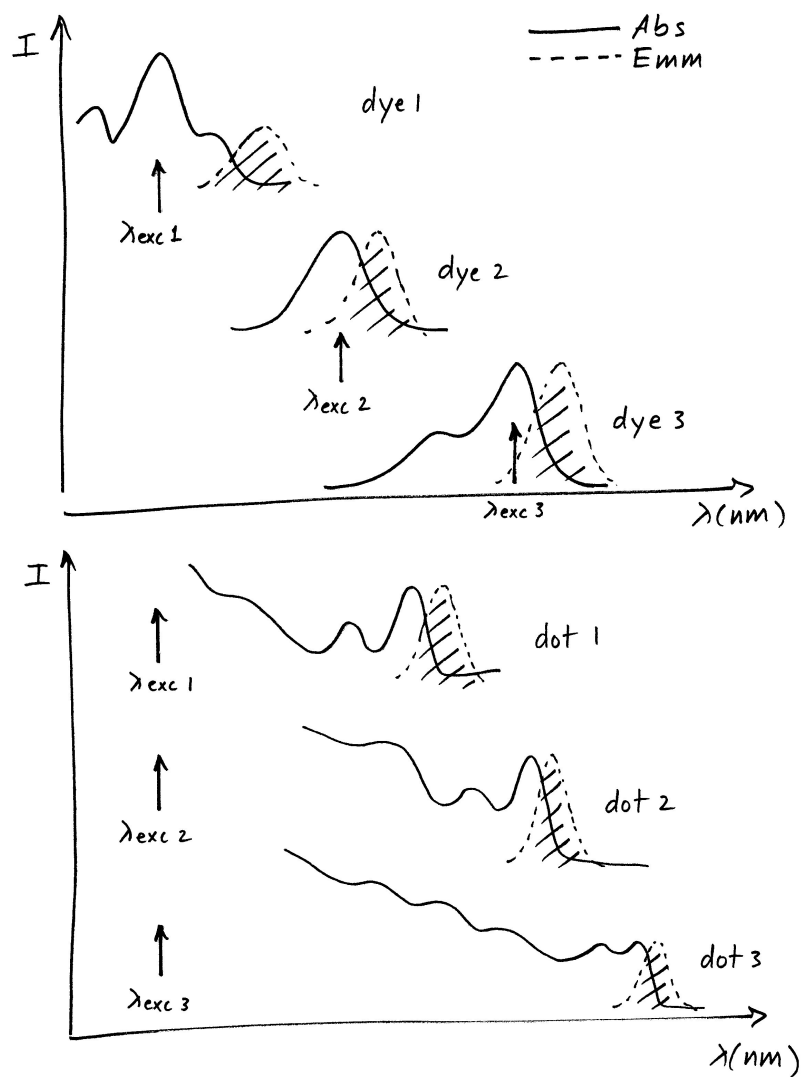


Figure 21.4: Comparison of quantum dot absorption/emission spectra to organic dye absorption emission spectra in light of multicolor labeling experiments.

quantum dots may also perturb the system in unintended ways. So for example if one is trying to watch protein folding in real time one needs to ask whether the dots attached to the protein are actually affecting the folding and unfolding pathways. Consider the size of a typical protein and the size of a typical quantum dot.

Lasing

Lasers are important devices used in everything from tomorrow's national missile defense system (Reagan Star Wars Version 2.0), the data reading element in your DVD or CD player, the red bar code scanner at the supermarket to an excitation source in the laboratory. Conventional lasing sources are based on gases, semiconductors and even organic dyes. With the general movement towards solid state lasers, semiconductors have received a lot of interest for diode laser applications. Further interest was generated with the realization of semiconductor nanostructures (also called low dimensional materials) since it was realized that these systems could potentially make even more efficient lasers than their bulk counterparts. This has to do with the density of states argument that we discussed in previous chapters. The density of states argument won't be repeated here but rather is briefly summarized in the accompanying figure. In this area, quantum well lasers have led the technology, producing some of the most efficient and tunable lasing systems to date. Nanowires have recently been made to lase but the technology in its infancy as with lasing in quantum dots. However, one can envision that the size dependent emission spectra of quantum dots, wires or wells make them attractive lasing elements. In the specific case of colloidal quantum dots, the emission from CdSe is shown to span the entire visible part of the spectrum. So, in principle, a single device could carry a CdSe blue laser, a CdSe green laser and a CdSe red laser. One potential drawback with this system though is a phenomenon called Auger ionization, which might ultimately limit the applicability of this material. However, we leave it to the reader to do some outside reading if they are interested in this subject.

Energy; Photovoltaics

Renewable energy has been an area of great interest since the 1973 OPEC oil embargo, in retaliation for our support of Israel in the 1973 Yom Kippur War. The idea for alternative sources of energy is to eventually move away from coal or petroleum based sources of energy. Motivating this are economic,

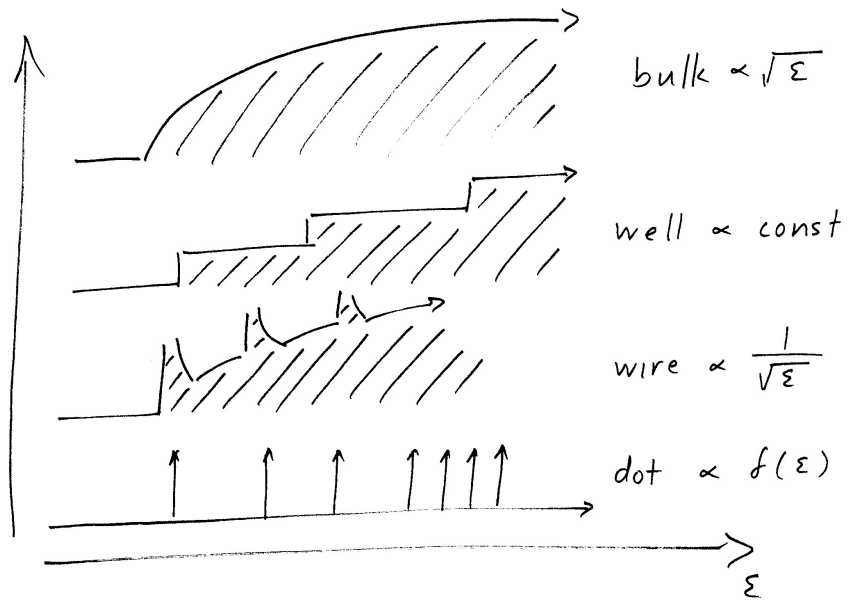


Figure 21.5: Comparison of the density of states for bulk versus a well versus a wire versus a dot.

political and environmental reasons. Solar energy is one facet of renewable energy with wind, methanol, and hydrogen being others. The idea here is to take advantage of the sun's abundant energy and convert it to usable energy much like how Nature has come up with photosynthesis in plants. What's needed, however, is an active material like chlorophyll that can absorb solar radiation and provide efficient charge separation to prevent radiative or nonradiative recombination in the material.

Commercial solar cells are currently made of silicon. Unfortunately, the efficiencies of these devices is typically on the order of 15%. So most of the solar energy collected by these devices is wasted. To make up for all of these losses, large tracts of land must be used for vast sprawling fields of solar cells (solar farms). Improved devices made of single crystal silicon have been shown to achieve conversion efficiencies of 30% but at the cost of being very expensive and impractical for commercial use. As a consequence solar energy has not broken through into mainstream use.

Quantum dots come into play for several reasons. They have tunable, size dependent, absorption and emission spectra. Different quantum dots can be made to absorb anywhere from the UV into the infrared. This tremendous dynamic range cannot be matched by organic dyes. Furthermore, there are few organic dyes that are efficient in the infrared. As a side note, one can imagine a quantum dot based solar cell that operates under cloudy conditions and rainy days where the overcast sky will block much of the visible yet still transmits most (if not all) of the infrared. In addition, the absorption cross section or extinction coefficient of quantum dots is generally an order of magnitude greater than conventional organic dyes. This means it take fewer dots to absorb the same amount of light. Dots are also more photostable, meaning that they are more likely to reach the 10,000 hour threshold needed for practical commercial devices. Furthermore, nanoparticles when used as substrates or electrodes in dye based solar cells have much larger surface areas than conventional bulk substrates. As a consequence, one can adsorb a greater number of dye molecule per unit area in these hybrid devices than in conventional cells. The efficiencies of these hybrid devices is consequently higher, reaching that of conventional silicon cells. One of the first of such devices is referred to as the Gratzel cell after its inventor.

Lighting; Light emitting diodes

Lighting hasn't changed all that much since the light bulb was invented by Edison and others close to a hundred years ago. More efficient fluorescent

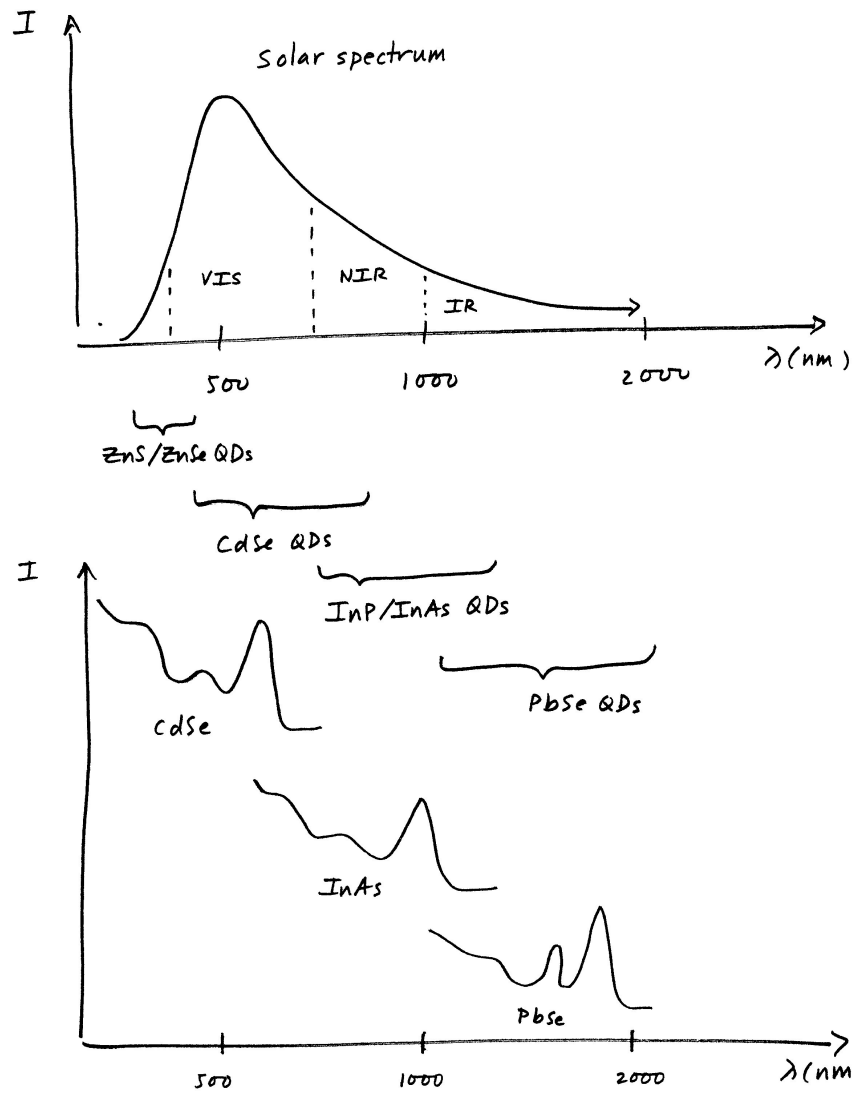


Figure 21.6: Comparison of the solar spectrum and representative quantum dot absorption spectra as well as achievable wavelength ranges.

lighting has since been developed but suffers from flicker and color purity issues. Recently solid state light emitting diodes (LED) have come on the market and are poised to revolutionize the lighting industry. LED devices that exhibit tremendous brightness (look at some of the new red and green traffic lights), consume little power, come in different colors, and emit little or no heat (museum quality lighting for paintings) are now commercially available. In this regard, a major goal of the LED industry is to eventually achieve affordable white light by mixing red, green and blue LEDs. The idea is to one day replace *all* incandescent and fluorescent light bulbs in homes and offices. Furthermore, along these lines, brighter, more efficient, flat panel displays using this technology, rather than inefficient backlit liquid crystal displays, may come out of these developments. Along the same lines, cheaper high definition digital televisions may also emerge from this technology.

A current problem with LEDs, however, is that different active semiconductor elements must be manufactured via potentially expensive processes such as MOCVD to achieve multiple colors. For example, GaN is used for blue light, indium doped GaN can be used to get green and so forth. One way to circumvent this problem is to take advantage of quantum confinement as in the case of quantum dots. Different sized quantum dots will emit different colors so, in principle, one material can cover the entire visible spectrum. They can also be manufactured using the same process potentially lowering overall manufacturing costs. One disadvantage with current colloidal quantum dots is that the heterojunction between the dot and the outside world is imperfect. There are organic ligands present as well as many quantum dot surface defects that open up undesired states and recombination pathways in addition to creating large resistances to carrier transport.

Memory; the Coulomb staircase

What would a chapter on devices and applications be if we didn't touch on computers. Back in 1965, Gordon Moore, one of the founders of Intel made an empirical observation that has since become known as "Moore's law" (or sometimes referred to as Moore's first law). The number of transistors per unit area on an integrated circuit doubles each year. Since then, Moore's law has generally held with some minor modifications. It now doubles every 18 months. However, as you might suspect, this wild growth cannot continue forever and it was realized that with current photolithographic techniques that we would be in trouble by 2010. To consistently get more transistors per unit area means that their size decreases yearly. Currently the features

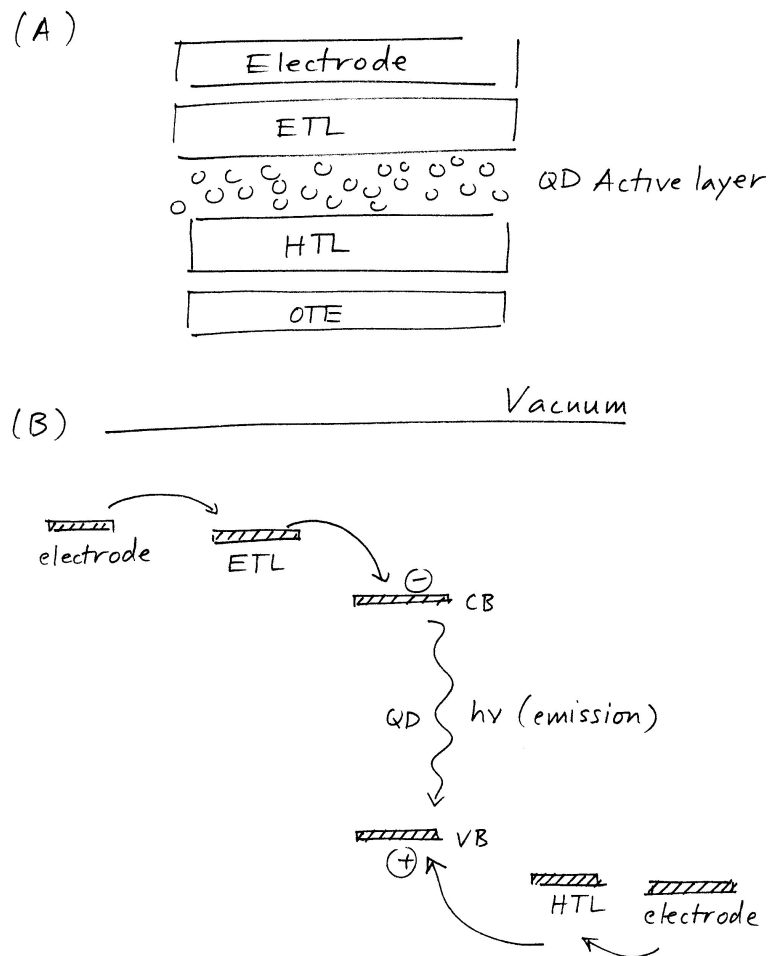


Figure 21.7: (A) Cartoon of a generic quantum dot light emitting diode. HTL (hole transport layer); ETL (electron transport layer); OTE (optically transparent electrode) (B) Ideal energy level diagram for injection of electrons and holes into a quantum dot device.

on a Pentium IV chip have spacings on the order of 0.11 microns (110 nm). Next generation chips will have features spaced by 0.09 microns (90 nm). How much lower can we go? Well, because of the diffraction limit we cannot continue to use existing techniques but are forced to invest in deep UV photolithography or x-ray lithography or even e-beam lithography if we are to get smaller transistors and stay on track with Moore's law. Such new technologies are very expensive and potentially too costly to scale up to the fab level (Moore's second law of costs). Because of this, researchers have looked to nano for a solution. Among the ideas people have come up with are what are referred to as single electron transistors.

Early on, researchers realized that if one has a very small metal nanoparticle, its capacitance might be large enough to store discrete charges. Lowering the temperature also helps. Both work because either raising the capacitance or lowering the temperature decreases the value of the thermal energy relative to the Coulomb energy between discrete charges. In turn, this allows one to store charges on the metal nanoparticle without having it thermally expelled. Alternatively, with semiconductor quantum dots, the discrete particle in a box-like energy levels with spacings large compared to kT also means discrete steps in the conductance of electrons through the dot and the additional possibility of storing charges just as with the metal nanoparticles. These effects could then form the basis of single electron electronics of which the single electron transistor is a member. We review the principles of what is known as the Coulomb blockade and Coulomb staircase model below because of its potential importance.

In the orthodox model for single electron tunneling, a simple circuit model is considered as shown in the accompanying figure. Basically the circuit consists of a perfect voltage source and two capacitors that may or may not have equivalent capacitances. In the orthodox model, one of the two capacitors is generally considered to have a much higher capacitance than the other. The region in between the capacitors is the "island" where electrons can be stored. This region represents a quantum dot or metal nanoparticle in real life.

The total electrostatic energy of the system is

$$\boxed{E_s = \frac{q_1^2}{2C_1} + \frac{q_2^2}{2C_2}} \quad (21.1)$$

where C_1 is the gate capacitance. At the same time the potential drops

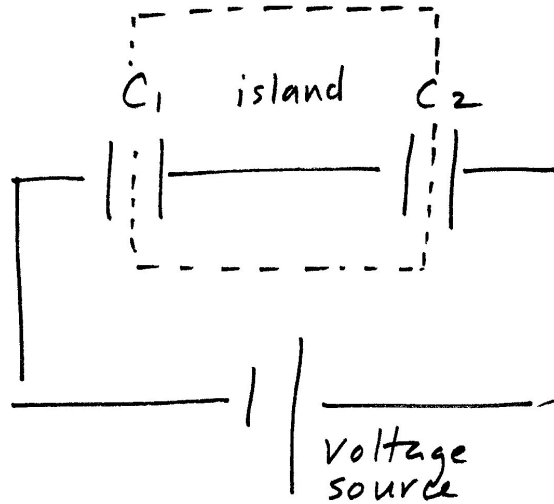


Figure 21.8: Orthodox model of single electron tunneling and Coulomb blockade/Coulomb staircase

across junctions 1 and 2 are

$$\begin{aligned}
 V_1 &= \left(\frac{C_2}{C_1 + C_2} \right) V + \frac{ne}{C_1 + C_2} \\
 &= \frac{C_2 V + ne}{C_1 + C_2} \qquad (21.2)
 \end{aligned}$$

$$\begin{aligned}
 V_2 &= \left(\frac{C_1}{C_1 + C_2} \right) V - \frac{ne}{C_1 + C_2} \\
 &= \frac{C_1 V - ne}{C_1 + C_2} \qquad (21.3)
 \end{aligned}$$

Therefore the total electrostatic energy of the system is

$$\begin{aligned}
 E_s &= \frac{q_1^2}{2C_1} + \frac{q_2^2}{2C_2} \\
 &= \frac{C_1^2 V_1^2}{2C_1} + \frac{C_2^2 V_2^2}{2C_2} \\
 &= \frac{C_1 V_1^2}{2} + \frac{C_2 V_2^2}{2} \\
 &= \frac{1}{2} (C_1 V_1^2 + C_2 V_2^2) \\
 &= \frac{1}{2} \left(C_1 \left(\frac{C_2 V + ne}{C_1 + C_2} \right)^2 + C_2 \left(\frac{C_1 V - ne}{C_1 + C_2} \right)^2 \right) \\
 &= \frac{1}{2} \left(\frac{1}{C_1 + C_2} \right)^2 (C_1 (C_2 V + ne)^2 + C_2 (C_1 V - ne)^2)
 \end{aligned}$$

The term in the last parenthesis can be expanded and reduced to give

$$\begin{aligned}
 E_s &= \frac{1}{2} \left(\frac{1}{C_1 + C_2} \right)^2 (C_1 C_2 V^2 (C_1 + C_2) + (ne)^2 (C_1 + C_2)) \\
 &= \frac{1}{2C_{tot}} (C_1 C_2 V^2 + (ne)^2)
 \end{aligned} \tag{21.4}$$

where $C_{tot} = C_1 + C_2$.

Now the net energy of the system, (or free energy) is the difference in energy between the total electrostatic energy stored and the work needed to shove an electron onto the island.

$$E_{tot} = E_s - W \tag{21.5}$$

where $W = \frac{C_1 Q V}{C_{tot}}$ is the work done by the system to load the island. This

results in

$$\begin{aligned}
 E_{tot} &= \frac{1}{2C_{tot}}(C_1C_2V^2 + (ne)^2) - \frac{C_1QV}{C_{tot}} \\
 &= \frac{(ne)^2}{2C_{tot}} - \frac{2C_1QV}{2C_{tot}} + \frac{C_1C_2V^2}{2C_{tot}} \\
 &= \frac{(ne)^2 - 2C_1QV}{2C_{tot}} + \frac{(C_1V)^2}{2C_{tot}} - \frac{(C_1V)^2}{2C_{tot}} + \frac{C_1C_2V^2}{2C_{tot}}
 \end{aligned}$$

Notice the trick consisting of adding and subtracting the middle terms

$$\begin{aligned}
 &= \frac{(ne)^2 - 2C_1QV + (C_1V)^2}{2C_{tot}} - \frac{(C_1V)^2}{2C_{tot}} + \frac{C_1C_2V^2}{2C_{tot}} \\
 &= \frac{(ne - C_1V)^2}{2C_{tot}} - \frac{1}{2C_{tot}}(C_1^2V^2 - C_1C_2V^2)
 \end{aligned}$$

where $C_2 = C_{tot} - C_1$

$$\begin{aligned}
 &= \frac{(ne - C_1V)^2}{2C_{tot}} - \frac{1}{2C_{tot}}(C_1^2V^2 - C_1(C_{tot} - C_1)V^2) \\
 &= \frac{(ne - C_1V)^2}{2C_{tot}} - \frac{1}{2C_{tot}}(C_1^2V^2 - C_1C_{tot}V^2 + C_1^2V^2) \\
 &= \frac{(ne - C_1V)^2}{2C_{tot}} - \frac{1}{2C_{tot}}(2C_1^2V^2 - C_1C_{tot}V^2) \\
 &= \frac{(ne - C_1V)^2}{2C_{tot}} - \frac{C_1V^2}{2C_{tot}}(2C_1 - C_{tot}) \\
 &= \frac{(ne - C_1V)^2}{2C_{tot}} - \frac{C_1V^2}{2C_{tot}}(C_{tot}) \\
 &= \frac{(ne - C_1V)^2}{2C_{tot}} - \frac{C_1V^2}{2}
 \end{aligned}$$

This gives the total free energy of the system

$$\boxed{E_{tot} = \frac{(ne - C_1V)^2}{2C_{tot}} - \frac{C_1V^2}{2}} \quad (21.6)$$

Island occupation number

The average number of electrons on the island is given by

$$\langle n \rangle = \frac{\sum_{-\infty}^{\infty} n e^{-\frac{E_n}{kT}}}{\sum_{-\infty}^{\infty} e^{-\frac{E_n}{kT}}}$$

where from before

$$E_n = \frac{(-ne + \bar{Q})^2}{2C_{tot}} - \frac{\bar{Q}^2}{2C_1}$$

Therefore

$$\begin{aligned} \langle n \rangle &= \frac{\sum_{-\infty}^{\infty} n e^{-\frac{(-ne+\bar{Q})^2}{2C_{tot}kT} + \frac{\bar{Q}^2}{2C_1kT}}}{\sum_{-\infty}^{\infty} e^{-\frac{(-ne+\bar{Q})^2}{2C_{tot}kT} + \frac{\bar{Q}^2}{2C_1kT}}} \\ &= \frac{e^{\frac{\bar{Q}^2}{2C_1kT}} \sum_{-\infty}^{\infty} n e^{-\frac{(-ne+\bar{Q})^2}{2C_{tot}kT}}}{e^{\frac{\bar{Q}^2}{2C_1kT}} \sum_{-\infty}^{\infty} e^{-\frac{(-ne+\bar{Q})^2}{2C_{tot}kT}}} \\ &= \frac{\sum_{-\infty}^{\infty} n e^{-\frac{(-ne+\bar{Q})^2}{2C_{tot}kT}}}{\sum_{-\infty}^{\infty} e^{-\frac{(-ne+\bar{Q})^2}{2C_{tot}kT}}} \\ &= \frac{\sum_{-\infty}^{\infty} n e^{-e^2 \frac{\left\{-n + \frac{\bar{Q}}{e}\right\}^2}{2C_{tot}kT}}}{\sum_{-\infty}^{\infty} e^{-e^2 \frac{\left\{-n + \frac{\bar{Q}}{e}\right\}^2}{2C_{tot}kT}}} \\ &= \frac{\sum_{-\infty}^{\infty} n e^{-\frac{\left\{-n + \frac{\bar{Q}}{e}\right\}^2}{2\Theta}}}{\sum_{-\infty}^{\infty} e^{-\frac{\left\{-n + \frac{\bar{Q}}{e}\right\}^2}{2\Theta}}} \end{aligned}$$

where $\Theta = \frac{C_{tot}kT}{e^2}$

Furthermore, let $x = \left(\frac{\bar{Q}}{e}\right)$ to get our final expression for the average island occupation

$$\boxed{\langle n \rangle = \frac{\sum_{-\infty}^{\infty} n e^{-\frac{(x-n)^2}{2\Theta}}}{\sum_{-\infty}^{\infty} e^{-\frac{(x-n)^2}{2\Theta}}}} \quad (21.7)$$

A plot of this below, shows the characteristic Coulomb staircase behavior.

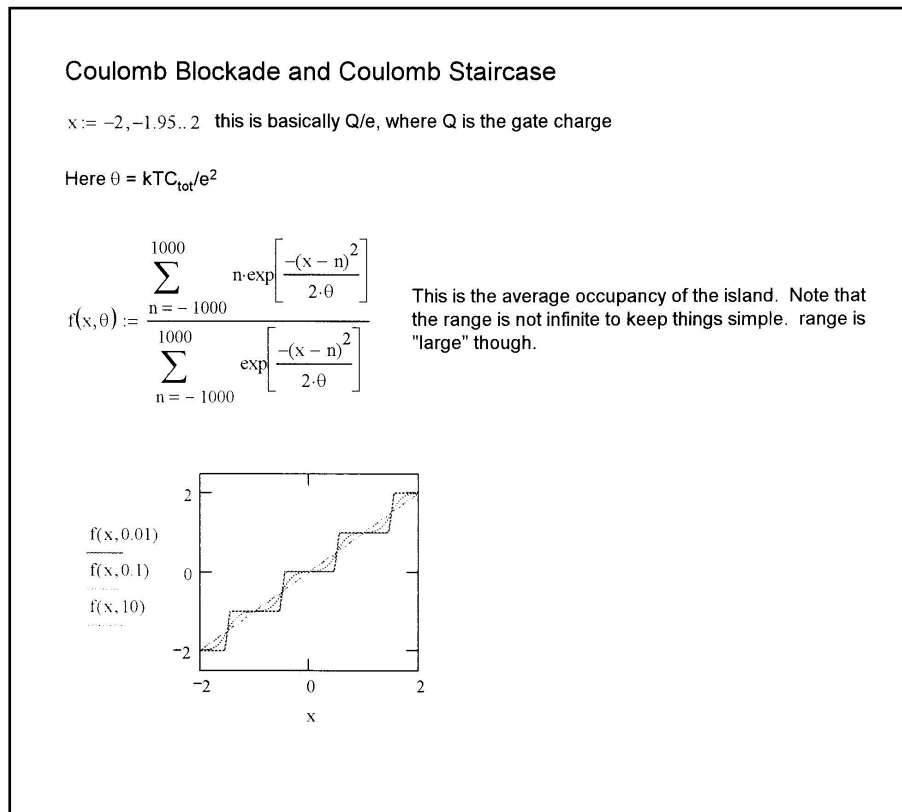


Figure 21.9: Coulomb staircase: Mathcad numerical solutions

Relevant literature

These are papers from the current literature in no particular order

- “Single-nanowire electrically driven lasers”
X. Duan, Y. Huang, R. Agarwal, C. M. Lieber
Nature, 421, 241 (2003).
- “Functional nanoscale electronic devices assembled using silicon nanowire building blocks”
Y. Cui, C. M. Lieber
Science, 291, 851 (2001).

- “Quantum dot bioconjugates for ultrasensitive nonisotopic detection”
W. C. Chan, S. Nie
Science, 281, 2016 (1998).
- “Semiconductor nanocrystals as fluorescent biological labels”
M. Bruchez Jr, M. Moronne, P. Gin, S. Weiss, A. P. Alivisatos
Science, 281, 2013 (1998).
- “Indium phosphide nanowires as building blocks for nanoscale electronic and optoelectronic devices”
X. Duan, Y. Huang, Y. Cui, J. Wang, C. M. Lieber
Nature, 409, 66 (2001).
- “Immunofluorescent labeling of cancer marker Her2 and other cellular targets with semiconductor quantum dots”
X. Wu et al.
Nature Biotechnology, 21, 41 (2003).
- “Nanowire nanosensors for highly sensitive and selective detection of biological and chemical species”
Y. Cui, Q. Wei, H. Park, C. M. Lieber
Science, 293, 1289 (2001).
- “Room-temperature ultraviolet nanowire nanolasers”
M. H. Huang et al.
Science, 292, 1897 (2001).
- “Hybrid nanorod-polymer solar cells”
W. U. Huynh, J. J. Dittmer, A. P. Alivisatos
Science, 295, 2425 (2002).
- “Quantum dot solar cells”
A. J. Nozik
Physica E. 12, 115 (2002).
- “Electroluminescence from single monolayers of nanocrystals in molecular organic devices”
S. Coe, W-K Woo, M. Bawendi, V. Bulovic
Nature, 420, 800 (2002).
- “A silicon nanowire with a coulomb blockade effect at room temperature”
S. Hu, W-Z. Wong, S-S. Liu, Y-C. Wu, C-L. Sung, T-Y. Huang, T-J.

Yang

Adv. Mater. 14, 736 (2002).

- “Quantum dot solar cells”
R. P. Raffaele, S. L. Castro, A. F. Hepp, S. G. Bailey
Progress in photovoltaics: research and applications, 10, 433 (2002).
- “A multiwavelength semiconductor laser”
A. Tredicucci, C. Gmachl, F. Capasso, D. L. Sivco, A. L. Hutchinson,
A. Y. Cho
Nature, 396, 350 (1998).
- “Bidirectional semiconductor laser”
C. Gmachl, A. Tredicucci, D. L. Sivco, A. L. Hutchinson, F. Capasso,
A. Y. Cho
Science, 286, 749 (1999).
- “Array-based electrical detection of DNA with nanoparticle probes”
S-J. Park, T. A. Taton, C. A. Mirkin
Science, 295, 1503 (2002)
- “Scanometric DNA array detection with nanoparticle probes”
T. A. Taton, C. A. Mirkin, R. L. Letsinger
Science, 289, 1757 (2000).
- “Efficient near-infrared polymer nanocrystal light-emitting diodes”
N. Tessler, V. Medvedev, M. Kazes, S. Kan, U. Banin
Science, 295, 1506 (2002).
- “Nanotube molecular wires as chemical sensors”
J. Kong, N. R. Frankliin, C. Zhou, M. G. Chapline, S. Peng, K. Cho,
H. Dai
Science, 287, 622 (2000).
- “Carbon nanotube actuators”
R. H. Baughman et. al.
Science, 284, 1340 (1999).
- “Logic circuits with carbon nanotube transistors”
A. Backtold, P. Hadley, T. Nakanishi, C. Dekker
Science, 294, 1317 (2001).

- “Single-and multi-wall carbon nanotube field-effect transistors”
R. Martel, T. Schmidt, H. R. Shea, T. Hertel, Ph. Avouris
Appl. Phys. Lett. 73, 2447 (1998).
- “Carbon nanotube electron emitter with a gated structure using back-side exposure processes”
D-S. Chung et. al.
Appl. Phys. Lett. 80, 4045 (2002).
- “Nano-particle transistors and energy-level spectroscopy in metals”
D. C. Ralph, C. T. Black, M. Tinkham
Superlattices and Microstructures, 20, 389 (1996).
- “A single-electron transistor made from a cadmium selenide nanocrystal”
D. L. Klein, R. Roth, A. L. L. Lim, A. P. Alivisatos, P. L. McEuen
Nature, 389, 699 (1997).
- “Gold nanoparticle single-electron transistor with carbon nanotube leads”
C. Thelander, M. H. Magnusson, K. Deppert, L. Samuelson, P.R. Poulsen, J. Nygard, J. Borggreen
Appl. Phys. Lett. 79, 2106 (2001).
- “Single-electron transistor made of two crossing multiwalled carbon nanotubes and its noise properties”
M. Ahlskog, R. Tarkiainen, L. Roschier, P. Hakonen
Appl. Phys. Lett. 77, 4037 (2000).
- “Single electron transistor using a molecularly linked gold colloidal particle chain”
T. Sato, H. Ahmed, D. Brown, B. F. G. Johnson
J. Appl. Phys. 82, 696 (1997).
- “Room temperature coulomb blockade and coulomb staircase from self assembled nanostructures”
R. P. Anders et al.
J. Vac. Sci. Technol. A 14, 1178 (1996).
- “Single electron tunneling through nano-sized cobalt particles”
C. Petit, T. Cren, D. Roditchev, W. Sacks, J. Klein, M-P. Pileni
Adv. Mater. 11, 1198 (1999).

- “Size-dependent tunneling and optical spectroscopy of CdSe quantum rods”
D. Katz, T. Wizansky, O. Millo, E. Rothenberg, T. Mokari, U. Banin
Phys. Rev. Lett. 89, 086801-1 (2002).
- “Scanning tunneling spectroscopy of InAs nanocrystal quantum dots”
O. Millo, D. Katz, Y. Cao, U. Banin
Phys. Rev. B 61, 16773 (2000).
- “A tunneling spectroscopy study on the single-particle energy levels and electron-electron interactions in CdSe quantum dots”
E. P. A. M. Bakkers, Z. Hens, L. P. Kouwenhoven, L. Gurevich, D. Vanmaekelbergh
Nanotechnology, 13, 258 (2002).

Acknowledgments

I'd like to take the opportunity to thank the Notre Dame Radiation Laboratory and the Department of Chemistry and Biochemistry for financial support over the spring and summer of 2003. I'd also like to thank Jean-Christophe Ducom in the chemistry department for figuring out how to insert a figure in a TEX coverage as well as Matt Meineke for alerting me to the use (and power) of the TEX label command. Finally, I'd like to thank Fred Mikulec, currently at Innovalight, for putting up with all my e-mail questions and always being open to my crazy ideas/garage projects. This text was written and typeset using the TEX environment withing PCTex Ver 4.2.05.

Index

- CdMe₂*, 212
- NaBH₄*, 212
- Ω value, 212
- ϵ cobalt, 27

- absorption coefficient, 85, 95
- AFM, 229
- Airy equation, 184, 189, 193
- Airy function, 184, 189, 193
- AOT, 211
- argon ion, 239
- atomic force microscopy, 229

- band gap, 115
- basis atoms, 11
- BCC, 13
- Bessel equation, 46, 54
- Bessel function, 52
- Bill Clinton, 1
- binomial expansion, 76, 82
- bis(2-ethylhexyl)phosphate, 211
- Bloch's theorem, 117
- Bravais lattice, 11

- CCP, 13
- centripetal force, 30
- colloidal growth, 211
- conventional unit cell, 12
- Coulomb blockade, 245
- Coulomb staircase, 245
- crystal lattice, 11
- CsCl, 17

- deBroglie wavelength, 29
- defocusing regime, 221
- density of states, 57, 58, 62, 63
- density of states, conduction band, 67, 75, 80
- density of states, valence band, 69, 75, 80
- diamond structure, 13
- dimethylcadmium, 212
- dip pen nanolithography, 231

- Einstein A and B coefficients, 97, 103, 107
- emission, 108
- Eric Drexler, 2
- exciton, 31
- exciton Bohr radius, 29

- FCC, 13
- Fermi Dirac distribution, 66, 73–75, 78, 80, 106
- Fermi Dirac integral, 66
- Fermi integral, 66
- Fermi level, 69, 72, 77, 83
- Fermi's golden rule, 95
- Fick's first law, 212
- field emission, 200
- flux, 167, 169, 212
- focusing regime, 220
- Foresight Institute, 2
- Fowler Nordheim, 200

- Gamma function, 67, 69, 82