ივანე ჯავახიშვილის სახელობის თბილისის სახელმწიფო უნივერსიტეტი

ზუსტ და საბუნებების,ეტყველო მეცნიერებათა ფაკულტეტი

მიკრო და ნანო სტრუქტურები

შესავალი

შოთა რუსთაველის ეროვნული სამეცნიერო ფონდი

SHOTA RUSTAVELI NATIONAL SCIENCE FOUNDATION

გეგმა

- რა არის მიკრო და ნანოსტრუქტურები
- რატომ არის საინტერესო
- რატომ დავაწყვილეთ

MICRO
10^{-6}მ

NANO
10^{-9}მ

Au
ლითონები

ელთვის ატარებს დენს
ელთვის არის თავისუფალი
ქტრონების დიდი რაოდენობა

Si
ნახევარგამტარები

ალმასი
იზოლატორები

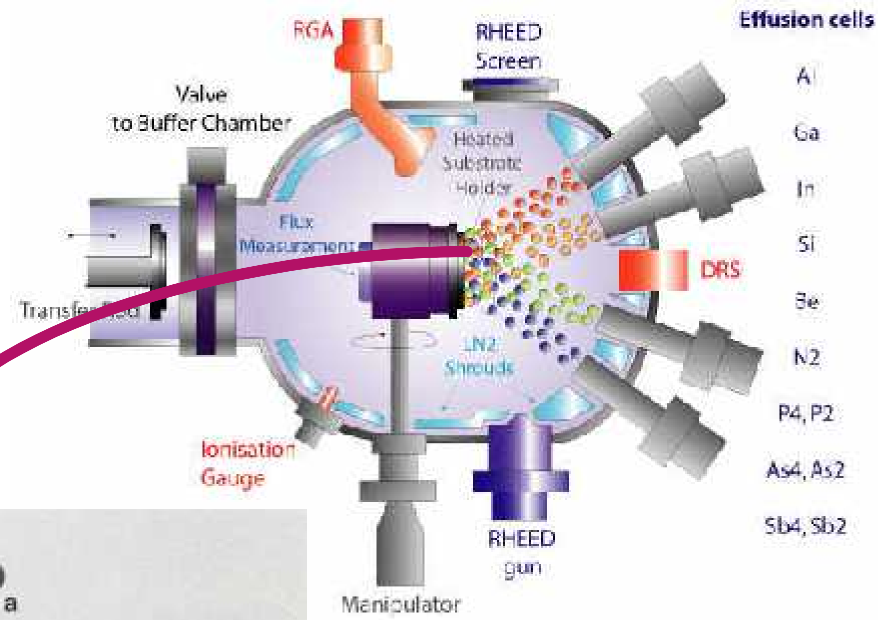
დენს ატარებენ მხოლოდ გარკვეულ პირობებში

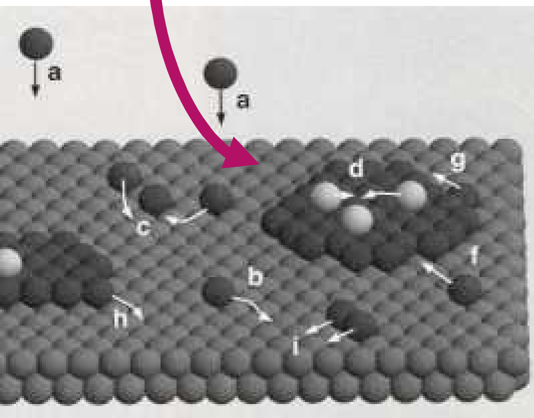
ნახევარგამტარები

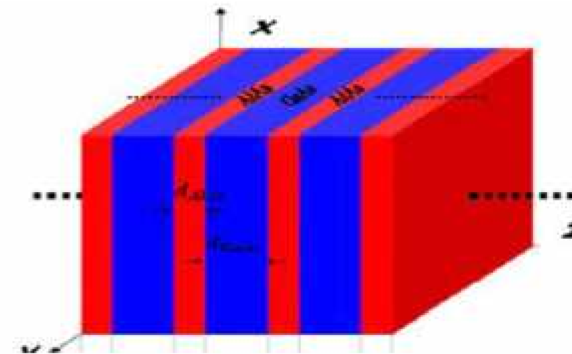
- მუხტის თავისუფალი მატარებლები ჩნდებიან აღზნებისას (გათბობა, სინათლის დასხივება, დეფორმაცია)
- ფიზიკური თვისებები ძალიან მგრძნობიარეა ქიმიური შემადგენლობის მიმართ

	Germanium	Silicon	Gallium Arsenide
300 K	2.02×10^{13}	8.72×10^{9}	2.03×10^{6}
400 K	1.38×10^{15}	4.52×10^{12}	5.98×10^{9}
500 K	1.91×10^{16}	2.16×10^{14}	7.98×10^{11}
600 K	1.18×10^{17}	3.07×10^{15}	2.22×10^{13}

სტრუქტურების ზრდა

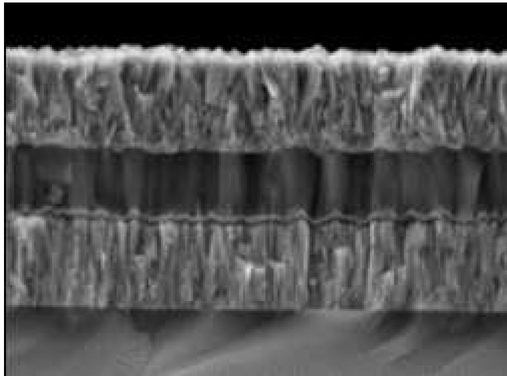


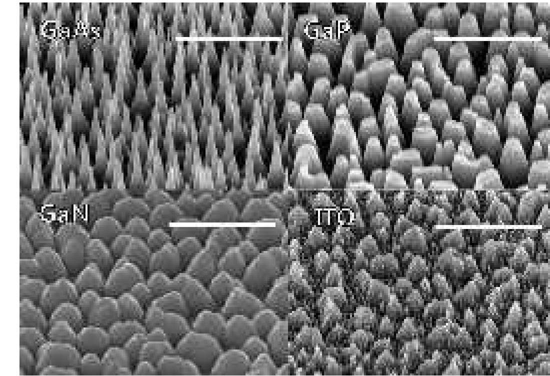




MICRO
10^{-6}მ

NANO
10^{-9}მ

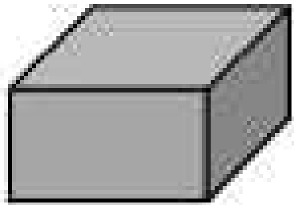


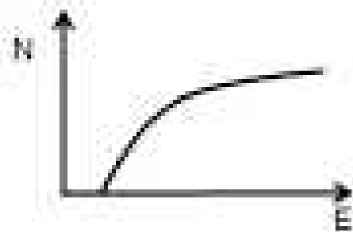


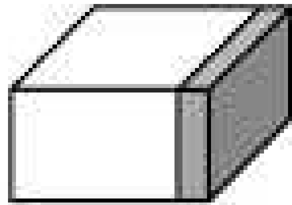
მიკრო ან ნანო მეტრში იგულისხმება აქტიურის არის ზომა, ანუ იმ არის ზომა სადაც მიმდინარე

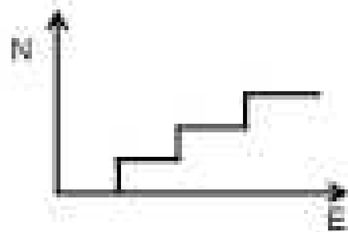
- დენის გატარება,
- ოპტიკური მოვლებები (შთანთქმა, გამოსხივება...)
- მაგნიტური მოვლენები (მაგნიტური მომენტების მოწესრიგება)

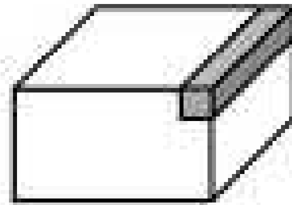
ნანოსტრუქტურები

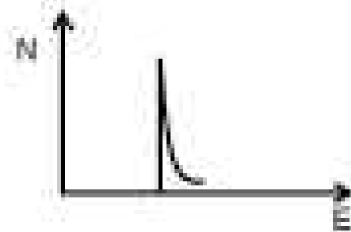


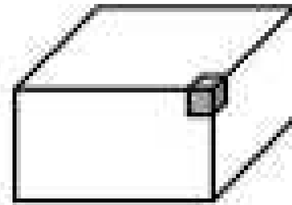


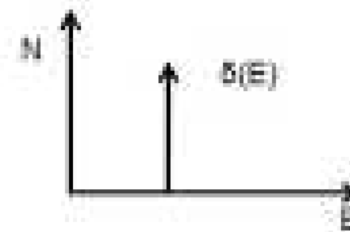








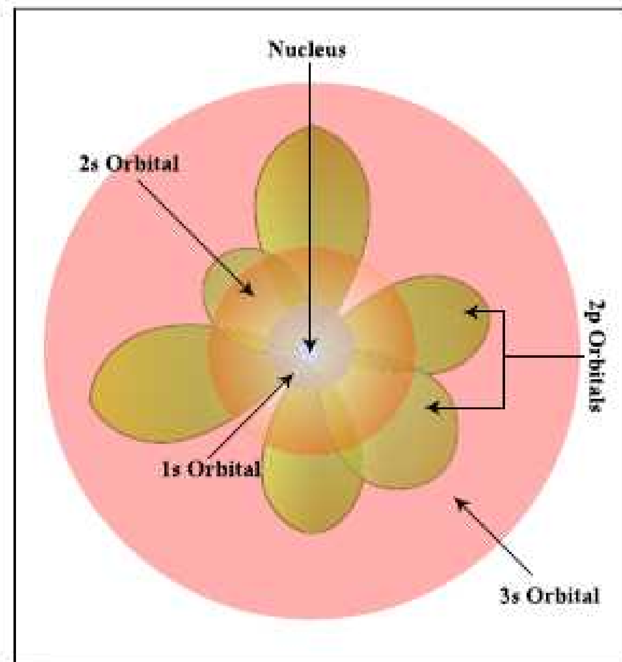




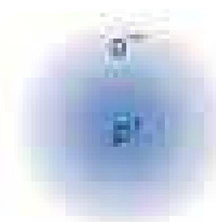
ატომებიდან კრისტალებამდე

1. ატომი - კვანტური ობიექტი
2. მოლეკულა - ქიმიური ბმა
3. კრისტალი - სტრუქტურა სიმეტრია

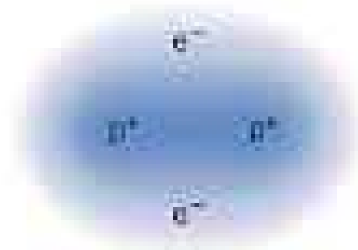
ატომი



ქიმიური ბმა



A No interaction

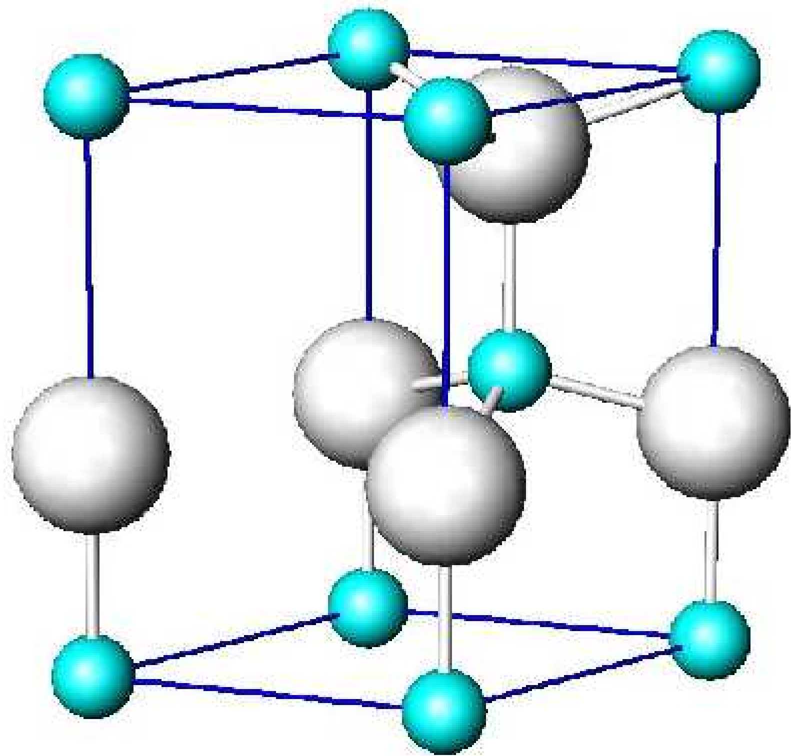


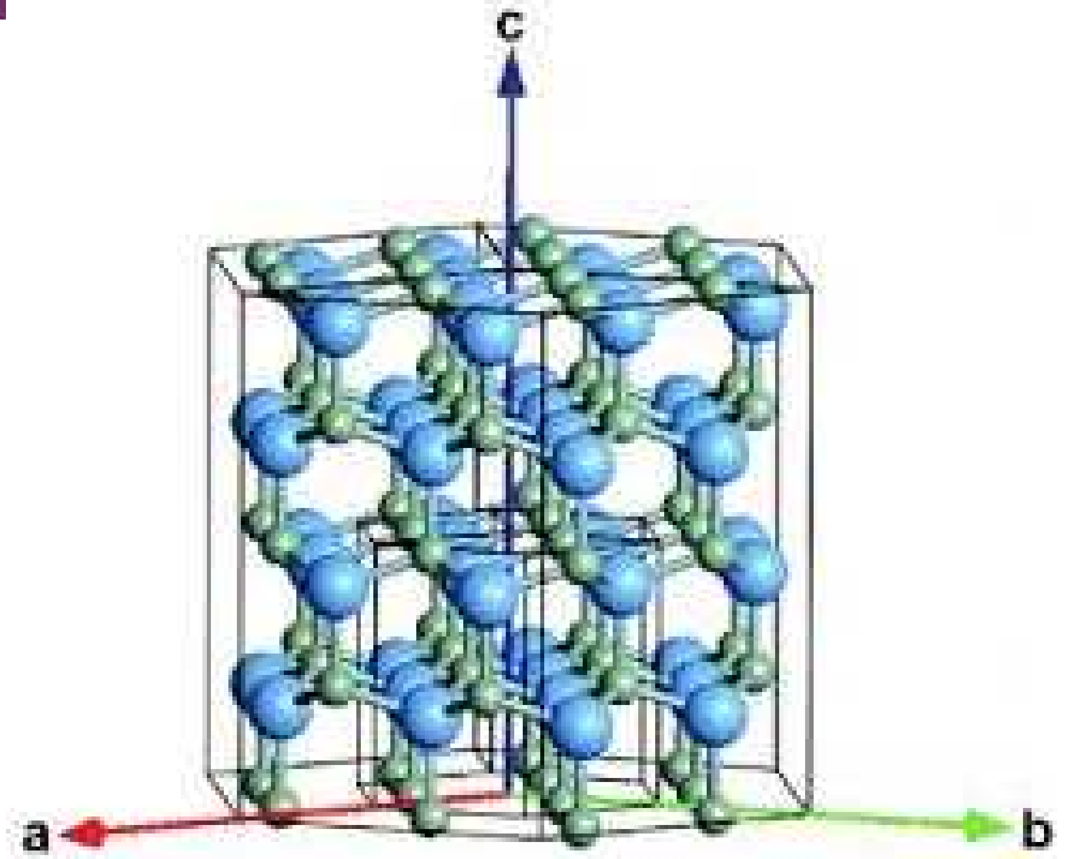
C Covalent bond

B Attraction begins

D Combination of forces

კრისტალური სტრუქტურა



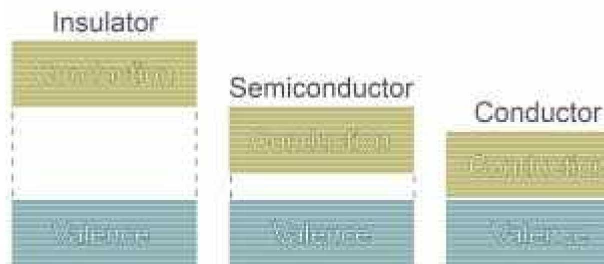


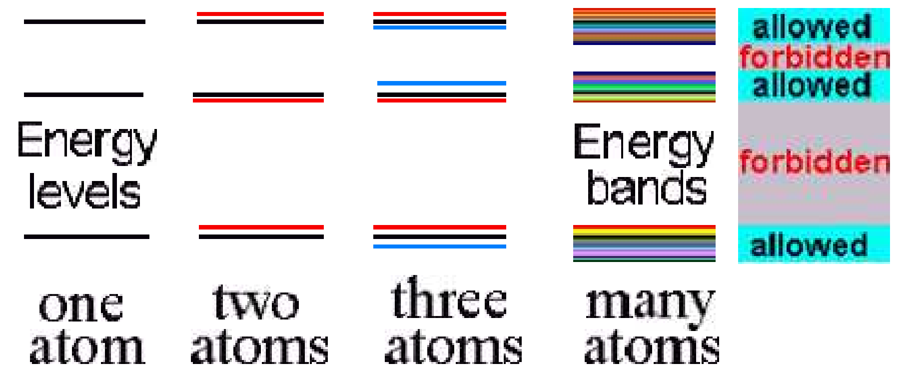
მყარი სხეულების ენერგეტიკული სტრუქტურა(ენერგეტიკული სპექტრი)

კრისტალების თვისებები განისაზღვერაბა ენერგეტიკული სტრუქტურით

ენერგეტიკული სტრუქტურა არის იმ ენერგიების ერთობლიობა, რომელიც შესაძლებელია მიიღოს ელექტრონმა

რა არის ენერგეტიკული სტრუქტურა

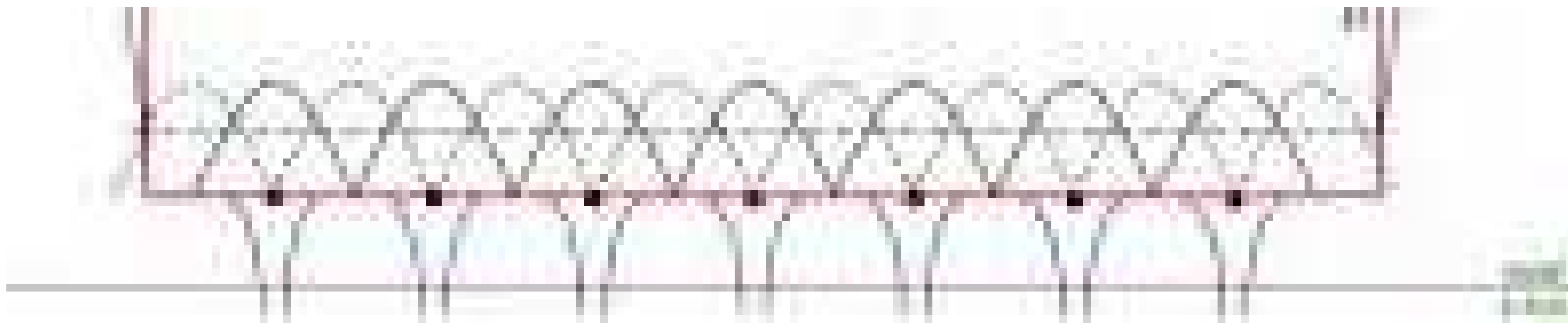




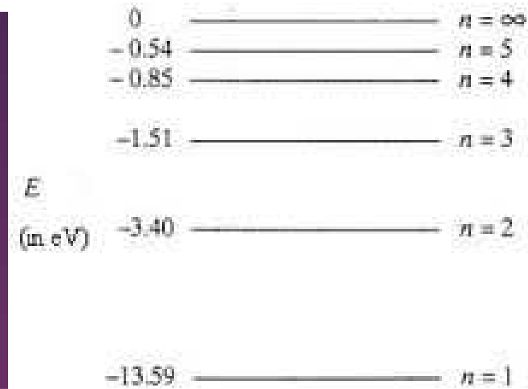
ელექტრონები კრისტალში

- თავისუფალი ელექტრონების მოდელი
- თითქმის თავისუფალი ელექტრონების მოდელი
- ძლიერი ბმის მოდელი

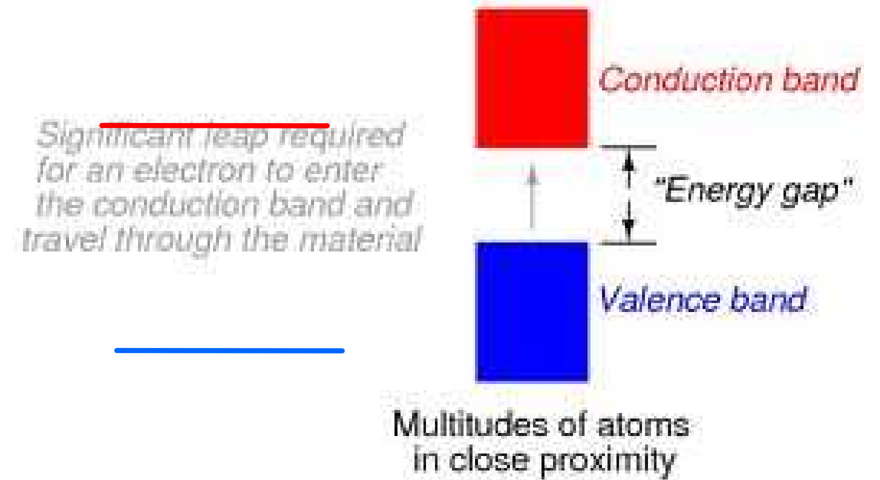
ელექტრონები კრისტალში



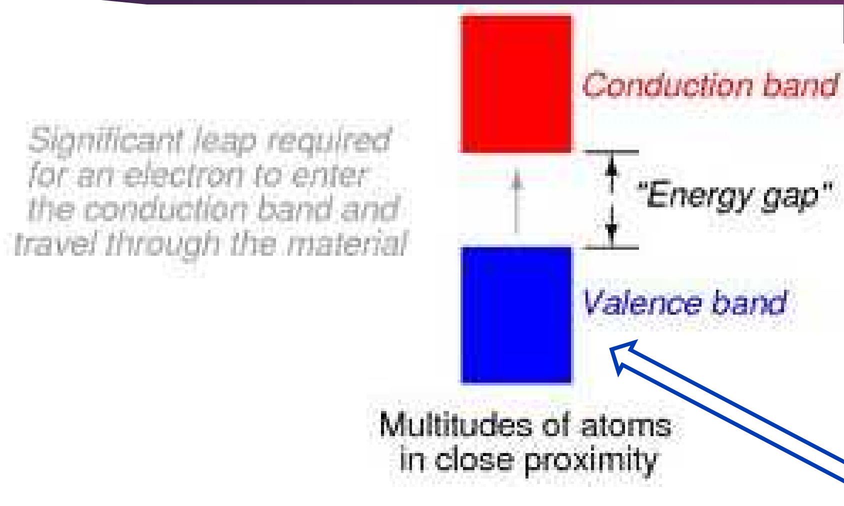
ატომების ენერგეტიკული სტრუქტურა წარმოადგენს დისკრეტულ (წყვეტილ) დონეებს



მყარ სხეულებში თითოეული ატომური დონე გარდაიქმნება ზონად. ზონები შეიცავენ $\sim 10^{22}$ დონეს, რომლებიც იმდენად ახლოს არიან ერთმანეთთან, რომ შეიძლება უწყვეტად ჩაითვალოს.



ენერგეტიკული ზონები გამოყოფილია ე.წ. ენერგეტიკული ღრეჩოებით.

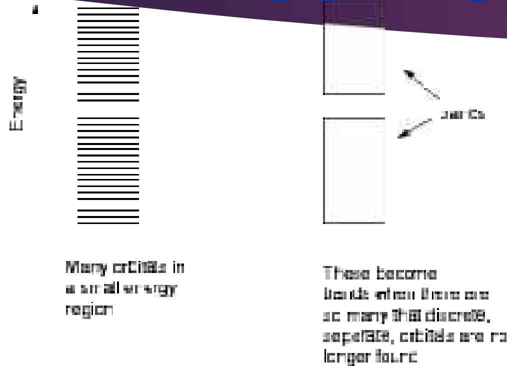


ელექტრონები არიან თავისუფლები და შეუძლიათ დენის გადატანაში მონაწილეობის მიღება. ქვედა ზონიდან გადასვლა ხდება ტემპერატურის ან დასხივების (ან სხვ.) საშალებით.

ელექტრონები რომლებიც მონაწილეობენ ქიმიურ ბმაში. დაბალ ტემპერატურებზე თითქმის ყველა ელექტრონი იმყოფება ამ ზონაში.

ამ ელექტრონებს არ შეუძლიათ დენია გადატანაში მონაწილეობის მიღება, იმიტომ რომ ისინი არ არიან თავისუფლები.

კვანტური ნანოსტრუქტურები



ჩვენ შეგვიძლია შევაფასოთ რამდენად ახლოს არიან ენერგეტიკული დონეები ერთმანეთთან.

$$E_n = \frac{\hbar^2 f^2 n^2}{2mL^2}$$

$\hbar = 1.054 \times 10^{-34}\ J \cdot s$ პლანკის მუდმივა

$m = 9.31 \times 10^{-31}\ kg$ ელექტრონის მასა

$$E_1 = \frac{\hbar^2 f^2 1^2}{2mL^2}$$

$$E_2 = \frac{\hbar^2 f^2 2^2}{2mL^2}$$

$L -$ კრისტალის ზომა

$$L = 10\sim m = 10^{-6} m,\ E_2 - E_1 \approx 10^{-6}\ eV$$

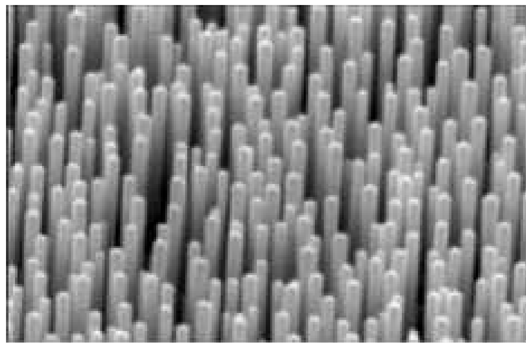
$$L = 10nm = 10^{-9} m,\ E_2 - E_1 \approx 1 eV$$

როცა კრისტალის ზომა შემცირებულია ნანომეტრის რიგამდე ენერგეტიკული სპექტრი ხდება დისკრეტული, ატომის მაგვარი.
ნანოსტრუქტურებს ხშირად უწოდებენ "ხელოვნურ ატომებს".

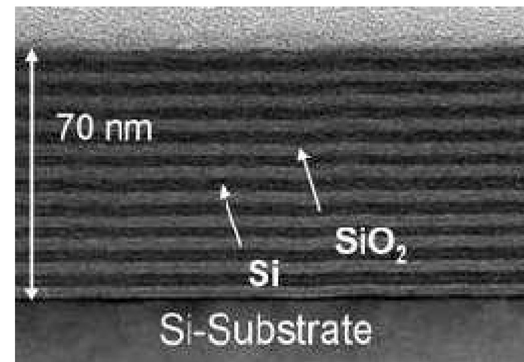
ნანოსტრუქტურების ტიპები

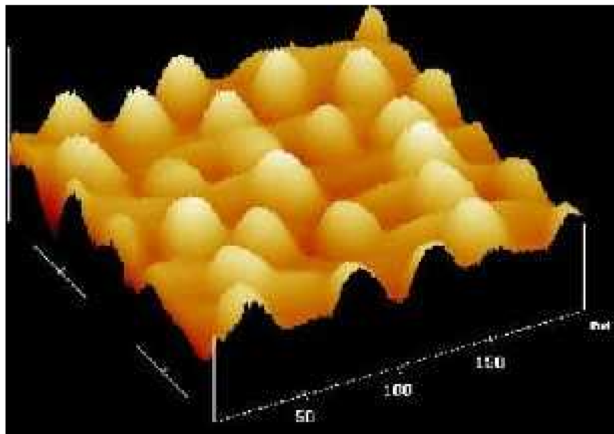
ნანოსტრუქტურების 3 ძირითადი ტიპი:

- **კვატური ჭები (ფენები)** მხოლოდ ერთი განზომილება არის ნანომეტრის რიგის, დანარჩენი 2 მაკროსკოპულია.
- **კვანტური ღეროები და სადენები** – 2 განზომილება არის ნანომეტრის რიგის, ერთი მ,აკროსკოპულია
- **კვანტური წერტილები** – სამივე განზომილება ნანომეტრის რიგისაა.



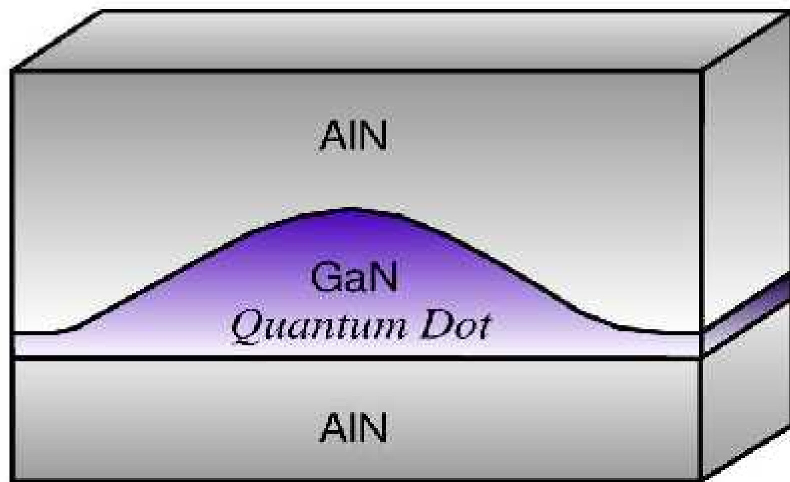
1um

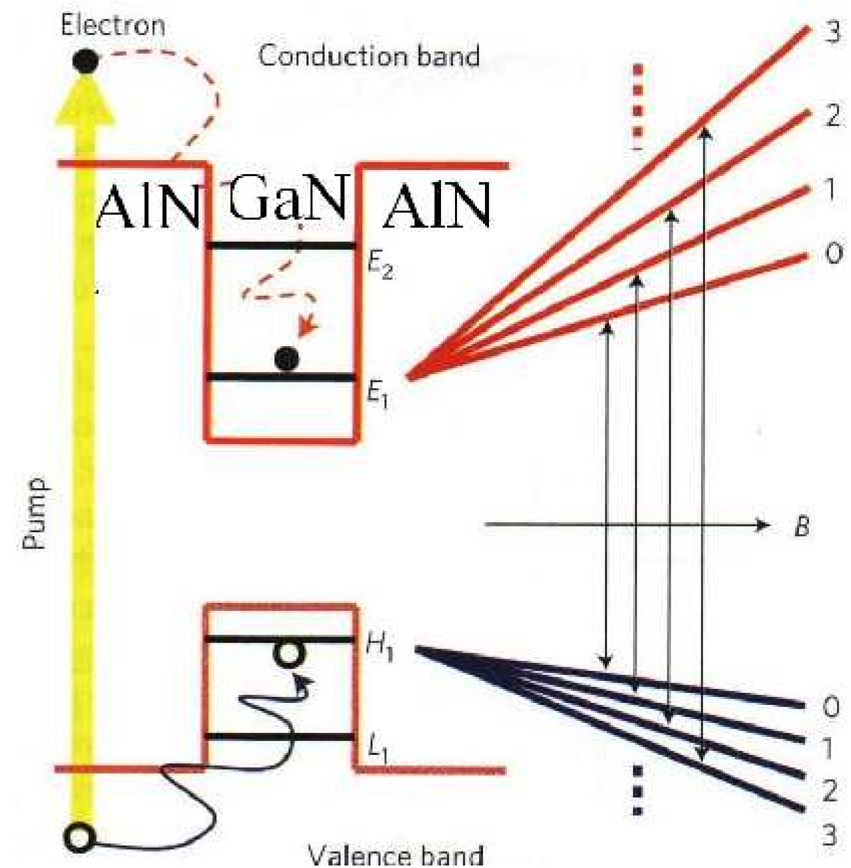




ნანოსტრუქტურების მთავარი უპირატესობა – მათი თვისებები დამოკიდებულია ზომაზე – ჩნდება მასალის თვისებების კონტროლის დამატებითი პარამეტრი

როგორ იქმნება ნანოსტრუქტურები





MATERIAL SCIENCE ON NANOSCALE

TAMAR TCHELIDZE

FACULTY OF EXACT AND NATURAL SCIENCES

Content

Introduction **3**

Chapter 1. Bulk Materials **5**

Chapter 2. Quantum nano-structures and their growth **19**

Chapter 3. Chapter 3 Electrons in low Dimensional Structures **43**

Chapter 4. Electrons in low Dimensional Structures of different shape and material distribution **66**

Chapter 5. Electrons in Quantum Semiconductor Structures: More Advanced Systems and Methods **83**

Chapter 6. Superlattices **113**

Chapter 7. Phonons **129**

Chapter 8. Optical Properties **139**

Chapter 9. Localization and Quantum Transport **157**

Chapter 10. Application **165**

Appendix 1. Crystal structure and Materials **175**

Appendix 2. **187**

INTRODUCTION

Let's begin our course with definition of nanoscience and nanotechnology that are main aspects of nanomaterial science.

Nanoscience is the study of phenomena and manipulation of materials at atomic, molecular and macromolecular scales, where properties differ significantly from those at a larger scale.

Bulk materials (object around us, crystals of micrometer size) possess continuous physical properties. But when particles assume nanoscale dimension, their quantities describing physical properties became quantized. The same material (e.g., gold) at the nanoscale can have properties (electrical, optical, etc.) which are very different form (even opposite to) the properties the material has at the macro-scale (bulk).

Nanotechnologies are the design, characterization, production and application of structures, devices and systems by controlling shape and size at nanometer scale.

In today's scientific realm, the prefix "nano" describes physical lengths that are on the order of a billionth of a meter (i.e. 10^{-9} m). The size range is set normally to be minimum 1nm to avoid single atoms or very small groups of atoms being designated as nano object. The upper limit is normally 100 nm. A valid question would be "why 100 nm, and not 150 nm?" The reason why the 1-100 nm range is approximate is that the definition "nano" itself focuses on the effect that the dimension has on a certain material. Nanoscience is not just the science of the small, but the science in which materials with small show new physical phenomena, collectively called quantum effects, which are size-dependent and dramatically differ from the properties of macro-scale materials. **In other words in nanoscience considers the size range where materials properties are sensitive to system size.** This size range is different for different materials. Bellow we will have more precise definition of lower and upper size limits of structures under consideration.

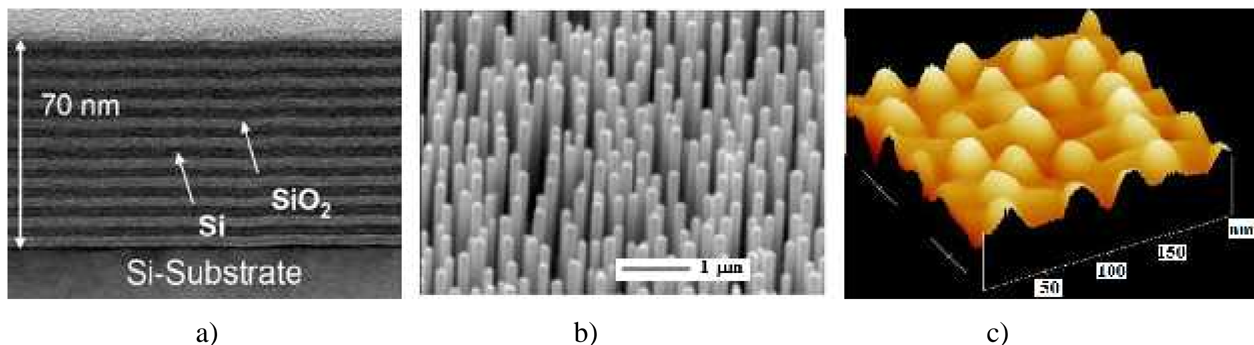
Nanoscale materials lie in a physical size regime between bulk, macroscale, materials and molecules of atoms. Nanoscale physics, chemistry, biology, and engineering ask basic, yet unanswered question such as how the optical and electrical properties of materials evolve from those of individual atoms or molecules. Other questions being asked include the following:

- How does one actually go about making a nanometer-sized objects and how does one make many such identical objects?
- How do their optical and electrical properties change with "dimensionality"?
- How do charges move in such nanoscale systems?
- Do these materials possess new and previously undiscovered properties?
- Are they useful?

In the presented course, basics physical concepts, which help us so answer these questions, are considered.

Nanostructures are classified according the number of reduced to nanometer dimensions:

1. **Quantum well** – with only one dimension of nanometer size. This actually is a sheet with nanometer thickness (Figure 1. a)) – so called 2 dimensional structures.
2. **Quantum wires** – with two dimensions of nanometer size. (Figure 2. b))
3. **Quantum dots** – all three dimensions are of nanometer size. (Figure 2. c))



a) b) c)

Figure 1. a) quantum well – electrons are confined in Si layers; b) quantum wires; c) quantum dots.

When we speak about nanometer size it is obvious that the whole samples cannot have such small sizes. Nanometer size is an area where the electron (the main player in electronics) is confined. Bellow we will see that when electron is confined in the area small area its energy levels become discrete (atomic like). The distance between the levels depends on size of the area where electron is confined. Quantum mechanical consideration (see below) tells us that the narrower localization area the higher is the kinetic energy of electron. **There exists the size for which electron has such high kinetic energy that it cannot be localized. Actually this is a lower size limit.** It varies from material to material. With increasing the localization size the distance between the neighboring energy levels ΔE . **The upper limit is usually defined from the condition** $\Delta E \cong k_B T$ **, i. e. energy level separation is order of thermal energy.**

In the next chapter we will discuss energetic structure of bulk materials, in order to understand how properties of materials change with size reduction, as well as to learn what the mechanism of electron localization in quantum nanostructures.

Chapter 1. Bulk Materials

The most of materials used in electronics are crystals. Therefore we will consider here properties of crystalline materials. The main question to ask when studying crystals is: how do properties of solid differ from those of its constituent atoms or molecules? We will highlight here five aspects that make the physics of the crystalline state interesting and different, namely

- Crystal symmetry
- Electronic bands
- Vibronic bands
- The density of states
- Delocalized states and collective excitations.

1.1 Crystal symmetry

Crystals have long range translational order, and short range order characterized by point groups. The point group symmetry refers to the group of symmetry operators that leaves the crystal invariant. Examples of this include rotations about particular axes, reflections about planes and inversion about points in the unit cell.

Let's consider crystal structure. We introduce the concept of crystal lattice. Crystal lattice is a set of equivalent points to which atoms or group of atoms (basis) are bound. The equivalence of crystal lattice points are conditioned by the identity of atoms or groups of atoms, and their symmetrical arrangement (Figure 1.1).

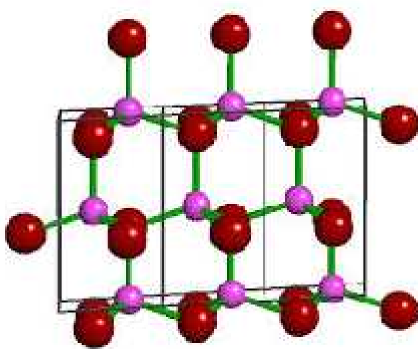


Figure 1.1 crystal structure = crustal structure + basis

From any point of crystal lattice with position vector $\overrightarrow{r_o}$ it is possible to transfer any other lattice point by means of lattice or translation vector defined as

$$\vec{n} = n_1\overrightarrow{a_1} + n_2\overrightarrow{a_2} + n_3\overrightarrow{a_3} \ . \tag{1.1}$$

In this expression n_1, n_2, n_3 are arbitrary integers, $\overrightarrow{a_1}, \overrightarrow{a_2}, \overrightarrow{a_3}$ are named as main translations. Proceeding from the mentioned above, if $\overrightarrow{r_o}$ is a lattice point, $\vec{r} = \overrightarrow{r_0} + \vec{n}$ corresponds to lattice point too.

$\overrightarrow{a_1}, \overrightarrow{a_2}, \overrightarrow{a_3}$ vectors define parallelogram, which we call unit cell. Proceeding from the definition of $\overrightarrow{a_1}, \overrightarrow{a_2}, \overrightarrow{a_3}$ vectors, by the repetition of unit cell to fill all space is possible (Figure.1.2).

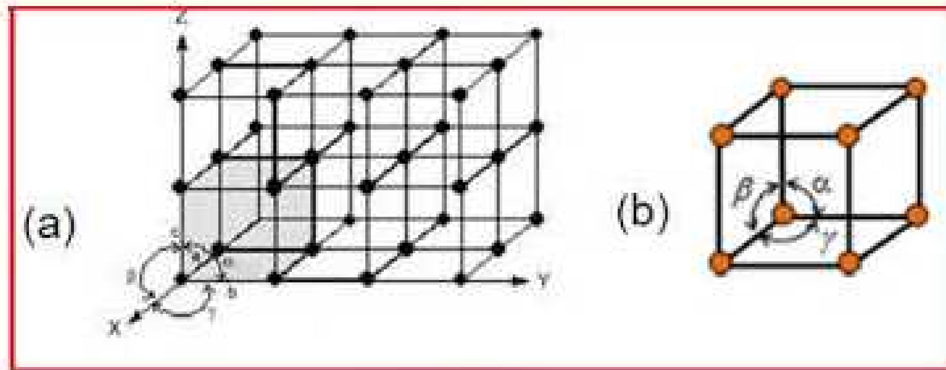


Figure 1.2 a) whole crystal; b) unit cell

Crystal point group symmetry is defined by relations between the length of $\overrightarrow{a_1}, \overrightarrow{a_2}, \overrightarrow{a_3}$ vectors and angles between them. $\overrightarrow{a_1}, \overrightarrow{a_2}, \overrightarrow{a_3}$ vectors identify the symmetry operators – rotations inversions reflections, which leave crystal invariant. Consequently, these relations define

the crystal systems, which differ from each other by set of symmetry operators. There are seven crystal systems and they are given in Figure 1.3.

The link between the measurable properties and the point group symmetry of a crystal can be made through *Neumann's principle.* This states that:

Any macroscopic physical property must have at least the symmetry of the crystal structure.

For example, if a crystal has four-fold rotational symmetry about a particular axis (4 rotations around of this axis leave crystal invariant), then we must get the same results in any experiment we might perform in four equivalent orientations.

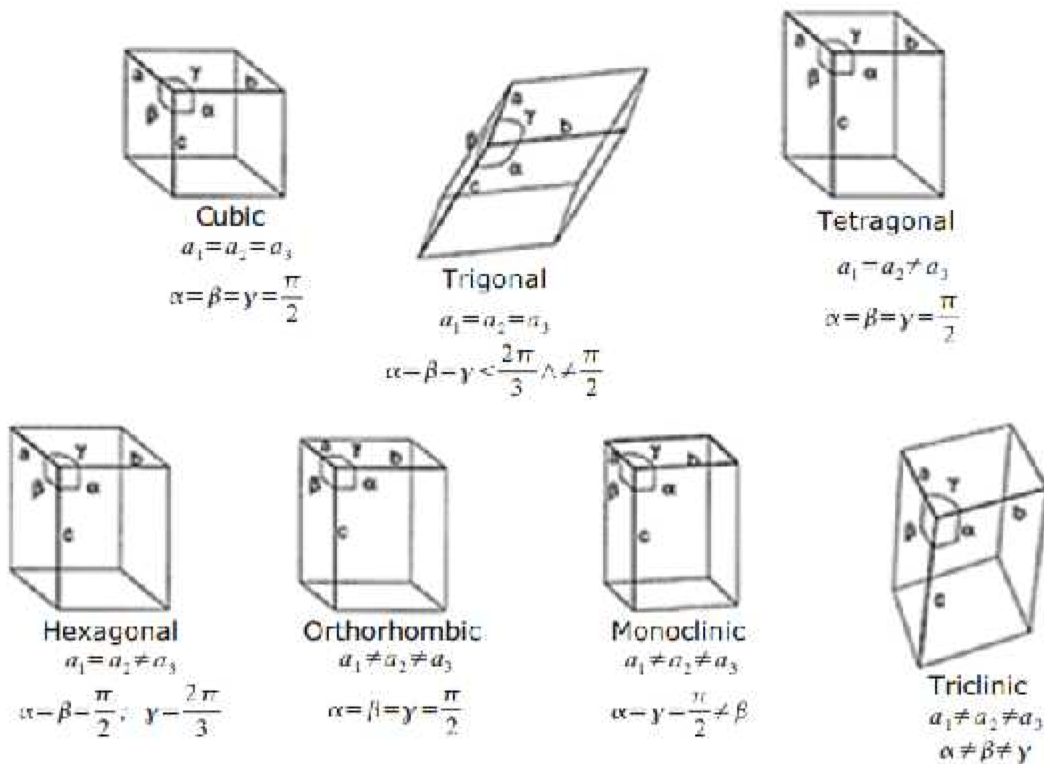


Figure 1.3 Seven crystal systems

It is instructive to compare the properties of a crystal to those of the atoms from which it has been formed. A gas of atoms has no translational order. Therefore in crystal we expect new properties connected to translational symmetry. Besides, the point group symmetry of a crystal is lower than that of the individual atoms, which have the highest possible symmetry due to their spherical invariance. We therefore expect to find other effects in the solid state that relate to the lowering of the symmetry on going from free atoms to the particular point group of the crystal class. The lifting of degeneracies by reduction of the symmetry is a well-known effect in atomic physics. Free atoms are spherically symmetric and have no preferred directions. The symmetry can be broken by applying an external magnetic or electric field which creates a preferred axis along the field direction. This can lead the lifting of certain level degeneracies that are present in the free atoms. The Zeeman effect, for example, describes the spitting of degenerate magnetic levels when a magnetic field is applied. If the same atom is introduced into a crystal, it will find itself in an environment with point group symmetry. This symmetry is lower than that of the free atom, and therefore some level degeneracies can be lifted. This point is illustrated in Fig 1.4. The splitting is caused by the interaction of the orbitals of the atoms with the electric fields of the crystalline environment. The character of splitting depends on the point group of a crystal. The higher the symmetry of a point group the higher is the remaining degeneracy.

It should be mentioned that many important materials do not posses long range translation symmetry. Glass is an obvious example. The energetic structure of these materials may be very similar to those of their constituent atoms and molecules. Point group symmetry mainly defines the peculiarities of electronic and vibronic spectra in such type of materials.

Long range order, which is characteristic for majority of solid materials, is revealed in that system is invariant with respect to $\vec{r} \to \vec{r} + \vec{n}$ transformation. This means that two points into crystal separated be lattice vector are equivalent. Obviously, because of this new type of symmetry new conserved measurable physical quantity should appear. Energies and wave functions of electrons become dependent on new additional quantum number. This new quantum number is named as quasi momentum, by analogy to ordinary momentum, which is connected to invariance with respect to $\vec{r} \to \vec{r} + d\vec{r}$ infinitely small translation.

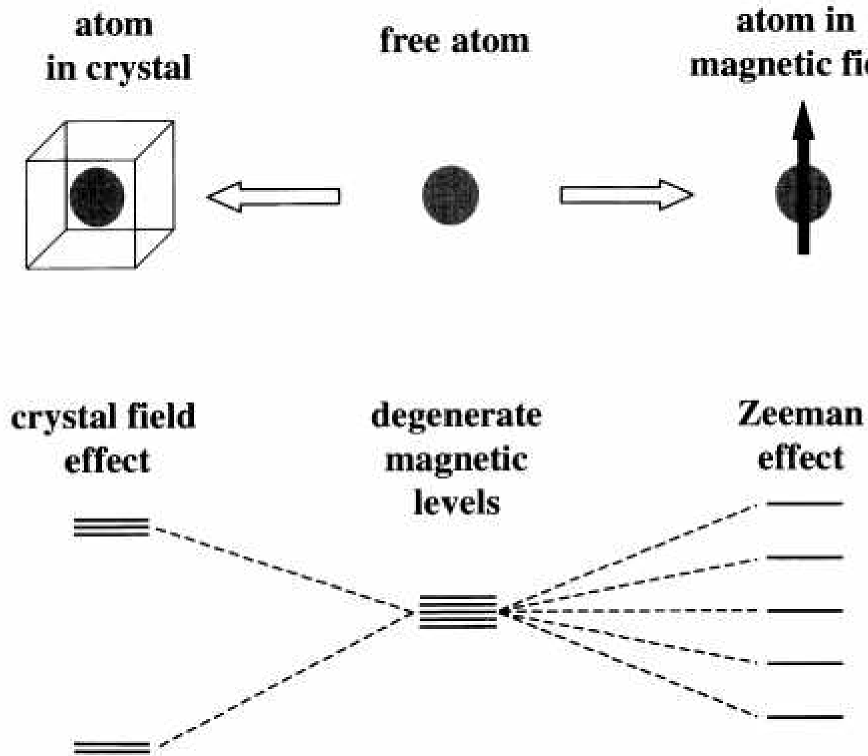


Figure 1.4 Splitting of the magnetic levels of free atom by the crystal field effect and external magnetic field.

1.2 Energetic bands

The atoms in a solids are packed very close to each other, with the interatomic separation approximately equal to the size of atoms. Hence the outer orbitals of the atoms overlap and interact strongly with each other. This broadens the discrete levels of the free atoms into bands, as illustrated in Fig.1.5

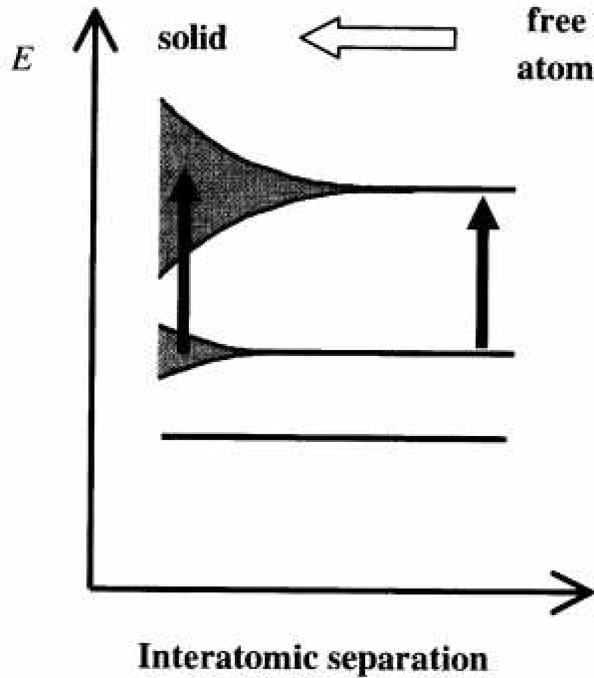


Figure 1.5 Schematic diagram of the formation of electronic bands in a solid from the condensation of free atoms

How does energy levels depend on quasi momentum (or wave vector)?

This can be estimated in different approximation.

1.3 Free electron model

Translational symmetry causes delocalization of electronic states, which is connected to the fact that due to the equivalence of crystal points probability of finding of electron in different lattice point is the same. However, in free elector model we did not account for microstructure of the crystal. Crystal can be treated as potential well for electrons with infinitely high walls. In this case (bulk samples) allowed energy values vary quasi-continuously, and are described by the formula:

$$E_k = \frac{\hbar^2}{2m}\left(k_x{}^2 + k_y{}^2 + k_z{}^2\right) \quad (1.2)$$

If we introduce wavefunctions that satisfy periodic boundary conditions that is if we require wavefunctions to be periodic in *x, y, z* with period

$$\text{Œ}\left(x+L,y,z\right)=\text{Œ}\left(x,y+L,z\right)=\text{Œ}\left(x,y,z+L\right)=\text{Œ}\left(x,y,z\right) \quad (1.3)$$

We obtain

$$k_x,k_y,k_z=0;\pm\frac{2f}{L};\pm\frac{4f}{L}..... \quad (1.4)$$

The corresponding wave functions are wave functions of free

$$\Psi_k\left(\vec{r}\right)=e^{i\vec{k}\vec{r}}. \quad (1.5)$$

This is a traveling wave caring momentum $\vec{p}=\hbar\vec{k}$.

1.4 Nearly free electron model

The band structure of a crystal can often be explained by the nearly free electron model for which the band electrons are treated as perturbed only weakly by the periodic potential of the ion cores. Let's see what introduction of periodic potential will give.

We consider the simplest case – one dimensional chain of atoms. Interatomic separation in this chain is a, i. e. we have one-dimensional crystal lattice with lattice constant a (Fig.1.6 a).

Introducing periodic potential in the model changes the picture in the way that the wavefunctions at $k=\pm f/a$ are not traveling waves $\exp\left(if\,x/a\right)$ or $\exp\left(-if\,x/a\right)$ of free electrons. At these special values of k the wavefunctions are made up of equal parts of waves traveling to the right and left. The wave traveling to the one direction interferes with the wave reflected from the neighboring crystal point.

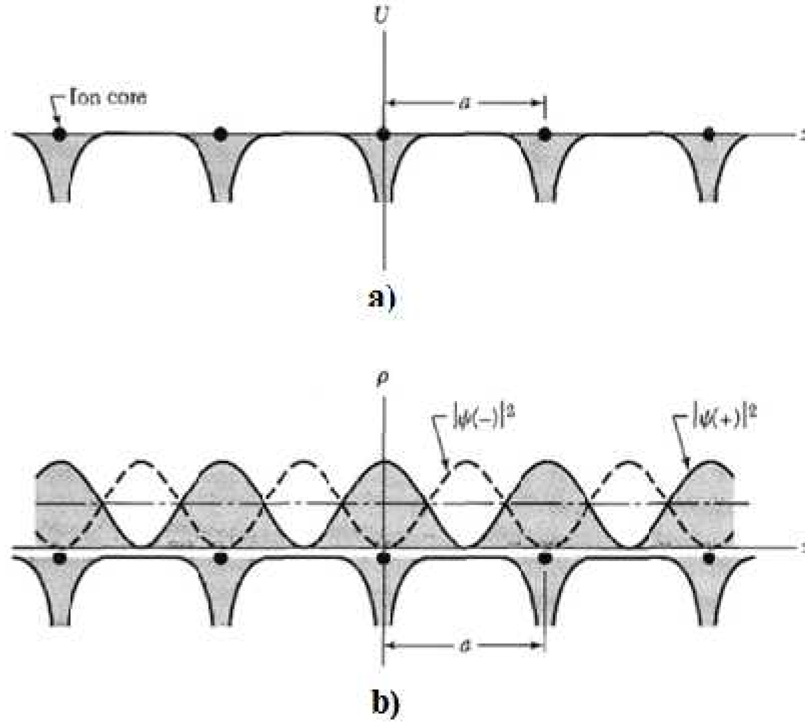


Figure.1.6 a) Variation of potential energy of a electron in the field of the ion cores of a linear lattice; b) Distribution of probability density in the lattice.

For this special value of wave vector constructive interference takes place, because the path difference is an integral number of wavelengths. That is traveling waves form two standing waves:

$$\psi(+) = e^{i\pi x/a} + e^{-i\pi x/a} = 2\cos\frac{\pi x}{a}$$
$$\psi(-) = e^{i\pi x/a} - e^{-i\pi x/a} = 2i\sin\frac{\pi x}{a} \qquad (1.6)$$

These two standing waves pile up electrons at different regions, and therefore they have different values of the potential energy in the field of the ions of the lattice. This is the origin of the energy gap. The probability density ρ of a particle is $|\psi(x)|^2$. For a pure traveling wave probability density is equal to 1. The charge density is not constant for linear combinations of plane waves. For which we have

$$\rho(+) = |\psi(+)|^2 \sim \cos^2 \frac{\pi x}{a}$$
$$\rho(-) = |\psi(-)|^2 \sim \sin^2 \frac{\pi x}{a} \qquad (1.7)$$

The first of these functions piles up electrons on the positive ions, where the potential energy is lowest, while the second one concentrates electrons away from the potential energy minima. Because of different localization of $\psi(+)$ and $\psi(-)$ states, corresponding energies are different. At $k = \pm\pi/a$ energy gap usually denoted as E_g occurs.

We see that using nearly free electron model instead of free electron models explains appearance of energy gap in dispersion relation. (Figure 1.7)

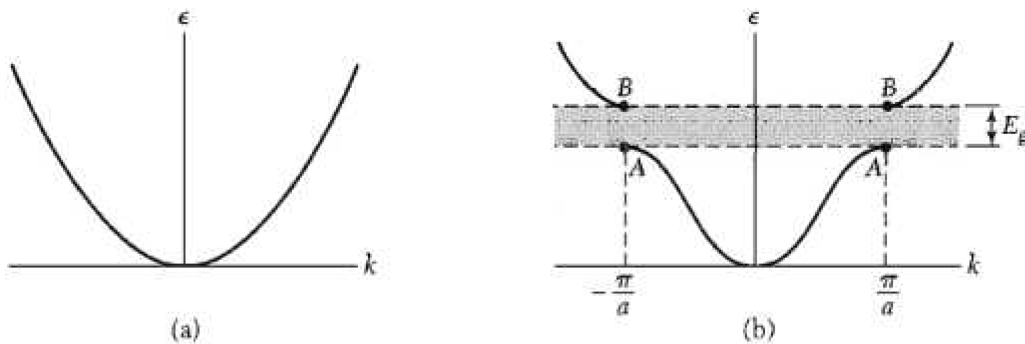


Figure.1.7 a) Plot of energy versus wavevector for a free electron; b) Plot of energy versus wavevector for a electron in a monoatomic chain.

1.5 k·P method

k·P model is one of the most popular in studying bulk and nano-structured semiconductor materials and has been used widely during past decades. This method enables one to find wave functions and energies for electron in periodic structure for any value of wavevector using known energies and wave function at particular point k_0.

For the electron in a periodic potential

$$V\left(\vec{r}\right)=V\left(\vec{r}+\vec{n}\right) \tag{1.8}$$

where $\vec{n}$ is the lattice vector defined by (1.1), the electron wave function satisfies the Schrödinger equation

$$\left[-\frac{\hbar^2}{2m}\Delta+V(\vec{r})\right]\Psi(\vec{r})=E(\vec{k})\Psi(\vec{r}) \tag{1.9}$$

The Hamiltonian is invariant under translation by the lattice vector $\vec{r}\rightarrow\vec{r}+\vec{n}$. If Œ $\left(\vec{r}\right)$ describes an electron moving in the crystal, Œ $\left(\vec{r}+\vec{n}\right)$ will also be a solution to (1.9). Thus Œ $\left(\vec{r}+\vec{n}\right)$ will differ from Œ $\left(\vec{r}\right)$ at most by a constant, which must have a unity magnitude; otherwise, the wave function may grow to infinity if repeat the translation infinitely. The general solution of the above equation is given by

$$\Psi_{nk}(\vec{r})=e^{i\vec{k}\vec{r}}u_{nk}(\vec{r}) \tag{1.10}$$

Where

$$u_{nk}(\vec{r})=u_{nk}(\vec{R}+\vec{n}) \tag{1.11}$$

is a periodic function. This result is the Bloch theorem. The energy is given by

$$E = E_n\left(\vec{k}\right) \tag{1.12}$$

Here n refers to the band and $\vec{k}$ the wave vector of the electron.

The k·P method is a useful technique for analyzing the band structure near a particular point $\vec{k}_0$, especially when it is near an extremum of the band structure. Here we consider that the extremum occurs at the zone center where $\vec{k}_0 = 0$.

Consider the general Schrödinger equation for an electron wave function in the n-th band

$$\left[-\frac{\hbar^2}{2m}\Delta + V(\vec{r})\right]\Psi_{nk}(\vec{r}) = E_n(\vec{k})\Psi_{nk}(\vec{r}) \ . \tag{1.13}$$

If substitute (1.10) in this formula

$$\left[-\frac{\hbar^2}{2m}\Delta + V(\vec{r})\right]e^{i\vec{k}\vec{r}}u_{nk}(\vec{r}) = E_n(\vec{k})e^{i\vec{k}\vec{r}}u_{nk}(\vec{r}) \tag{1.14}$$

we obtain equation in terms of $u_{nk}(\vec{r})$

$$\left[-\frac{\hat{p}^2}{2m} + V(\vec{r}) + \frac{\hbar}{m}\left(\vec{k}\hat{p}\right)\right]u_{nk}(\vec{r}) = \left(E_n(\vec{k}) - \frac{\hbar^2 k^2}{2m}\right)u_{nk}(\vec{r}) \ . \tag{1.15}$$

The above equation can be expanded near a particular point $\vec{k}_0$ of interest in the band structure. When $\vec{k}_0 = 0$ the above equation is expanded near $E_n(0)$,

$$\left[\widehat{H}_0 + \frac{\hbar}{m}\left(\vec{k}\,\widehat{p}\right)\right]u_{nk}(\vec{r}) = \left(E_n(\vec{k}) - \frac{\hbar^2 k^2}{2m}\right)u_{nk}(\vec{r}) \tag{1.16}$$

Where

$$\begin{aligned} H_0 &= -\frac{\widehat{p}^2}{2m} + V(\vec{r}) \\ \vec{k}\cdot\widehat{p} &= k_x\left(i\hbar\frac{\partial}{\partial x}\right) + k_y\left(i\hbar\frac{\partial}{\partial y}\right) + k_z\left(i\hbar\frac{\partial}{\partial z}\right) \end{aligned} \tag{1.17}$$

Our goal is to find $u_{nk}(\vec{r})$ functions and corresponding energies, which enable one to find dispersion relation for definite zone (for definite n) For $\vec{k} = 0$ (1.16) equation has the form

$$\widehat{H}_0 u_{no}(\vec{r}) = E_n(0)u_{no}(\vec{r}) \tag{1.18}$$

Let's assume that we know the solution of (1.18). If interaction between the zones is small (distance between them is large) the solution of (1.16) can be found in the frame of perturbation theory. In this case perturbation is equal to

$$W = \frac{\hbar}{m}\left(\vec{k}\,\widehat{p}\right) + \frac{\hbar^2 k^2}{2m} \tag{1.19}$$

Using standard procedure of perturbation theory, close to $\vec{k} = 0$, E_{nk} and corresponding u_{nk} are written as

$$E_{nk} = E_{n0} + \frac{\hbar^2 k^2}{2m} + \frac{\hbar^2}{m^2}\sum_{n\neq n'}\frac{\left|\left\langle u_{n0}\left|\vec{k}p\right|u_{n'0}\right\rangle\right|^2}{E_{no} - E_{n'0}} \tag{1.20}$$

$$u_{nk} = u_{n0} + \frac{\hbar}{m}\sum_{n\neq n'}\frac{\left\langle u_{n0}\left|\vec{k}p\right|u_{n'0}\right\rangle}{E_{no} - E_{n'0}} \tag{1.21}$$

As $\langle u_{n0} | \vec{k} p | u_{n0} \rangle$ matrix elements are equal to 0. In the expression for energy there are no linear terms of k.

Formula (1. 20) can be rewritten as

$$E_{nk} = E_{n0} + \frac{\hbar^2 k^2}{2m^*} \tag{1.22}$$

where

$$\frac{1}{m^*} = \frac{1}{m}\left[1 + \frac{2}{mk^2}\right] \sum_{n \neq n'} \frac{\left| \langle u_{n0} | \vec{k} p | u_{n'0} \rangle \right|^2}{E_{no} - E_{n'0}} \tag{1.23}$$

This expression can be treated as so called inverse effective mass of electron in crystal. It indicates that in crystal, electron effective mass is different from mass of free electron. The reason is mixing the states of different because of mixing of states of different zones by means of $\vec{k} p$ term. As can be seen the effect depends on

- $\langle u_{n0} | \vec{k} p | u_{n'0} \rangle$ matrix element
- $E_{no} - E_{n'0}$ differences. The mixing with the zones energy of which is less than E_{no} decrease the effective mass, while mixing with the states with energy higher than E_{no} increase the effective mass, or may make it even negative.

Let's consider conductive band. The conductive band states mainly interact with the valence band states. From the symmetry considerations valence band states, which are usually p-type (generated from l=1 atomic states) and are denoted as $|X\rangle, |Y\rangle$ and $|Z\rangle$. We denote the conductive band state as $|S\rangle$ (they generate mainly from l =0 atomic orbitals). From the symmetry considerations we can estimate matrix elements. Only those given bellow are not equal to zero.

$$\langle S | p_x | X \rangle = \langle S | p_y | Y \rangle = \langle S | p_z | Z \rangle \equiv iP \quad . \tag{1.24}$$

Using (1.24) from (1.23) we obtain

$$\frac{m}{m^*} = 1 + \frac{2P^2}{mE_g} \quad . \tag{4.14}$$

Here E_g is so called band gap – the difference between the conductive and highest valence band at $\vec{k} = 0$ we see that in this approximation effective mass does not depend on wave vector – we have parabolic dispersion relation. Generally an effective mass is wave vector dependent and dispersion relation is quite complicated. In Figure 1.8 the band structure of Si is given. , , , , L denote the points of different symmetry in k space. We can see that close to point ($\vec{k} = 0$) dispersion relation is parabolic indeed. However, away from this point, there is deviation from the parabolic shape.

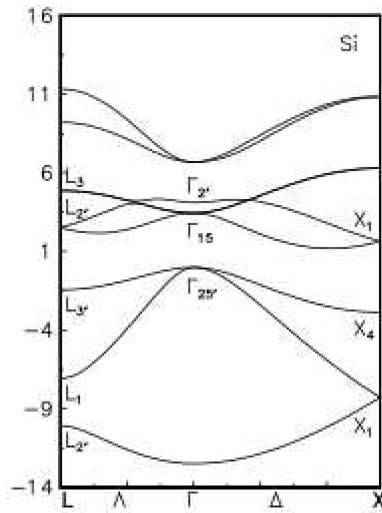


Figure 1.8 E=E(k) relations for Si.

Finally, it should be mentioned that the values of quasi wave vector equal to $\pm\frac{f}{a}$ limit the area called the first Brillouin zone. In this region all un-equivalent values of wave vector are places.

Chapter 2. Quantum Nano-structures, Their Growth and Structural Characterization

In the late 1960s new electronic and optical phenomena was founded on a suggestion by Leo Esaki and Raphiel Tsu, than working at the IBM Research Laboratories. They proposed that structures composed of layered regions of semiconductors with different band gaps would have a spatially varying potential energy surface that would confine carriers to the narrower band-gap material. If there were few enough adjacent layers of this material, then the carriers could be confined within regions comparable to their de Broglie wavelength – the natural length scale that governs their quantum mechanical behavior. For this reason, these narrow regions are now called "quantum wells". Electrons and holes in quantum wells were predicted to exhibit remarkable optical and transport properties that could be controlled by varying the width of the well and their barriers. Figure 2.1 presents AlGaAs/GaAs quantum well and its energetic diagram.

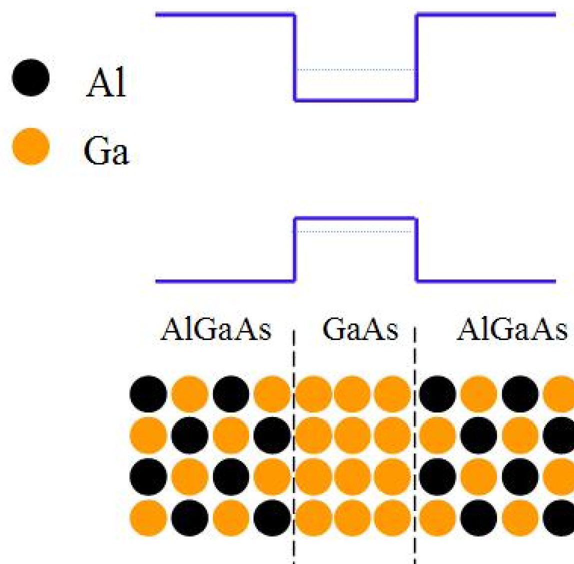


Figure 2.1 AlGaAs/GaAs quantum well

At the time that Esaki and Tsu made their proposal, the available technology could not produce materials of sufficient quality to verify the predicted effects. Nowadays **Epitaxial growth techniques** gives us the possibility to obtain layered structures, which have already gained practical application. Advances in epitaxial growth techniques have played a pivotal role in the realization of ever more ambitiously designed quantum heterostructures. In such structures the confinement of electron motion in one, two or three dimensions occurs, and the effective reduction of dimensionality takes place. In this chapter we present epitaxial growth techniques, epitaxial growth modes, as well as experimental methodologies that have been developed to achieve the required control over composition, doping, and interface characterization.

2.1 Molecilar Beam Epitaxy

Molecular beam epitaxy (MBE) is the simplest and one of the widely used way of fabricating semiconductor heterostructures. MBE is essentially two-step process carried out in an ultra-high vacuum environment. In the first step, atoms or simple homoatomic molecules which are constituent of the growing material (e.g. atomic Ga and either As2 for GaAs) are evaporated from solid sources and directed toward a heated substrate, which is typically a few centimeter in size (Figure 2.2). The particles within these beams neither react nor collide with one another, i.e. the deposition onto the deposition onto the substrate is ballistic and the particles are said to undergo molecular flow – hence the name *molecular-beam epitaxy*. The substrate is often rotated for more uniform deposition rates across the substrate.

The second step of MBE is the migration of the deposited species on the surface prior to their incorporation into the growing material. This determines the profile, or *morphology,* of the film and its effectiveness depends on a number of factors, including the deposition rates of the constituent species, the surface temperature, the surface material, and its crystallographic orientation. The dependence of the morphology on the deposition rate of new material means that MBE (as well as other epitaxial growth techniques) are inherently *nonequilibrium*, or *driven,* process. Growth near equilibrium is governed almost exclusively by thermodynamics. For epitaxial growth, thermodynamics still provides the overall driving force for the

morphological evolution of the surface, but the extent to which equilibrium is attained even locally is mediated by kinetics, i. e. the rates of processes that determine how a system evolves.

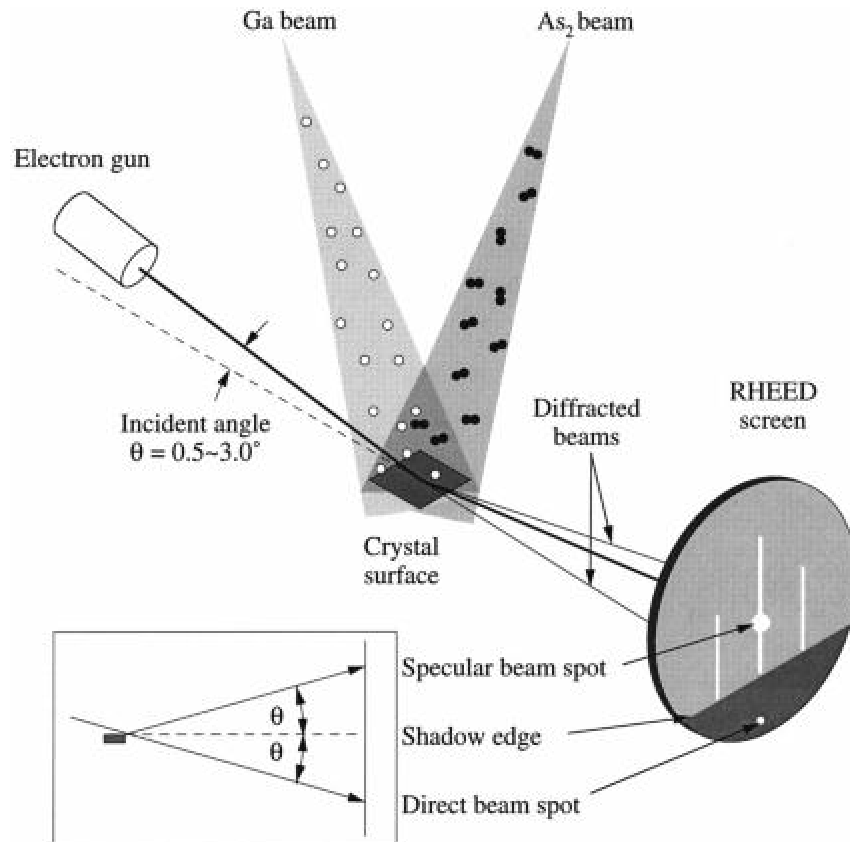


Fig. 2.2 The arrangement of the substrate, the reflection high-energy electron diffraction (see below) measurement apparatus, consisting of an electron gun and collector screen.

A major strength of MBE is that the ultra-high vacuum environment enables the application of *in situ* surface analytical techniques to characterize the evolution of the growing material at various level of resolution – from microns to the arrangements of atoms. These techniques will be discussed below.

2.2 Vapour-phase Epitaxy

An alternative to deposition by molecular beams is hydrodynamic transport of material to the substrate from gas source. In the scenario, which is called **vapour-phase epitaxy (VPE),** the constituent of the growing surface are delivered within heteroatomic molecules called *precursors*. For group IV group materials, the precursors are hydrides, chlorides, or chloro-hydrides. The growth of III-V materials uses precursors for the group species, which contain carbon, and the V elements are supplied as hydrides.

The pressures inside a vapour-phase reactor can vary from 10^{-2} torr up to atmospheric, so the flow of the gas is viscous and the chemicals reach the substrate by diffusion through a boundary layer. Thus, the delivery of material to the growing film encompasses gas phase and surface chemical reactions, as well as mass transport within fluid as it flows through the reactor, the latter being highly dependent on the system pressure and reactor design.

The use of gas sources has several attractive features for the epitaxial growth of semiconductor heterostructures. They can be used at room temperature, thus causing less contamination than high-temperature sources, and with a very simple reactor design, can give a more uniform flux than of a molecular beam, so that the surface does not need to be rotated. An operational advantage over MBE is that, because there is no depletion, the growth chamber does not need to be opened and exposed to air to replenish the source material. An important practical disadvantage of vapour-phase is that the gas sources can be highly-toxic.

2.3 Epitaxial growth Modes

As mentioned for obtaining nanostructures two different materials with different band gap are needed. Numerous experiments have revealed that, for small amounts of one material deposited onto the surface of another material (similar or different), the epitaxial growth is one of the three distinct types:

- *Frank-van Merve* morphology, with flat single crystal films consisting of successive complete layers;
- *Volmer-Weber* morphology, with three-dimensional (3D) islands that leave part of the substrate exposed;
- *Stranski-Krastanov* morphology, with 3D islands atop athin flat “wetting” film that completely covers the substrate. These morphologies are illustrated in Figure 2.3

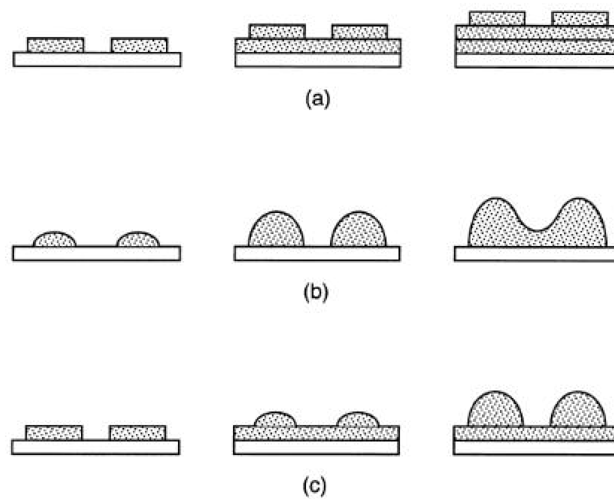


Figure 2.3 (a) - Frank-van der Merve, (b) - Volmer-Weber, (c) Stransky-Krastanov – morphologies.

For lattice-matched systems, the Frank-van der Merve and Volmer-Weber morphologies can be understood from thermodynamic wetting arguments based on interfacial free energies. We denote the free energy of the epilayer/vacuum interface by χ_e, that of the epilayer/substrate interface by χ_i, and that of the substrate/vacuum interface by χ_s, the Frank-van der Merve growth mode is favored if

$$\chi_e + \chi_i < \chi_s \qquad (2.1)$$

In this case, as the epilayers are formed, the free energy *decreases* initially before attaining a steady-state value for thicker films. Alternatively, if

$$\chi_e + \chi_i > \chi_s \tag{2.2}$$

then Volmer-Weber growth is favooured. Here, the free energy *increases* if epilayers are formd on the substrate, rendering a uniform layer thermodynamically unstable against a break-up into regions where substrate is covered and those it is uncovered.

The Stanski-Krastanov morphology is observed in systems where there is appreciable lattice mismatch between the epilayer and the substrate. This growth mode is thought to be related to the accommodation of the resulting misfit strain, which changes the balance between the surface and interfacial free energies as the strain energy increases with the film thickness. Thus, although the growth of "wetting" layers is favored initially, the build-up of strain energy eventually makes subsequent layer growth unfavorable. Beyond this point, the deposition of additional material leads to the appearance of 3D islands within which atrain is relaxed through the formation of misfit dislocations. However, there is another scenario within the Stanski-Krastanov morphology: the formation of islands without dislocations – called coherent islands (Fig. 2.4) – atop one or more wetting layers. This phenomenon, which has been observed for a number of systems, has many applications.

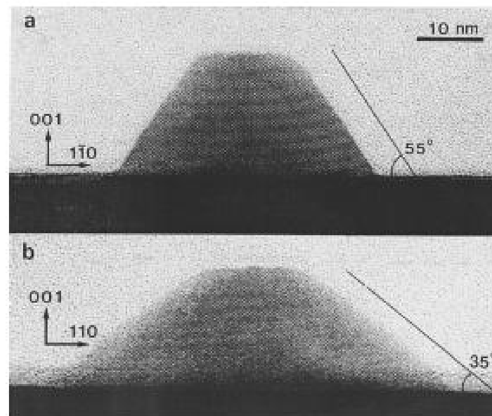


Figure 2.4 Coherently strained islands atop wetting layer.

Figure 2.5 presents the schematic dependence of the film chemical potential on the film thickness in number of minolayers for the three modes of growth: Volmer-Weber (VW), Stransky-Krastanov (SK), and Frank-van der Merwe (FM).

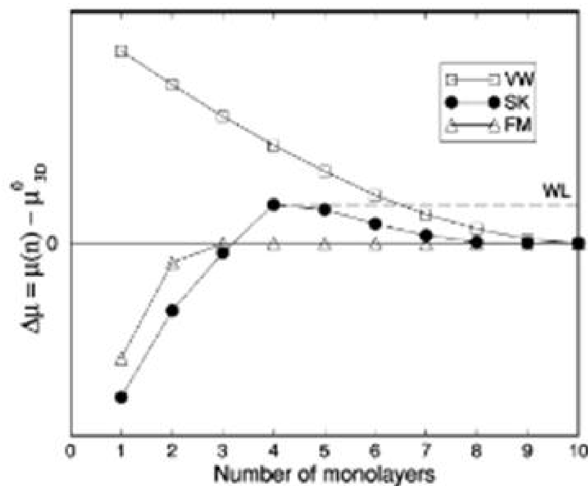


Figure 2.5 dependence of the film chemical potential on the film thickness for the three growth modes. The dashed line gives the chemical potential of the unstable wetting layer.

2.4 Growth Kinetics

The fabrication of heterostructures requires growing crystalline materials on the surface of *different* materials, a process which is known as *heteroepitaxy.* But a useful starting point for understanding heteroepitaxial phenomena is *homoepitaxy* – the growth of a material on a substrate of the same material. Many atomistic processes that occur during heteroepitaxy have direct analogues in homoepitaxy. Many atomistic processes that occur during heteroepitaxy have direct analogues in homoepitaxy.

We develop two most analytic descriptions of epitaxial kinetics: the Burton-Cabrera-Frank (BCF) theory and homogeneous rate equations.

Burton-Cabrera-Frank theory

The BCF theory describes growth on a vicinal surface of a monoatomic crystal in terms of the deposition and migration of single adatoms. The central quantity is, therefore, the adatom concentration $c(\mathrm{x},\mathrm{t})$ at position x and time t. The processes which causes this quantity to change are t

Diffusion of adatoms, which have diffusion constant D, and flux J of adatoms onto the surface. From the molecular beam. We will assume that the desorption of the atoms from the surface can be neglected, but this can be readily included in the theory if required. In the simplest form of the BCF theory, the equation determining $c(\mathrm{x},\mathrm{t})$ on a terrace is a one-dimensional diffusion equation with a source term:

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} + J \qquad (2.3)$$

This equation is supplemented by boundary conditions at the step edges which bound the terrace

$$C(0,t) = 0, \qquad C(L,t) = 0 \qquad (2.4)$$

where L is the terrace length. These boundary conditions stipulates that adatoms are absorbed at step and immediately incorporated into the growing crystal with no possibility of subsequent detachment. We will focus on the steady-state solution of equation (2.3). With (2.4) boundary condition (2.4), we obtain

$$C(x) = (J/2D)x(L-x) \qquad (2.5)$$

which is a parabola with its maximum at the centre of the terrace and which vanishes at the terrace edge, according to boundary condition. The scale of the adatom concentration is set by the growth conditions (substrate temperature and flux) through the ratio J/D. This quantity is a measure of the competition between the deposition flux, which drives the surface away from equilibrium and *increases* the adatom density, and the relaxation of the surface toward equilibrium through adatom diffusion, which *decreases* the adatom density. Since the theory neglects interactions between adaatoms, the growth conditions must be chosen to ensure that the adatom concentration is maintained low enough to render their interactions unimportant. Thus th BCF theory is valid only for relatively small values of J/D, i.e. high temperatures and/or low fluxes, where growth is expected to occur by so called *step flow*.

Homogenous rate equations

With increasing temperature or decreasing deposition rate, growth by the nucleation, aggregation and coalescence on the terraces of substrate becomes increasingly likely and the BCF picture is no longer appropriate. One way of providing a theoretical description of this regime within an analytical framework that complements the BCF theory is with equations of motion for the densities of adatoms and islands. These are called *rate equations*.

Here we consider the simplest rate equation of growth, where adatoms are only mobile surface species and the nucleation and growth of islands proceeds by the irreversible attachments of adatoms, i.e. once an adatom attaches to an island or another adatom, subsequent detachment of that adatom cannot occur. We will signify the density of surface atons by $n_1(t)$ and the density of s-atom islands by $n_S(t)$ where $s>1$. The rate equation for n_1 is

$$\frac{dn_1}{dt} = J - 2D\dagger_1 n_1^2 - Dn_1 \sum_{s=2}^{x} \dagger_s n_s \tag{2.6}$$

In common with most formulations of rate equations, the adatom and island densities are taken to be spatially homogeneous. In particular, there is no diffusion term $D\nabla^2 n_1$, despite the fact that adatoms are mobile. This description is most suitable for singular surfaces, where there are no pre-existing steps to break the translational symmetry of the system and induce a spatial dependence in the adatom and island densities.

The first term on the right-hand side of equation (2.6) is the deposition of atoms onto thr substrate, which increases the adatom density, and so has a positive sign. The next term describes the nucleation of a two-atom island by the irreversible attachment of two migrating adatoms. This term decreases the number of adatoms (by two) and thus has a negative sign. The rate for this process is proportional to the square of the adatom density because two adatoms are required to form a two-atom island, and D, the adatom diffusion constant, because these adatoms are mobile. The third term accounts for the depletion rate of adatoms due to their capture by islands. This term is proportional to the product of the adatom and total island densities and must also have a negative sign. The quantities $\dagger_i$, called capture cross section, accounts for the diffusion flow of atoms into the islands.

The rate equation for the density of s-atom islands $n_S(t)$ is

$$\frac{dn_s}{dt} = Dn_1 \dagger_{s-1} - Dn_1 \dagger_s n_s \tag{2.7}$$

The first term on the right-hand side of this equation is the creation rate of s-atom islands due to the capture of adatoms by (s-1)-atom islands. Similarly, the second term is the depletion rate of s-atom islands caused by their capture of adatoms to become (s+1)-atom islands. There is an equation of this form for every island comprised of two or more atoms, so we have an infinite set of coupled ordinary differential equations. However, since the density of large islands decreases with their size, in practice the hierarchy in (2.7) is truncated to obtain solution for n_1 and the remaining n_S to any required accuracy. Notice that in writing (2.7) we have omitted any direct interactions between islands. This restricts us to a regime where there is no appreciable coalescence of these islands.

To illustrate the calculus of rate equation, we consider a limiting case where all of the capture numbers are set equal to unity. Then, by introducing the total island density, $N = \sum_{S>1} n_S$, using this definition in (2.6), and summing the equation in (2.7) over *s*, we obtain a closed set of two equations for n_1 and N :

$$\begin{aligned} \frac{dn_1}{d\theta} &= 1 - 2Rn_1^2 - Rn_1 N \\ \frac{dN}{d\theta} &= Rn_1^2 \end{aligned} \tag{2.8}$$

where $R = D / J$ and we have used the relation between the coverage and the flux in the absence of desorption, $\theta = Jt$, to replace the time by the coverage as the independent variable. This replacement is made because the coverage is the more “natural” variable, since it can be measured directly with greater accuracy that the deposition time and the flux.

Equations (2.8) are straightforward to integrate numerically and their silution are shown in Fig. 2.6. we will focus here on the initial and longtime behavior of the adatom and islands densities, where analytical solutions are easily obtained.

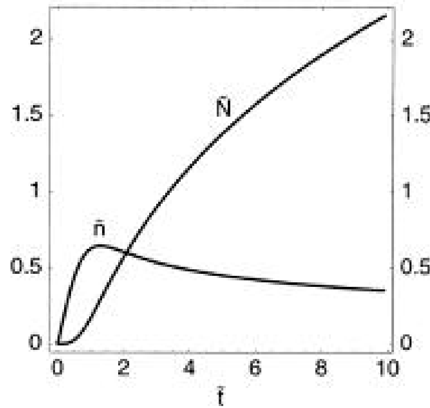


Figure 2.6 The dimensionless adatom and island densities, denoted by $\tilde{n}$ and $\tilde{N}$, as a function of dimensionless time, $\tilde{t}$, obtained by integrating the rare equation (2.8)

At short times ($\theta \ll 1$),

$$n_1 \sim \theta, \quad N \sim R\theta^3 \tag{2.9}$$

The density of adatoms initially shows a linear increase with coverage (or time), which is due entirely to the deposition flux. The islands are somewhat slower in their early development, showing a cubic time–dependence, because the adatom density is too low for there to be appreciable island nucleation. Equations (2.8) show that N continues to increase for all later times, although n_1 increases initially, it eventually begins to decrease (Fig. 2.6). This continues until we reach a regime where $n_1 \ll N$ and $dn_1/d\theta \approx 0$. In this regime, we obtain

$$\begin{aligned} n_1 &\sim \theta^{-1/3} R^{-2/3} \\ N &= \theta^{1/3} R^{-1/3} \end{aligned} \tag{2.10}$$

Notice that, just as in equation (2.5), the ratio D/J is the controlling parameter for quantities which characterize the surface morphology. In particular, the equation for N indicates that increasing the temperature (i.e. decreasing D) and/or decreasing the flux J causes the island density to decrease, resulting in fewer, but larger islands.

2.4 Mechanism of Heteroepitaxial Growth

The morphology that results during the growth of a material on the substrate of a different material is central to the fabrication of all quantum heterostructures. This morphology is determined by the surface and interface energies of the materials, the manner in which strain is accommodated if the materials have different lattice constants, and any effects of alloying and segregation. Controlling the morphology during heteroepitaxy involves understanding the atomistic mechanisms by which these factors assert themselves.

GaAs, AlAs and their alloys are the simplest semiconductor heteroepiraxial systems because of the very small lattice mismatch between AlAs and GaAs and similar values of thermal expansion coefficients. But this situation is not typical. The fabrication of heterostructures from other combination of materials with potentially attractive properties requires identifying (or even utilizing) the morphological and electronic consequences of any lattice mismatch.

There are abundant data available for several heteroepitaxial systems, but there is no theoretical approach which captures the essence of morphological evolution if there is lattice misfit. There are two reasons for this. The kinetics of atomic processes on the surface of strained systems are not determined simply by local environment of the atoms, as in the case of homoepitaxy, but may incorporate non-local information, such as the *height* of a terrace above the initial substrate or the *size* and *shape* of 2D and 3D islands. Then there is the issue of lattice relaxation and any resulting defect formation. The theoretical description of such effects at heterogenous interfaces has relied largely on the minimization of energy functional in order to determine equilibrium atomic positions near the interface as a function of the lattice mismatch.

The Frenkel-Kontorova model has been used to address several general aspects of the accommodation of misfit srtain and the formation of dislocations in heteroepitaxial system within a simply analytic framework.

Frenkel-Kontorova model

In the Frenkel-Kontorova model, the equilibriumpositions of atoms within the growing layer result form the competition between the preferred interatomic separation of these atoms, which interact through harmonic springs, and the periodicity imposed by the rigid potential of the substrate. This potential induces elastic strain in the epilayer and can result in the formation of misfit dislocations.

Many of the characteristic features of strained islands can be captures by the simplest calculation of a one-dimensional monolayer island consisting of N adjacent atoms. The harmonic springs connecting these atoms have a natural length, b the lattice constant of the deposited material, and a force constant k. The interaction between the atoms within the island and the substrate is described by a rigid sinusoidal potential which has periodicity a, the lattice constant of the substrate:

$$V(x) = \frac{1}{2}W[1 - 2\cos(2\pi\, x/a)] \tag{2.11}$$

The ground state of this system is determined by calculation the energy as a function of the atomic positions within the island and then minimizing this expression with respect to these positions. We denote this distances from the origin to the nth and (n+1)th atoms by $X_n = na + x_n$ and $X_{n+1} = (n+1)a + x_{n+1}$, where x_n and x_{n+1} are the displacements of the atoms from the bottoms of their substrate potential troughs (Fig. 2.7).

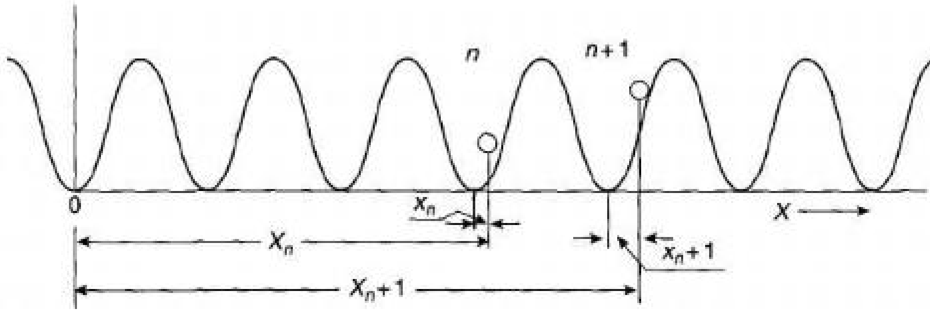


Figure 2.7 Layout of atoms in epitaxial layer with respect of their equilibrium position. Curve shows periodic potential.

The distance ΔX_n between the (n+1)th and nth atoms can be written as

$$\Delta X_n = X_{n+1} - X_n = (n+1)a + x_{n+1} - na - x_n = x_{n+1} - x_n + a \tag{2.12}$$

The strain $\varepsilon(n)$ of the bond between these atoms is

$$\text{v}(n) = \Delta X_n - b = x_{n+1} - x_n + a - b = a\left(\frac{x_{n+1}}{a} - \frac{x_{n+1}}{a} - \frac{b-a}{a}\right) =$$
$$= a\left(<_{n+1} - <_n - f\right), \quad <_{n+1} = \frac{x_{n+1}}{a},\ <_n = \frac{x_n}{a},\ f = \frac{b-a}{a} \qquad (2.13)$$

where f is the misfit between the epilayer and the substrate. The energy of the N-atom island is now written as the sum of the potential energy (2.11) and the strain energy due to the changes in the length of the springs in (2.13):

$$\frac{E}{W} = l_0^{\,2}\sum_{n=1}^{N-1}(<_{n+1} - <_n - f)^2 + \frac{1}{2}\sum_{n=1}^{N}[1-\cos(2f<_n)]\ ,\ l_0 = \left(\frac{ka^2}{2W}\right)^{1/2} \qquad (2.14)$$

l_0 is the ratio of the interaction energy between the atoms in the epilayer to that between the epilayer and the substrate.

The partial derivatives of E potential energy with respect to the variable $<_n$ give the force acting on the nth atom. In the equilibrium condition $\partial E/\partial <_n = 0$. By minimizing (2.14) expression the recurrent equations are obtained:

$$<_1 - <_0 - f = \frac{f}{2l_0^2}\sin 2f<_0$$
$$<_{n+1} - 2<_n - <_{n-1} = \frac{f}{2l_0^2}\sin 2f<_n \qquad (2.15)$$
$$<_{N-1} - <_{N-2} - f = \frac{f}{2l_0^2}\sin 2f<_{N-1}$$

These equations can be solved numerically, However it is possible to solve them analytically if make definite assumptions. The obtained results gives us the possibility to calculate the energy of epitaxial layer.

Figure 2.8 shows the equilibrium configuration for islands of different size calculated with the parameters $f = 0.1$ and $l_0 = 10$, i.e. for atoms in the epilayer that are much strongly bonded to each other than th the substrate and which have a 10 % larger lattice constant than the substrate.

Several general issues can be discusses with reference to this figure

- Atoms near the centre of the island adopt positions close to the minimum of the nearest potential energy trough. Atoms further from the centre of the island are correspondingly further away from their nearest minima. Thus, strain relaxation occurs predominantly at the edges of islands.
- As the number of atoms in an island increases, the strain energy within the island builds up and the energy difference between a coherent island and an island with a single dislocation diminishes until, at some critical size, a dislocation is formed.
- With the parameters we have chosen, an island with 12 or more atoms minimizes its energy by forming a dislocation, which is located in the centre of the island. Since strain relaxation is largest at island edges, we would expect the dislocation to form at the island edge and to migrate toward the centre, with the interface between the epilayer and the substrate providing the one-dimensional slip “plane”

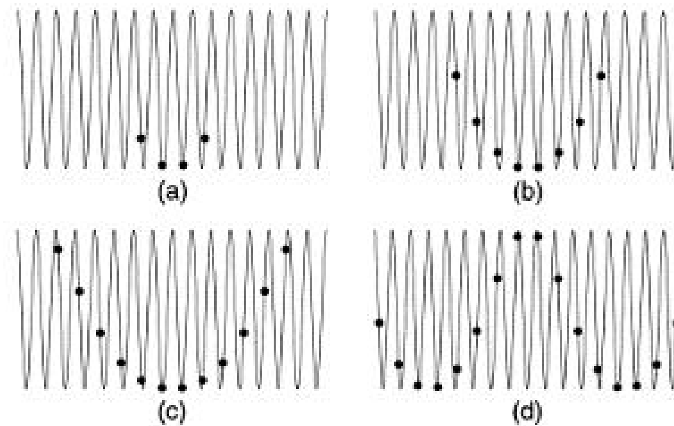


Figure 2.8 Equilibrium position of atoms in a one-dimensional epilayer within the Frenkel-Kontorova model. Islands are shown with (a) 4 atoms, (b) 8 atoms, (c) 12 atoms, and (d) 16 atoms. The islands in (a), (b) and (c) are coherent, but the island in (d) has a dislocation located in its centre.

Figure 2.9 demonstrate formation of dislocation in heterostructure. One can see that number of atomic layers in epilayers is less than that in substrate by one.

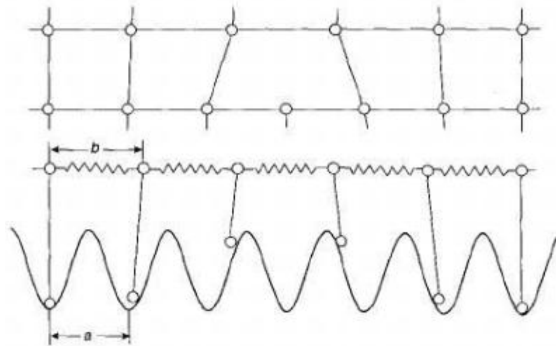


Figure 2.9 Formation of dislocation in heterostructure. Curve shows the periodic potential which minima coincides to the equilibrium position of atoms.

2.5 Characterization Techniques

The advances in nanoscience and nanotechnology are uniquely connected to the development of characterization techniques. The techniques we are discussing here are: reflection high-energy electron diffraction (RHEED), transmission electron microscopy (TEM), scanning electron microscopy (SEM), Scanning Tunneling Microscopy (STM), and atomic force microscopy (AFM).

Reflection High-energy Electron Diffraction

Surface electron diffraction is a standard method for examining the structure of surface both in equilibrium and in the presence of a deposition flux, and thus represents one of the techniques used during epitaxial growth. RHEED measurement is carried out by directing a high-energy (10-20keV) beam of electrons at a glancing angle (0.5-3^0) toward the surface (Fig.2.2). The electrons

penetrate a few layers into the surface and those which emerge are recorded on a phosphorescent screen. RHEED has three main advantages: (i) it is a simple measurement to set up, requiring only an electron gun and a recording screen, (ii) it is geometrically compatible with the molecular beams (during MBE), and so does not interfere with the growth process, and thus, (iii) it can be carried out *in situ* under normal growth conditions.

RHEED provides several type of information about a surface, including its crystallographic symmetry (from the symmetry of the diffraction pattern), the extent of long-range order (from the sharpness of the pattern), and whether growth is proceeding in a 2D or a 3D mode. One of the most common applications of RHEED is based on measuring the intensity of the specular beam (equal incident and reflected angles). A typical example, taken during the growth of GaAs, is shown in Fig. 2.10. Most apparent in this trace are the oscillations and their decaying envelope. The oscillations are due to the repeated formation of bi-atomic Ga-As layers and provided the first direct evidence of layer-by-layer growth in this system. The decaying envelope results from the fact that this layer-by-layer growth is *imperfect,* i. e. subsequent layers began to form before the preceding layers are complete.

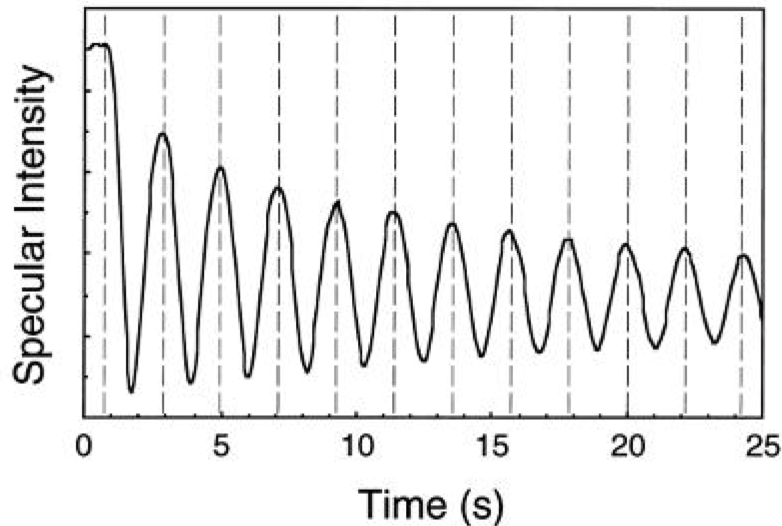


Figure 2.10 Specular RHEED intensity oscillations from a singular GaAs. The broken lines indicate the points where the specular intensity is a local maximum; these correspond to the deposition of additional Ga-As bilayers

The period of the oscillations in Fig. 2.10 indicates that the time required to form a complete bi-layer is of the order of seconds. Since the molecular beams can be turned on and off mechanically with a shutter, the amount of material deposited can be controlled to within a fraction of a layer. Thus, a prescribed number of layers of one material (e.g. GaAs) can be deposited onto a surface, followed by a prescribed number of layers of a second material (e.g. AlAs). This process can be repeated to obtain multi-layer structure. The electronic properties of such structures can be engineered by varying the number of deposited layers of different materials.

Transmission Electron Microscopy

Transmission electron microscopy is analogous to conventional light microscopy in that one obtains pictures or micrographs of a specimen of interest. However, instead of photons, electrons are used to construct the images.

In the TEM experiment, a thin or diluted sample is bombarded under high vacuum with a focused electron beam. Electromagnetic lenses steer the beam and focus it onto the specimen. Transmitted electrons then form contrast patterns that created images of the sample. In this regard, thicker regions of the specimen occlude more of the incident beam than thinner regions, causing intensity variations that help define the image. It should be mentioned, that absorption is generally small in these experiments, since samples are inherently thin. As a consequence, most micrographs are actually constricted from scattered electrons that arise from interactions with the material. This scattering can be either elastic or inelastic and, in many cases, it is elastically scattered electrons that are used to construct the micrographs. The transmitted electron beam image is then magnified onto a detector or phosphorescent screen to yield a picture of the specimen. Fig.2.11 shows TEM image of silver nanoparticles.

A key feature of the TEM is its inherently high degree of spatial resolution. Specifically, the classical diffraction limit of electromagnetic waves is roughly half their wavelength. As a consequence, a beam of electrons resolves much finer things than photons.

To illustrate, assume that the TEM has an accelerating voltage V .

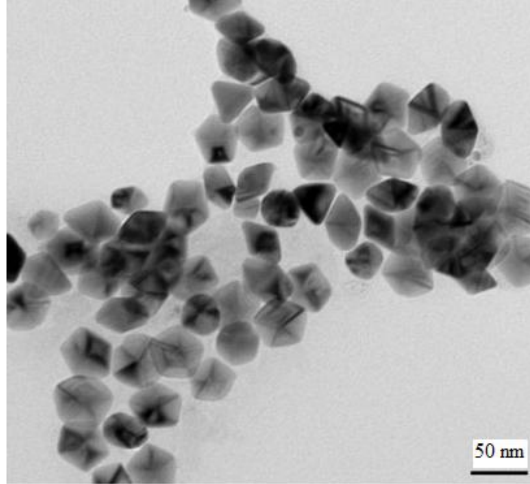


Figure 2.11 TEM image of silver nanoparticles randomly assembled on a grid.

Electrons extracted from the filament or other sources then acquire a kinetic energy $E = eV$. Therefore for momentum and de Broglie wavelength we obtain:

$$p = \sqrt{2m_0 eV}, \qquad \lambda = \frac{h}{\sqrt{2m_0 eV}} \tag{2.16}$$

where m_0 is electron mass, e - its charge, h - Planck constant. The expression for the de Broglie wavelength can be written a more convenient way substituting the values of constants:

$$\lambda\,(\text{nm}) = \frac{1.225(\text{nm})}{\sqrt{V\,(\text{volt})}} \tag{2.17}$$

Next, the diffraction limit is defined as

$$l = \frac{0.612\lambda}{NA} \tag{2.18}$$

which is the length below which we cannot distinguish two objects next to each other. The relation is called Abbé equation. In it, $NA = n_m \sin\theta$ is the numerical aperture of the imaging lens, n_m is the refractive index of the medium, and θ is the half-angle of the impinging rays from the normal incidence. Substituting (2.17) in (2.18) we obtain

$$l(\text{nm}) = \frac{0.612(1.225\,\text{nm})}{\sqrt{V(\text{volt})}\,n_m \sin\theta} \ . \tag{2.19}$$

Since TEM experiments are conducted in vacuum, $n_m = 1$; θ is generally small, therefore $\sin\theta \simeq \theta$. A critical length scale is therefore

$$l(\text{nm}) = \frac{0.750}{\sqrt{V}\,\theta} \ . \tag{2.20}$$

For a 100 keV system, the theoretical diffraction limited resolution is $l \simeq 0.237 nm$, assuming $\theta \simeq 0.01\,\text{rad}$. This value is comparable to the interatomic spacing.

Scanning Tunneling Microscopy

The scanning tunneling microscope, invented in 1982 by Gerd Binning and Heinrich Rohrer, uses an atomically sharp tip placed sufficiently close (a few angstroms) to a surface to produce an electron tunneling current between the tip and the surface. By measuring this current as the tip scans the surface, images of the surface are obtained which, under favorable circumstances, have a lateral resolution of $\sim 1\mathring{A}$ and vertical resolution of $\sim 0.1\mathring{A}$.

The basic principle of the STM can be understood with a model introduced by Tersoff and Hamann in 1983. The tip is represented by a spherical potential well within which the Schrödinger equation is solved. By retaining only the spherical symmetric solutions, a simple expression is obtained for the tunneling current I at low bias voltage of V : $I \sim eV\rho\left(\vec{r}, E_F\right)$,

where $\ldots\left(\vec{r}, E_F\right)$ is the local density of states at the Fermi energy, E_F, of the surface at the position $\vec{r}$ of the tip. Thus, scans taken at constant current measure contours of constant Fermi-level charge density at the surface. It should be mentioned that STM is sensitive to *charge density*, rather than simply *atomic positions*.

STM is very important technique, and its impact on the development of epitaxial growth and investigation of its fundamentals is huge. However, utilization of this technique during the growth processes faces some technical problems. If an STM is placed in a conventional growth chamber, the tip shadows the incoming molecular beam. Thus, the imaging of growing surfaces has had to rely on one of two indirect strategies. The most common is to image a surface that has been quenched after a prescribed period of growth, thereby providing a "snapshot" of the surface. It has become possible also to arrange scan and growth rates within specially designed growth chambers to image the same region of a surface during growth Though technically more demanding, this approach is the more desirable in principle because particular kinetic processes can be tracked and no quenching is required, thus providing a more faithful record of surface evolution. But, because of the very slow growth rates used in current implementations of this "in vivo" method, the growing surface is closer to equilibrium than for more typical growth rates and, moreover, is exposed for relatively long times to the ambient impurities which are always present in any growth chamber. These factors can affect the growth in several ways, so care must be taken when interpreting such images to ensure that they reflect the intrinsic growth characteristics of the material.

STM images of the surfaces of Si and GaAs are shown in Fig. 2.12. These images reveal an important feature that is typical of semiconductor surfaces (and surfaces of many other materials). The creation of a surface produces broken, or dangling, bonds which leave the surface in an unstable high-energy state. The formation of new bonds to lower the surface free energy results in displacements of surface atoms from their bulk-terminated positions.

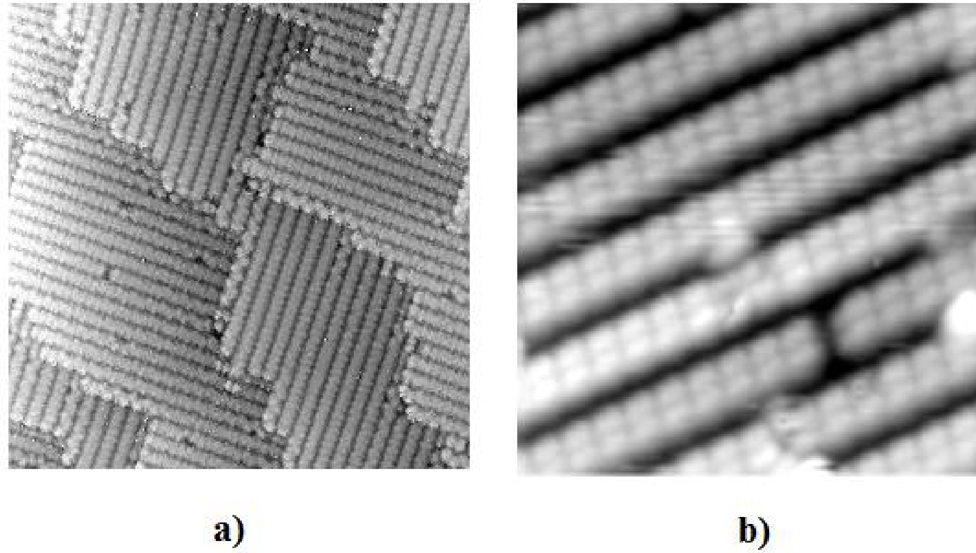


Figure 2.12 a) STM images of Si; b) STM images of GaAs

Atomic Force Microscopy

When the tip of an STM is brought close to a surface, the atoms near the apex of the tip exert a force on that surface which is of the same order of magnitude as the interatomic forces within the surface. This effect is the principle behind the atomic force microscope. An STM tip, mounted on a flexible beam, is brought just above a surface. The force between the surface and the tip causes a small deflection of the beam. The surface is then scanned while a constant force is maintained between the tip and the surface with a feedback loop similar to that used in the operation of an STM.

The AFM complements the STM in several ways. Because the STM relies on a tunneling current for its operation, it is sensitive mainly to the density of electronic states near the Fermi level of the sample. Thus, this density of states must be non-zero, i.e. the sample must be conducting. However, since the AFM tip responds to interatomic forces, which is a cumulative effect of all electrons, the sample need not be a conductor. Additionally, since the tunneling current decreases exponentially with the tip-sample distance, an STM tip must be placed within a few angstroms of the surface to maximize the resolution of the image.

The AFM most commonly operates in this mode (the “contact” mode) as well, but it can also operate at much larger distances from the surface (50—150 A) for samples susceptible to damage or alteration by being in close proximity to the tip (the “non-contact” mode). But, even in the contact mode, attaining atomic resolution is much more demanding technically than with the STM. Thus, many applications of the AFM involve scanning large areas (up to microns) to image the gross morphology of the sample. This has the advantage of not requiring a UHV environment and AFMs often operate in ambient atmosphere or in a liquid.

Chapter 3 Electrons in low Dimensional Structures

Quantum effects arise in systems which confine electrons to regions comparable to their de Broglie wavelength. When such confinement occurs in one dimension only (say, by a restriction on the motion of the electron in the z-direction), with free motion in the x- and y-directions, a “two-dimensional electron gas” (2DEG) is created. Confinement in two directions (y and z, say), with free motion in the x-direction, gives a “one-dimensional electron gas” (1DEG) and confinement of its x-, y-, and z-motions at once gives a “zero-dimensional electron gas” (ODEG). In this section, we consider the description of ideal electron gases in these cases, i.e. electron gases in which there is no motion in the confining direction and where we neglect interactions between the electrons. We will then use these results in the following section to characterize the density of states in real low-dimensional structures, in which there is some degree of lateral mobility.

3.1 Free Electrons in Three Dimensions

An unconfined electron in free space is described by the Schrodinger equation

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)\Psi = E\Psi \qquad (3.1)$$

where m is the free-electron mass. The solutions of this equation,

$$\Psi_k(\vec{r}) = \frac{1}{(2\pi)^3} e^{i\vec{k}\vec{r}} \qquad (3.2)$$

are plane waves labeled by the wave vector

$$\vec{k}\left(k_x, k_y, k_z\right)$$

and correspond to the energy

$$E = \frac{\hbar^2 k^2}{2m} = \frac{\hbar^2 (k_x{}^2 + k_y{}^2 + k_z{}^2)}{2m} \tag{3.3}$$

The vector components of $\vec{k}$ are the quantum numbers for the free motion of the electron, one for each of the classical degrees of freedom.

The number of states in a volume $d\vec{k} = dk_x dk_y dk_z$ of $\vec{k}$-space is

$$g(\vec{k})d\vec{k} = \frac{2}{(2\pi)^3} d\vec{k} \tag{3.4}$$

with the factor of 2 accounting for the spin-degeneracy of the electrons. To express this density of states in terms of energy states, we use the fact that the energy dispersion (3.3) depends only on the magnitude of $\vec{k}$. Thus, by using spherical polar coordinates in $\vec{k}$-space, $d\vec{k} = k^2\,\mathrm{dk}\sin\theta\, d\theta\, d\phi$ and, integrating over the polar and azimuthal angles, we are left with an expression that depends only on the magnitude k:

$$g(\mathrm{k})\,\mathrm{dk} = \frac{1}{\pi^2} k^2 dk \tag{3.5}$$

By invoking (3.3), we can perform a change of variables to cast the right-hand side of this equation into a form involving the differential of the energy:

$$\frac{1}{\pi^2} k^2 dk = \frac{1}{\pi^2}\left(\frac{2mE}{\hbar^2}\right)\frac{dk}{dE} dE = \frac{1}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2} \sqrt{E} dE \tag{3.6}$$

From this equation, we deduce the well-known density of states $g\left(E\right)$ of a free- electron gas in three dimensions:

$$g(\mathrm{E}) = \frac{1}{2f^{2}}\left(\frac{2m}{\hbar^{2}}\right)^{3/2}\sqrt{E} \tag{3.7}$$

Notice the characteristic square-root dependence on the energy.

3.2 Ideal Two-dimensional Electron Gas

An ideal 2DEG differs from free electrons in three dimensions in that the electrons have unrestricted movement in only two dimensions (x and y) with complete confinement in the z-direction, i.e. there is no freedom of movement at all in this direction. The energy of an electron in a 2DEG is therefore

$$E = \frac{\hbar^{2}k^{2}}{2m} = \frac{\hbar^{2}(k_{x}^{\,2} + k_{y}^{\,2})}{2m} \tag{3.8}$$

The number of states within an area in k-space box of $d\vec{k} = dk_{x}dk_{y}$ is

$$g(\vec{k})d\vec{k} = \frac{2}{(2f)^{2}}d\vec{k} \tag{3.9}$$

with the factor of 2 again inserted to account for the spin degeneracy of the electrons. We proceed as above, but now use circular polar coordinates to obtain

$$g(\mathrm{k})\,\mathrm{dk} = \frac{1}{f}kdk \tag{3.10}$$

where $k=\sqrt{k_x^{\,2}+k_y^{\,2}}$. We again use the relationship between energy and wave vector (3.8) to express the density of states in terms of the energy:

$$\frac{1}{f}kdk=\frac{1}{f}\left(\frac{2mE}{\hbar^2}\right)^{1/2}\frac{dk}{dE}dE=\frac{m}{f^{\,2}\hbar^2}dE \tag{3.11}$$

The density of states $g\left(E\right)$ of a 2DEG is therefore given by

$$g(\mathrm{E})=\frac{m}{f\,\hbar^2} \tag{3.12}$$

Thus, for a 2DEG the density of states is a constant, i.e. independent of the energy. This is one of the fundamental features of electrons in planar hetero- structures which make such structures useful for applications.

3.3 Ideal Zero- and One-dimensional Electron Gases

When an electron is allowed only one-dimensional motion (along, say, the x-direction), the energy is given by

$$E=\frac{\hbar^2k^2}{2m}=\frac{\hbar^2k_x^{\,2}}{2m} \tag{3.13}$$

A procedure analogous to that used in the preceding two sections then yields for the density of states the expression

$$g(\mathrm{E})=\frac{1}{f}\left(\frac{2m}{\hbar^2}\right)^{1/2}\frac{1}{\sqrt{E}} \tag{3.14}$$

This shows that the density of states of a one-dimensional electron gas (1DEG) has a square-root singularity at the origin.

An ideal zero-dimensional electron is one that exists in a single state of fixed energy E_0. The density of states is then given by

$$g(\mathrm{E}) = \mathrm{u}\left(E - E_0\right) \tag{3.15}$$

Figure 3.1 shows energy dependence of density of states in 1, 2 and 3 dimensional cases.

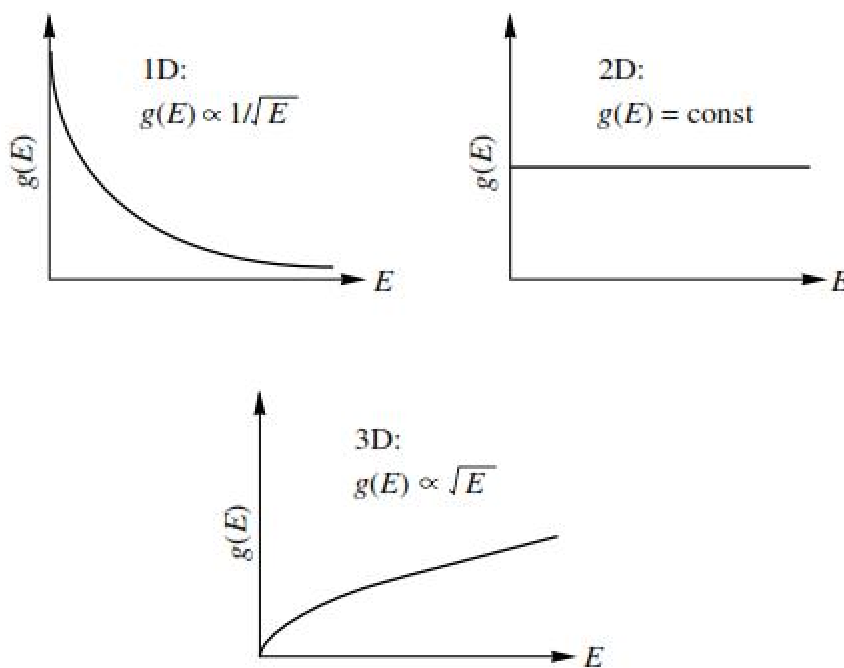


Figure3.1. Density of states for an ideal electron gas in 1, 2 and 3 dimensions.

In real low dimensional crystals i) electrons interact with ions in lattice sites. This is accounted in the frame of *effective mass approximation* described in Chapter 1; ii) the dimension in which the motion is restricted in reality is finite. This point is discussed in next sections.

3.4 Real Electron Gases: Single Particle Models

By a quantum well we mean any structure in which an electron (or hole) is strongly confined in one dimension. A practical example of great importance is obtained when a plane layer of GaAs lies within a sample of bulk Al_xGa_{l-x} As. These materials may be grown, e.g. by molecular-beam epitaxy (MBE) in a layer- by-layer fashion to form such a structure. The materials are lattice-matched (the same lattice structure and very similar lattice spacing). Moreover, their band structures are qualitatively similar if the aluminum proportion x is less than approximately 0.4. However, the band gap of Al_xGa_{l-x} increases linearly with increasing x. What results (when $x \neq 0$) is a discontinuity in the conduction and valence band edges, E_c and E_v, as shown in Fig. 3.2. The precise proportion of the discontinuity taken up by the conduction band alone must be known beforehand, from experiment, or else from theory. Quantum confinement of an electron within the thin layer of GaAs will happen if its energy is below that of the conduction-band edge in the AlGaAs. This is an example of a compositional quantum well.

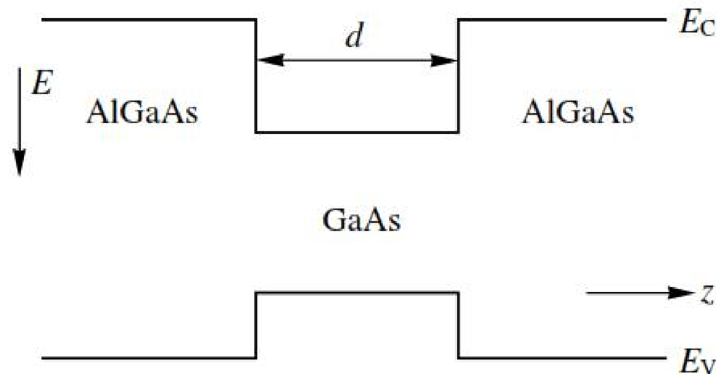


Figure 3.2. Band-edge diagram for a typical AlGaAs/GaAs quantum well. The fraction x, of Al is less than 0.35.

The envelope function obeys a Schrodinger-like equation which, in the simplest materials, such as GaAs, and near k = 0, takes the form

$$\left[-\frac{\hbar^2}{2m^*}\nabla^2 + V\left(\vec{r}\right)\right]\psi\left(\vec{r}\right) = E\psi\left(\vec{r}\right) \tag{3.16}$$

where E is measured from the conduction-band edge and V does not include the crystal potential. The entire effect of the crystal potential is to change the electron mass from m to m^*, the effective mass. (In GaAs, for instance, m^* = 0.067 m.) The potential V in equation (3.16) contains the effect of all external potentials, and in particular, that due to changes in the conduction band edge.

Ideal Square Well

Most simply, the square-well potential produced in a compositional quantum well can be approximated by that of an infinite square well, i.e. that in which the potential is constant within the well and infinite outside the well (Figure 3.3 a))

$$V\left(z\right) = \begin{cases} 0 & for\ 0 \le z \le d \\ \infty & otherwize \end{cases} \tag{3.17}$$

Since motion is unrestricted in the y- and z-directions, the Schrodinger equation (3.16) is separable in rectangular coordinates, so the coordinate dependence of the wave function in the (x, y) plane can be separated from that in the z-direction. This results in plane-wave solutions for the motion of the electron in the x- and y-directions,

$$\Psi(\mathrm{x,y,z}) = \mathrm{e}^{ik_x x}\mathrm{e}^{ik_y y}\,\psi(\mathrm{z}) \tag{3.18}$$

where { (z) obeys the one-dimensional Schrodinger equation for a particle in an infinite square well

$$\left[-\frac{\hbar^2}{2m^*}\frac{d^2}{dx^2}+V(z)\right]\{ (z)=E\{ (z) \tag{3.19}$$

where $V(z)$ is given by (3.17).

As is well known, the general solution of (3.19) must be a linear combination of sines and cosines chosen to satisfy the boundary conditions imposed by the well (Fig. 3.3 b)). Since $\{ (z)$ must vanish at $z=0$, the solutions must be of the form $\sin(kz)$ and, since it must also vanish at $z=d$, we must choose $k=nf\ /d$, $n=1,2,3....$. This restriction results in the quantization of the energy. The allowed energies associated with the motion of the electron along the z-direction are

$$E_n=\frac{\hbar^2 f^{\,2} n^2}{2m^* d^2} \tag{3.20}$$

The total energy of the electron is the sum of this quantized energy and the kinetic energy due to its (x, y)-motion:

$$E=\frac{\hbar^2 (k_x{}^2+k_y{}^2)}{2m^*}+\frac{\hbar^2 f^{\,2} n^2}{2m^* d^2} \tag{3.21}$$

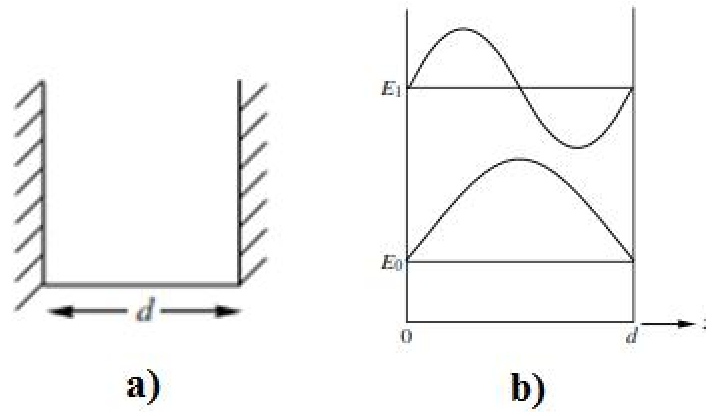


Figure 3.3. a) Infinite square-well potential; b) Energies and wave functions for the first two quantized states in a square-well potential.

Already, differences from the ideal 2DEG case are evident: (i) there can be several different quantized energies E_n, i.e. several possible states of z-motion, and (ii) the electron wavefunctions have a finite spread in the z-direction.

The $E-\vec{k}$ dispersion relation for an infinite quantum well is thus a generalization of simple parabolic form shown in Fig. 3.4 and corresponding to ideal 2D electron systems.

One obtains instead the situation shown in Fig. 3.5 , where from equation (3.20) for

$$n=1 \quad E_0 = \frac{\hbar^2 f^2}{2m^* d^2}$$

For energies $E < E_0$ (A in Fig. 3.5), there are no states; for energies B ($E_0 < E < 4E_0$) the density of states (per unit area) is just that for a perfect two-dimensional electron gas, namely $g_0(\mathrm{E}) = \mathrm{m}^*/f\hbar^2$. For energies C the density of states (DOS) is $2g_0$; energies D - $9E_0 < E < 16E_0$ have $3g_0$ for the DOS, and so forth.

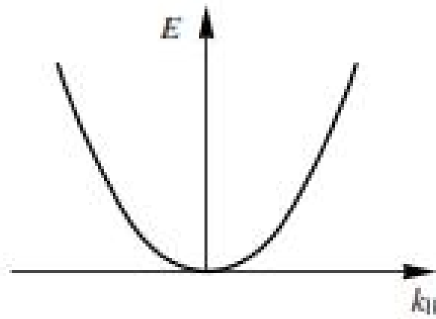


Figure 3.4. Parabolic $E-\vec{k}$ relation for ideal 2D system

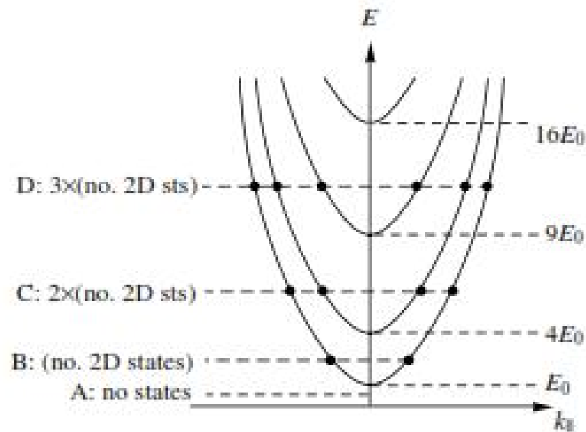


Figure 3.5. Energy versus wave number k for an infinite quantum well

To convert this $g(\mathrm{E})$, which is the density of states per unit area of real space, to a density of states per unit volume, one must divide by an appropriate length in the z-direction, in this case by the well width d. This three-dimensional DOS then rises in steps of $2\,\mathrm{m}^*/f\,\hbar^2$, as shown in Fig. 3.6, where it is also compared with the ordinary bulk DOS. If states in the well are filled up to some Fermi energy E_F, then states at the Fermi level (the ones of most interest for transport) will have different kinetic energies of motion in the x-y plane, and therefore different

Fermi velocities, depending upon their quantum state in the well. Thus, for instance, if E_F lies above some level E_n, then the Fermi wavenumber for states in level n is given by

$$E_F - E_n = \frac{\hbar^2 k_F^2}{2m^*} \tag{3.22}$$

with the corresponding Fermi velocity

$$v_F = \frac{\hbar k_F}{m^*} \tag{3.23}$$

related to k_F in the usual way.

The density of states can be written as

$$\begin{aligned} g(E) &= \frac{m}{f\hbar^2}\sum_n {}_{''}(\mathrm{E}-\mathrm{E}_n), \\ {}_{''}(\mathrm{E}-\mathrm{E}_n) &= \begin{cases} 1, E > E_n \\ 0, E < E_n \end{cases} \end{aligned} \tag{3.24}$$

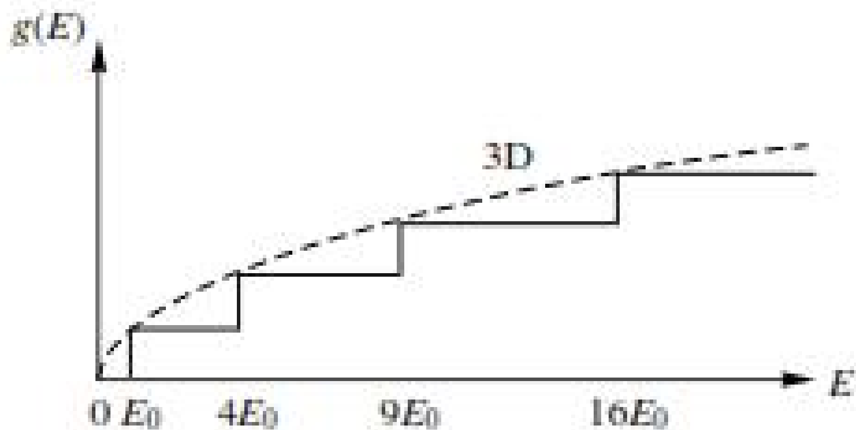


Figure 3.6. Density of states for an infinite square well. The corresponding density of states for an unconfined 3D system is also shown for comparison (broken line).

Figure 3.7 also shows dispersion relation in quantum well.

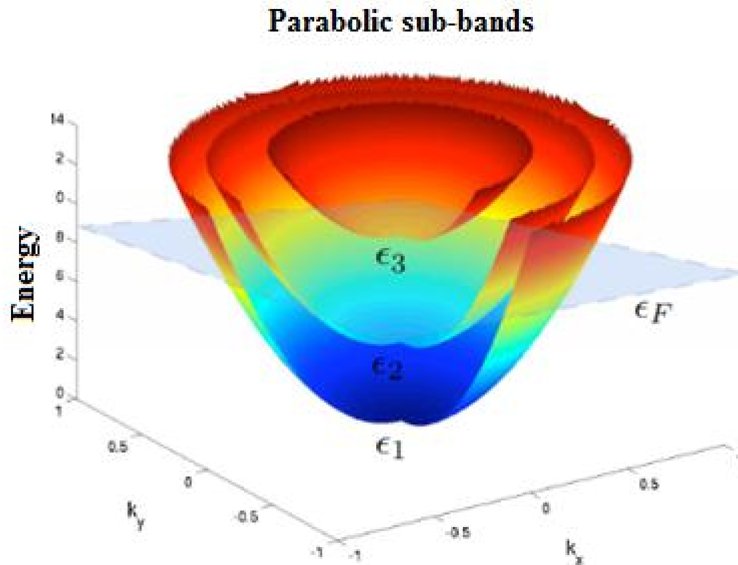


Figure 3.7 $E-\vec{k}$ relation in quantum well

Holes in Quantum Wells

Holes as well as electrons can be confined strongly in one or more dimensions. In a GaAs quantum well in the GaAs/AlGaAs system (Fig. 3.2), there is a quantum well for holes wherever there is a well for electrons. In other systems, there may be a well just for electrons. In the first case we have type-I quantum structure, in the second one - type-II quantum structures (Figure 3.8)

There is a three-fold degeneracy in the hole bands at the -point ($k=0$), which is the highest point in the valence bands. In the bulk one deals with three sorts of holes. Spin-orbit splitting depresses one of these bands to create the spin-split-off band, which is then often ignored. The other two bands correspond to heavy and light holes (so called because of their greater or

smaller effective mass); these states are degenerate in energy at the -point. In the bulk one can treat the two (or three) sorts of holes separately (Figure 3.9).

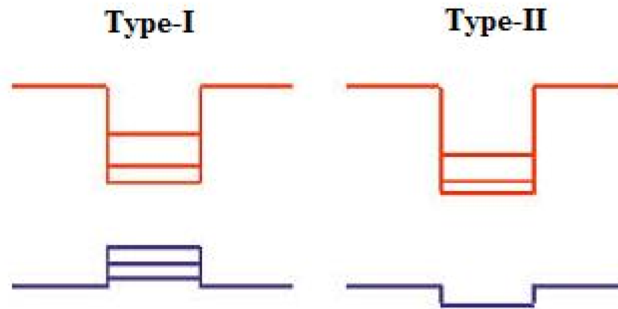


Figure 3.8 type-I and type-II quantum structures

In quantum wells and other strongly confined systems, however, the confinement breaks the symmetry which caused the degeneracy in the first place. The hole states then mix and, in general, they will mix differently in different directions.

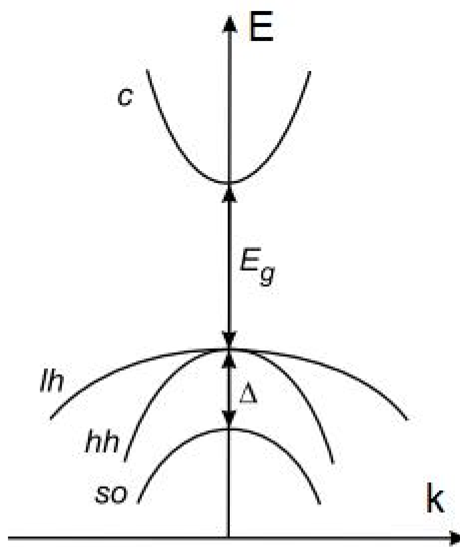


Figure 3.9 Dispersion relation for bulk crystal near $k = 0$ point

Non-parabolicity

For electrons with energies near the bottom of the conduction band, the $E(k)$ band structure is parabolic: $E = \hbar^2 k^2 / 2\,\mathrm{m}^*$. For somewhat higher energies, this relation is no longer true (the importance of non-parabolicity, at a given energy, will depend on the material in question). However, one can still define an energy-dependent effective mass $m(E)$ by

$$E(k) = \frac{\hbar^2 k^2}{2\,\mathrm{m}^*(\mathrm{E})} \tag{3.25}$$

We can suppose that $m(E)$ is known in the bulk. Then, in an infinite square well, the quantized energies E_n will be given by

$$E_n\,\mathrm{m}(E_n) = \frac{1}{2}\hbar^2 k^2,\ k = \frac{nf}{d} \tag{3.26}$$

Again the full energy E E for an electron in state n is given by

$$E = \frac{\hbar^2 k_{\parallel}{}^2}{2m(\mathrm{E})} + E_n \tag{3.27}$$

The situation is illustrated schematically in Fig. 3.10.

Non-parabolicity will modify the quantized energies E_n' from their original value, though only negligibly for low energies near the conduction-band edge (i.e., for low n). Moreover, for each level the $E-k$ relation will be parabolic near $k=0$, but the curvature of the parabolas will become broader as n becomes higher (non-parabolicity is known to increase $m(E)$ as E becomes higher). And finally, the $E-k$ relation for each level will itself become non-parabolic as k becomes large (Figure 3.11).

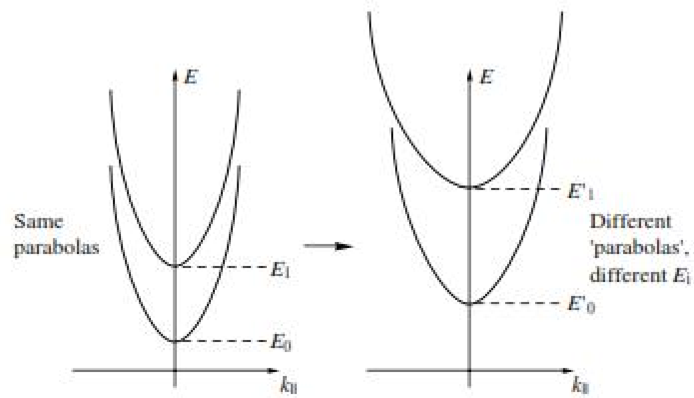


Figure 3.10 Parabolic and non-parabolic $E-k$ relations for an ideal square well

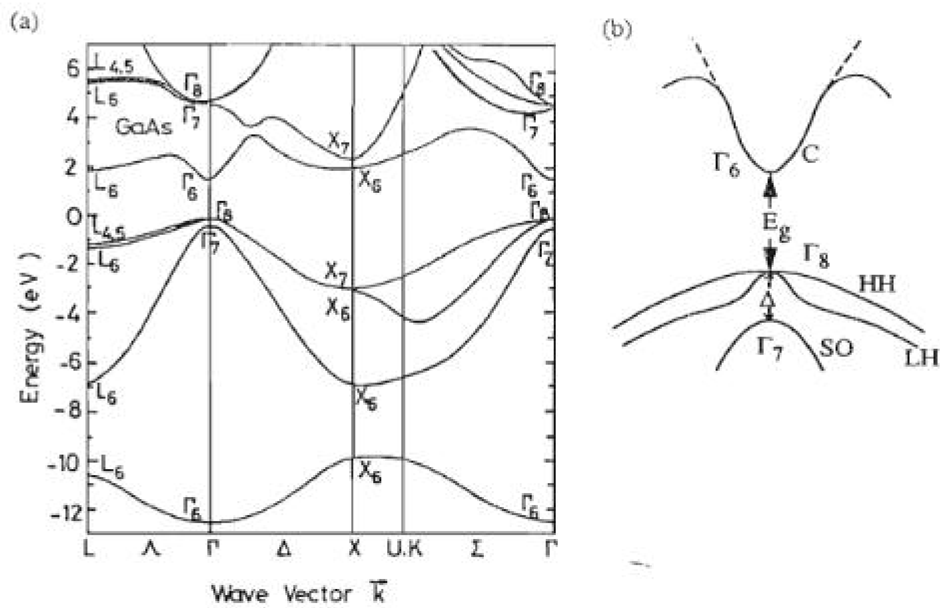


Figure 3.11 $E-k$ relations at whole range of wave vector

3.5 Finite Quantum Wells and Real Systems

One of the most crucial differences between this case and that of an infinite square well is that the electron wavefunction need not be zero in the barrier region (Figure 3.12). This fact that electrons can penetrate into the barrier region will be particularly important when it comes to a discussion of the physics of superlattices (see below)

An electron in a finite square well is confined by two potential steps. These could be the finite conduction-band discontinuities, for instance, in Fig. 3.2. Usually one solves for the wavefunction { at a potential step using the following assumptions: (i) { must behave suitably at infinity (usually, decaying to zero), (ii) { must be continuous at the interface (say, at the potential step), (iii) the first derivative of { must also be continuous at the interface (Fig.3.13). These conditions can be stated as

$$\{_1(0)=\{_2(0), \qquad \left.\frac{d\{_1}{dz}\right|_{z=0}=\left.\frac{d\{_2}{dz}\right|_{z=0} \tag{3.28}$$

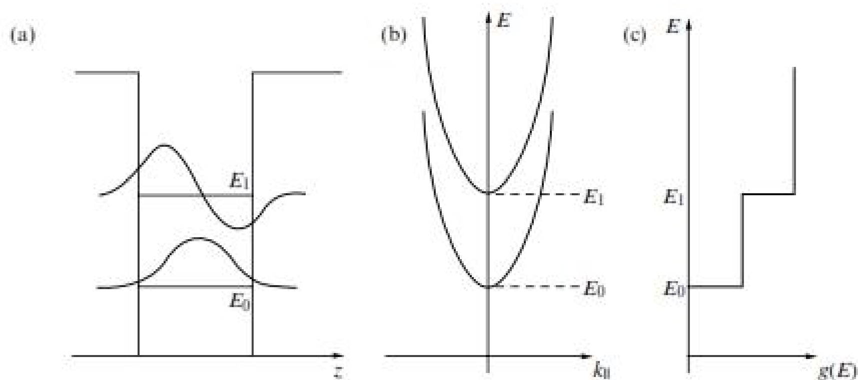


Figure 3.12 Schematic illustration of (a) wave functions, (b) energies and (c) density of states for a particle confined to a finite square well.

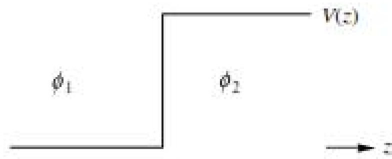


Figure 3.13 The wavefunction at an ordinary potential step.

A semiconductor interface, however, is more subtle than an ideal potential step for various reasons. In spite of this fact, the first assumption one might make is that the envelope function { for the electron responds to a conduction band offset in much the same way as the complete wavefunction to an ideal potential step. This begs a number of questions, some of which we shall mention here. What one really wants to know is the correct matching conditions for the envelope function at a material interface (Fig. 3.14). Although there are many such questions still to be answered, it is the case that a simple effective-mass approach is surprisingly good.)

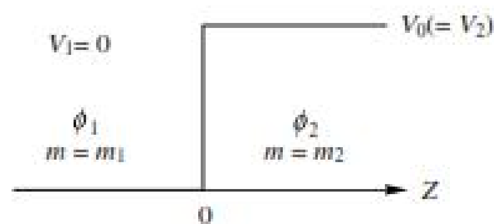


Figure 3.14. Boundary conditions near a material interface. Electrons in the material on the left (right)-hand side have effective mass m_1 (m_2).

Various prescriptions have been proposed for the matching conditions of a wavefunction at a semiconductor interface. Which is correct is still a matter of debate, and it may be the case that there is not a unique answer. We present here the matching conditions in common use, which are known as the Bastard conditions.

At an interface, the effective mass and the conduction band edge potential are effectively discontinuous. Using the notation of Fig. 3.14, we have approximately that

$$m^* = m(z) = m_1 + (m_2 - m_1)\Theta(z)$$

$$V = V(z) = V_0\Theta(z) \tag{3.29}$$

$$\Theta(z) = \begin{cases} 0, & z < 0 \\ 1, & z > 0 \end{cases}$$

The usual effective-mass Hamiltonian

$$H = -\frac{\hbar^2}{2m^*}\nabla^2 + V \tag{3.30}$$

is not Hermitian for such a z-dependent effective mass, but must be made so for quantum-mechanical consistency. This can be done by using instead the Hamiltonian

$$H = -\frac{\hbar^2}{2}\nabla\frac{1}{m(z)}\nabla + V \tag{3.31}$$

We assume that the envelope function $\psi(z)$ is continuous at $z = 0$. We next suppose that, in fact, $m(z)$ and $V(z)$ change very rapidly but continuously over a small distance $\Delta z = 2\varepsilon$ (Fig. 3.15).

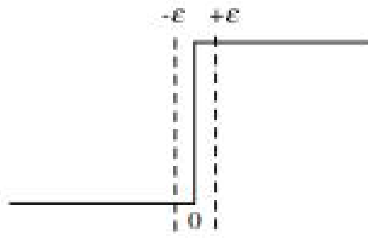


Figure 3.15. The quantities $m(z)$ and $V(z)$ show smooth but rapid variation over a small distance $\pm\varepsilon$ on either side of the interface, at $z = 0$.

We now take the wavefunction $\Psi(x,y,z) = e^{ik_x x}e^{ik_y y}\psi(z)$ and then integrate the effective mass equation $\hat{H}\psi = E\psi$ from $z = -\varepsilon$ to $z = +\varepsilon$:

$$\int_{-\varepsilon}^{+\varepsilon}\left\{\frac{d}{dz}\left[\frac{1}{m(z)}\frac{d\psi}{dz}\right]\right\}dz + V_0\psi(+\varepsilon)\varepsilon = E\left[\psi(+\varepsilon) - \psi(-\varepsilon)\right]\varepsilon \qquad (3.32)$$

We now let $\varepsilon \to 0$. Since V_0, ψ and E are all finite, the two terms multiplied by ε both approach zero as $\varepsilon \to 0$. Thus, the first term on the left in (3.328) must also approach zero, giving

$$\left.\frac{1}{m_1}\frac{d\psi_1}{dz}\right|_{z=0} = \left.\frac{1}{m_2}\frac{d\psi_2}{dz}\right|_{z=0} \qquad (3.33)$$

This suggests that the proper boundary conditions, at the (plane) interface ($z = 0$) between materials 1 and 2, are

$$\psi_1(0) = \psi_2(0), \qquad \left.\frac{1}{m_1}\frac{d\psi_1}{dz}\right|_{z=0} = \left.\frac{1}{m_2}\frac{d\psi_2}{dz}\right|_{z=0} \qquad (3.34)$$

3.6 Quantum Wires

Above we have considered confinement of electrons to two dimensions. To create systems of yet smaller dimension, it is usual to start with a well-confined two-dimensional electron gas, so tightly confined that electrons are present in only a single energy level. A new confinement is then imposed on this system, in a direction perpendicular to the original confinement (Fig. 3.16). A direct way to do this lateral confinement is by cutting the material containing the 2DEG, for instance by etching, to remove all but a thin strip of the 2DEG (a typical width L

might be approximately 1000 A. The electrons, now confined in two dimensions but free to move in the third, form a quantum wire.

The simplest theoretical picture of such a quantum wire is given by the confinement of an ideal two-dimensional electron gas in an infinite square well. Figure 3.17 illustrates the wavefunctions of the first two quantum states of such a quantum wire. The original 2DEG lies in the plane z = 0, with the additional confinement in the x-direction. The energies of x-confinement are

$$E_{n_x} = \frac{\hbar^2 f^2 n_x^2}{2m^* L^2} \tag{3.35}$$

and the total energy is

$$E_{n_x n_z k_y} = \frac{\hbar^2 f^2 n_x^2}{2m^* L^2} + \frac{\hbar^2 k_y{}^2}{2m^*} \tag{3.36}$$

In this formula first terms corresponds to confined motion, the last one to unrestricted motion in y-direction.

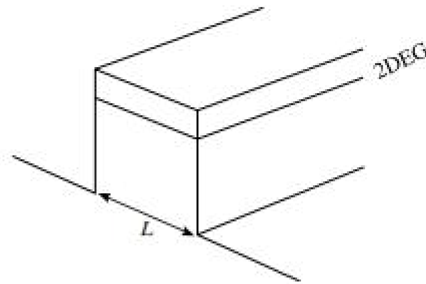


Figure 3.16 A two-dimensional electron gas as the basic ingredient for forming a quantum wire.

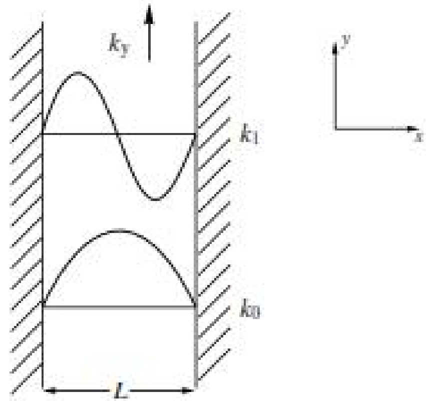


Figure 3.17 Wavefunctions for an electron in an ideal quantum wire. An ideal 2DEG in the plane z = 0 has undergone additional confinement by infinite potential steps in the x-direction

Figure 3.18 shows the density of states $g(E)$ for such an ideal quantum wire. $g(E)$ shows the characteristic singularity in $E^{-1/2}$ which was derived for a 1DEG in equation (3.14). In a quantum wire, such a singularity will occur at each energy E_n of quantization in the x-direction. For real quantum wires, the spacing of the quantized energies E_n, and the corresponding wavefunctions, will depend on the precise shape of the potential $V(x, y)$, just as they depended, for a 2DEG, on the shape of the potential $V(z)$.

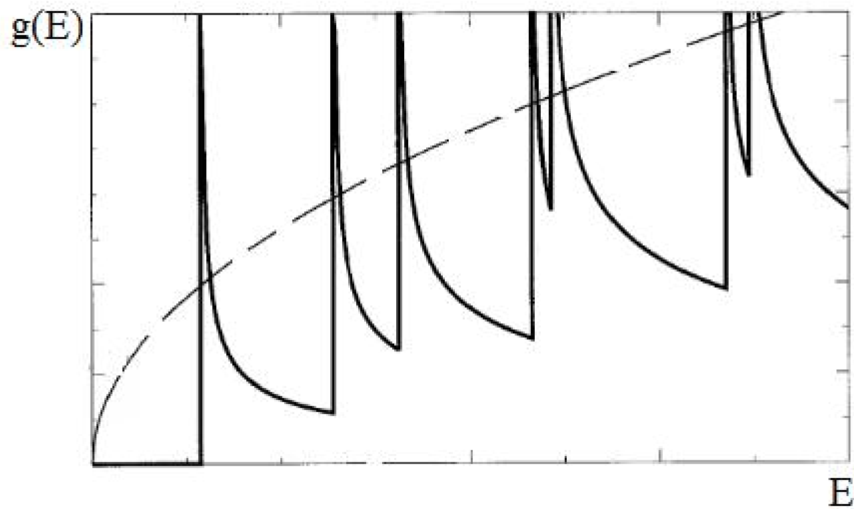


Figure 3.18. Density of states for an ideal one-dimensional quantum wire.

This density of states can be written as

$$g(E) = \sum_{n_x n_y} \frac{(2m)^{1/2}(\mathrm{E}-\mathrm{E}_{n_x n_y})^{-1/2}}{f\hbar}\,〃(\mathrm{E}-\mathrm{E}_{n_x n_y}),$$

$$〃(\mathrm{E}-\mathrm{E}_{n_x n_y}) = \begin{cases} 1, E > \mathrm{E}_{n_x n_y} \\ 0, E < \mathrm{E}_{n_x n_y} \end{cases}$$

3.7 Quantum Dots

Electrons can be confined in all three dimensions in a “dot” or “quantum box”. The situation is analogous to that of a hydrogen atom: only discrete energy levels are possible for electrons trapped by such a zero-dimensional potential. The spacing of these levels depends, as always, on the precise shape of the potential. The development and application of quantum dot systems is an increasingly important research topic at the time of writing for a number of reasons, both technological and theoretical. Figure 3.19 presents schematic image of quantum well, quantum wires, and quantum dots.

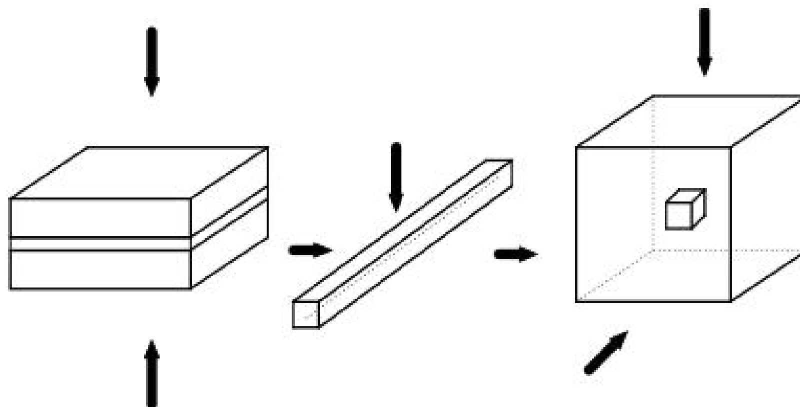


Figure 3.19 Cartoon of confinement along 1, 2 and 3 dimensions. Analogous to a quantum well, quantum wire and quantum dot.

For quantum dots of rectangular shape (with sizes $d_{1,}$ $d_{2,}$ d_3) motion is quantized in all three dimensions. Energies and corresponding wave functions are given by

$$
\begin{aligned}
E_{n_1 n_2 n_3} &= \frac{\hbar^2 \pi^2 n_1^2}{2m^* d_1^2} + \frac{\hbar^2 \pi^2 n_2^2}{2m^* d_2^2} + \frac{\hbar^2 \pi^2 n_3^2}{2m^* d_3^2} \\
\Psi_{n_1 n_2 n_3} &= \sin\left(\frac{n_1 \pi}{d_1} x\right) \sin\left(\frac{n_2 \pi}{d_2} y\right) \sin\left(\frac{n_3 \pi}{d_3} z\right) \quad n_1, n_{2,} n_3 = 1, 2, 3.....
\end{aligned}
\tag{3.37}
$$

Chapter 4 Electrons in low Dimensional Structures of different shape and material distribution

Electronic states in quantum structures are strongly influenced by shape of structures and material distribution profile. In this chapter we consider spherical quantum dots and cylindrical nanowires, as well as quantum wells with confinement potential of non-rectangular shape.

4.1 Electrons in an Infinite Spherical Box

Consider a particle of mass m and energy $E > 0$ moving in the following simple central potential (Figure 4.1):

$$V(r) = \begin{cases} 0 & 0 \le r \le a \\ \infty & otherwise \end{cases} \tag{4.1}$$

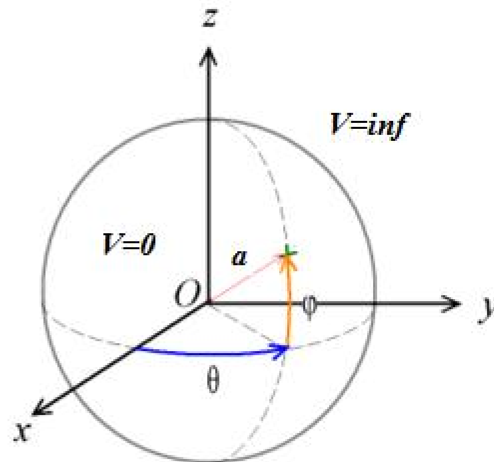


Figure 4.1 Infinite spherical box

Clearly, the wavefunction Ψ is only non-zero in the region $0 \le r \le a$. Within this region, it is subject to the physical boundary conditions that it be well behaved (*i.e.*, square-integrable) at $r = 0$, and that it be zero at $r = a$. The wavefunction is to be written in the standard form – as a product of radial and angular part (for details see textbooks of *Quantum Mechanics*)

$$\Psi\left(r, \theta, \varphi\right) = R_{n,l} Y_{l,m}\left(\theta, \varphi\right) \tag{4.2}$$

where $Y_{l,m}\left(\theta, \varphi\right)$ is a spherical function; r, θ, φ are spherical coordinates. Taking into account that Hamilton operator in spherical coordinates can be written as

$$\hat{H} = -\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r} - \frac{\hat{L}^2}{\hbar^2 r^2}\right) + V\left(r\right) \tag{4.3}$$

where $\hat{L}$ is an angular momentum operator

$$\hat{L}^2 Y_{l,m}\left(\theta, \varphi\right) = \hbar^2 l\left(l+1\right) \tag{4.4}$$

we obtain the equation for the radial part of the wave function (4.2)

$$\frac{d^2 R_{n,l}}{dr^2} + \frac{2}{r}\frac{dR_{n,l}}{dr} + \left(k^2 - \frac{l\left(l+1\right)}{r^2}\right) R_{n,l} = 0 \tag{4.5}$$

In the region $0 \le r \le a$, where

$$k^2 = \frac{2mE}{\hbar^2} \tag{4.6}$$

Defining the scaled radial variable $z = kr$, the above differential equation can be transformed into the standard form

$$\frac{d^2R_{n,l}}{dz^2}+\frac{2}{z}\frac{dR_{n,l}}{dz}+\left(1-\frac{l(l+1)}{z^2}\right)R_{n,l}=0 \tag{4.7}$$

The two independent solutions to this well-known second-order differential equation are called *spherical Bessel functions*, and can be written as

$$\begin{aligned} j_l(z) &= z^l\left(-\frac{1}{z}\frac{d}{dz}\right)^l\left(\frac{\sin z}{z}\right) \\ y_l(z) &= -z^l\left(-\frac{1}{z}\frac{d}{dz}\right)^l\left(\frac{\cos z}{z}\right) \end{aligned} \tag{4.8}$$

Thus, the first few spherical Bessel functions take the form

$$\begin{aligned} j_0(z) &= \frac{\sin z}{z}, \\ j_1(z) &= \frac{\sin z}{z^2}-\frac{\cos z}{z}, \\ y_0(z) &= \frac{\cos z}{z}, \\ y_1(z) &= -\frac{\cos z}{z^2}-\frac{\sin z}{z} \end{aligned} \tag{4.9}$$

These functions are also plotted in Fig. 4.2. It can be seen that the spherical Bessel functions are oscillatory in nature, passing through zero many times. However, the $y_l(z)$ functions are badly behaved (*i.e.*, they are not square-integrable) at $z=0$, whereas the $j_l(z)$ functions are well behaved everywhere. It follows from our boundary condition at $r=0$ that the $y_l(z)$ are unphysical, and that the radial wavefunction $R_{n,l}$ is thus proportional to $j_l(z)$ only. In order to satisfy the boundary condition at $r=a$ [*i.e.*, $R_{n,l}=0$], the value of k must be chosen such

that $z = ka$ corresponds to one of the zeros of $j_l(z)$. Let us denote the nth zero of $j_l(z)$ as $z_{n,l}$. It follows that

$$z_{n,l} = ka, \quad n = 1, 2, 3... \tag{4.10}$$

Hence, from (4.6), the allowed energy levels are

$$E_{n,l} = z_{n,l}{}^2 \frac{\hbar^2}{2ma^2} \tag{4.11}$$

The first few values of $z_{n,l}$ are listed in Table 1. It can be seen that $z_{n,l}$ is an increasing function of both n and l.

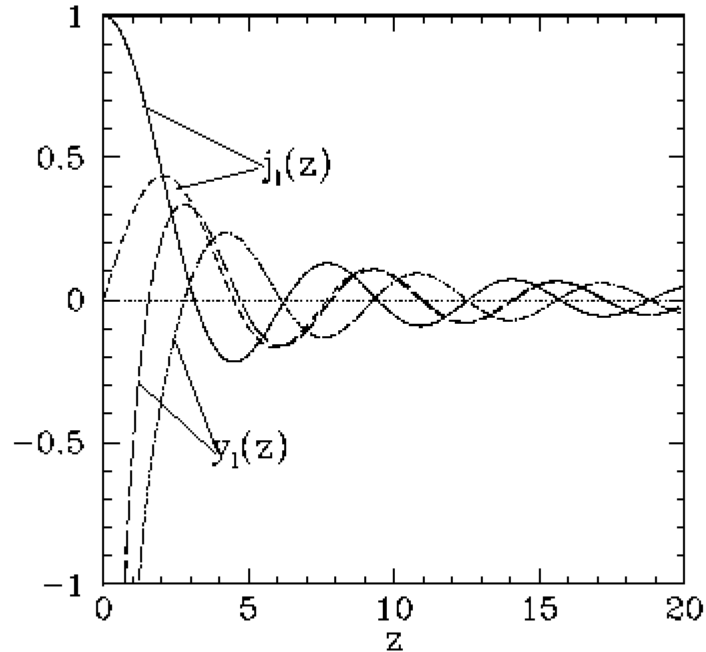


Figure 4.2 The first few spherical Bessel functions.

We are now in a position to interpret the three quantum numbers - n, l, and m --which determine the form of the wavefunction specified in Eq. (4.2). The azimuthal quantum number m determines the number of nodes in the wavefunction as the azimuthal angle ϕ varies between 0 and 2π. Thus, $m=0$ corresponds to no nodes, $m=1$ to a single node, $m=2$ to two nodes, *etc*.

Table.1 *The first few zeros of the spherical Bessel function* $j_l(z)$

	n=1	n=2	n=3	n=4
l=0	3.142	6.283	9.425	12.566
l=1	4.493	7.725	10.904	14.066
l=2	5.763	9.095	12.323	15.515
l=3	6.988	10.417	13.698	16.924

Likewise, the polar quantum number l determines the number of nodes in the wavefunction as the polar angle θ varies between 0 and π. Again, $l=0$ corresponds to no nodes, $l=1$ to a single node, *etc*. Finally, the radial quantum number n determines the number of nodes in the wavefunction as the radial variable r varies between 0 and a (not counting any nodes at $r=0$ or $r=a$). Thus, $n=1$ corresponds to no nodes, $n=2$ to a single node, $n=3$ to two nodes, *etc*. Note that, for the case of an infinite potential well, the only restrictions on the values that the various quantum numbers can take are that n must be a positive integer, l must be a non-negative integer, and m must be an integer lying between $-l$ and l. Note, further, that the allowed energy levels (4.11) only depend on the values of the quantum numbers n and l.

4.2 Electrons in an Infinite Cylindrical Box

Let's first consider ideal system - particle in cylindrical box with infinite walls, so it is assumed that potential energy is 0 inside the box and infinity outside the box. In this case we

only need to solve Schrodinger equation only inside the box, outside the particle wave function is equal to zero (Figure 4.3).

In cylindrical coordinates the Schrodinger equation for $V_{pot} = 0$ has the form:

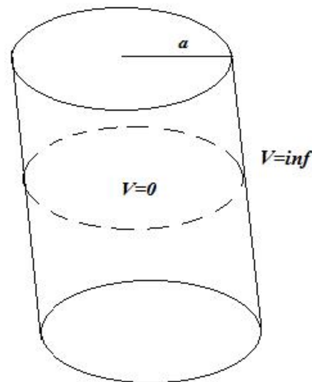


Figure 4.3 Infinite cylindrical box

$$\left[\frac{1}{\ldots}\frac{\partial}{\partial\ldots}\left(\ldots\frac{\partial}{\partial\ldots}\right)+\frac{1}{\ldots^{2}}\frac{\partial^{2}}{\partial_{''}{}^{2}}+\frac{\partial^{2}}{\partial z^{2}}\right]\Phi=-\frac{2mE}{h^{2}}\Phi \tag{4.12}$$

let

$$\frac{2mE}{h^{2}}=k^{2} \tag{4.13}$$

than

$$\left[\frac{1}{\ldots}\frac{\partial}{\partial\ldots}\left(\ldots\frac{\partial}{\partial\ldots}\right)+\frac{1}{\ldots^{2}}\frac{\partial^{2}}{\partial_{''}{}^{2}}+\frac{\partial^{2}}{\partial z^{2}}\right]\Phi=-k^{2}\Phi \tag{4.14}$$

We can represent

$$k^{2}=k^{2}-k_{z}^{2}+k_{z}^{2} \tag{4.15}$$

$$\Phi(\rho,\theta,z)=\Psi(\rho,\theta)\varphi(z) \qquad (4.16)$$

and separate z and (ρ,θ) variables:

$$\left[\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial}{\partial\rho}\right)+\frac{1}{\rho^2}\frac{\partial^2}{\partial\theta^2}\right]\Psi=-(k^2-k_z^2)\Psi \qquad (4.17)$$

$$\frac{d^2\varphi}{dz^2}=-k_z^2\varphi \qquad (4.18)$$

As known, the solution of eq. (4.18) is

$$\varphi(z)=\sqrt{\frac{2}{l}}\sin\frac{n_z\pi}{l}z \qquad (4.19)$$

$$n_z=1,2,3,\ldots.$$

This function describes z motion of particle, which brings $\frac{n_z^2\pi^2 h^2}{l^2}$ part in total energy. Here l is length of wire. If it is too long we can represent z motion as motion of free particle:

$$\varphi(z)=\sqrt{\frac{1}{l}}e^{ik_z z} \qquad (4.20)$$

In Equation (4.17) ρ,θ variables can also be separated. If we multiply the both sides on ρ^2, represent

$$\Psi(\rho,\theta)=P(\rho)Q(\theta) \qquad (4.21)$$

and replace $k^2-k_z^2=k^2-k_z^2+m^2-m^2$ we obtain separate equations for ρ and θ:

$$\frac{\partial^2 Q}{\partial „^2} + m^2 Q = 0 \qquad (4.22)$$

$$\left\{ \ldots^2 \frac{d^2 P}{d\ldots^2} + \ldots \frac{dP}{d\ldots} + \left[\left(k^2 - k_z^2 \right) \ldots^2 - m^2 \right] \right\} P = 0 \qquad (4.23)$$

The solution of (4.22) is

$$Q(„) = Ce^{im\{} \qquad (4.24)$$

Because of the requirement of periodicity $Q(„ + 2f) = Q(„)$ m get the values: $m = 0, \pm 1, \pm 2$.

Normalization condition makes $C = \sqrt{\frac{1}{2f}}$.

By changing variable

$$\sqrt{k^2 - k_z^2} \ldots = < \qquad (4.25)$$

equation (4.23) in transformed into

$$<^2 \frac{d^2 P}{d^2} + < \frac{dP}{d} + \left(<^2 - m^2 \right) P = 0 \qquad (4.26)$$

The solutions of (4.26) owing desired behavior in zero is first kind Bessel function $J_m(<)$. The second Bessel function that satisfies equation (4.26) has singularity in the origin (Fig.4.4). As our wave function should vanish at the surface of the cylinder, we are interested in solution whci has correct behavior at the origin. Substituting (4.25) we obtain

$$P(\ldots) = CJ_m(<) = CJ_m\left(\sqrt{k^2 - k_z^2} \ldots \right) \qquad (4.27)$$

At $\ldots = a$ (a is the radius of cylinder) wave function must be 0. This determines k and, consequently E.

As

$$J_m\left(\sqrt{k^2 - k_z^2}\,a\right) = 0$$

$$\sqrt{k^2 - k_z^2}\,a = \mathrm{r}_{m,n} \tag{4.28}$$

where $\alpha_{m,n}$ is the n-th root of J_m function. Substituting (4.28) in (4.13) we obtain the expression for the total energy of a particle in a cylinder.

$$E = \frac{h^2 k^2}{2m} = \frac{h^2\left(\frac{\mathrm{r}_{m,n}^2}{a^2} + k_z^2\right)}{2m} = \frac{h^2}{2m}\left(\frac{\mathrm{r}_{m,n}^2}{a^2} + \frac{n_z^2 f^2}{l^2}\right) \tag{4.29}$$

The radial part of wave function can be written as:

$$P(\ldots) = CJ_m(<) = CJ_m\left(\frac{\mathrm{r}_{m,n}}{a}\ldots\right) \tag{4.30}$$

C is determined from normalization:

$$C^2\int_0^a J_m\left(\frac{\mathrm{r}_{m,n}}{a}\ldots\right) J_m\left(\frac{\mathrm{r}_{m,n}}{a}\ldots\right)\ldots d\ldots = 1 \tag{4.31}$$

If we introduce new dimensionless variables radius: $\bar{\ldots} = \frac{\ldots}{a}$, and use the properties of Bessel functions

$$\int_0^a J_m\left(\frac{\mathrm{r}_{m,n}}{a}\ldots\right) J_m\left(\frac{\mathrm{r}_{m,n}}{a}\ldots\right)\ldots d\ldots = a^2\int_0^1 \left[J_m\left(\mathrm{r}_{m\,n}\bar{\ldots}\right)\right]^2 \bar{\ldots}\, d\bar{\ldots} = \frac{a^2}{2}\left[J_m'\left(\mathrm{r}_{m\,n}\bar{\ldots}\right)\right]^2 \tag{4.32}$$

We obtain

$$C=\frac{\sqrt{2}}{aJ_m^{'}\left(\alpha_{m,n}\bar{\rho}\right)} \tag{4.33}$$

and

$$P\left(\rho\right)=\frac{\sqrt{2}}{aJ_m^{'}\left(\alpha_{m,n}\bar{\rho}\right)}J_m\left(\alpha_{m,n}\bar{\rho}\right) \tag{4.34}$$

Total wave function, as well as energy depends on three quantum numbers m , n and n_z :

$$\Phi_{n,m,n_z}\left(\rho,\theta,\varphi\right)=\Psi_{m,n}\left(\rho,\varphi\right)\varphi_{n_z}\left(z\right)=\frac{e^{im\varphi}}{a\sqrt{\pi}\,J_m^{'}\left(\alpha_{m,n}\bar{\rho}\right)}J_m\left(\alpha_{m,n}\bar{\rho}\right)\sqrt{\frac{2}{l}}\sin\frac{n_z\pi}{l}z \tag{4.35}$$

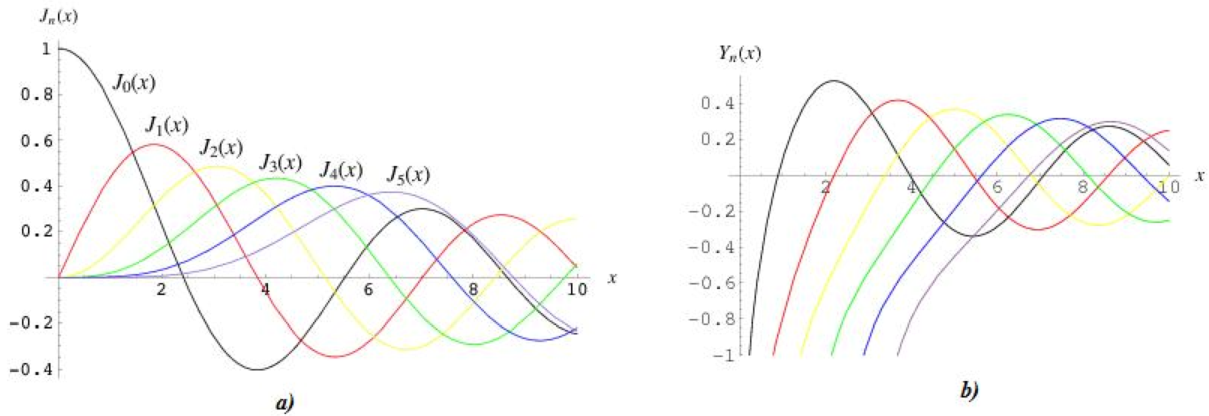


Figure 4.4 a) Bessel function of the first kind ; b) Bessel function of the second kind.

The first few values of $\alpha_{n,m}$ are listed in Table 2.

Table.2 *The first few zeros of the spherical Bessel function* $J_m(x)$

	n=1	n=2	n=3	n=4
m=0	2.4048	5.5201	8.6537	11.7915
m=1	3.8317	7.0156	10.1735	13.3237
m=2	5.1356	8.4172	11.6198	14.7960
m=3	6.3802	9.7610	13.0152	16.2235

4.3 The Effect of Material Distribution Profile

Very often in quantum structures composition variation takes place, which is connected to the material inter diffusion. In this case band edges change smoothly across the interface and confinement potential does not have rectangular shape. Figure 4.4 shows so called core-shell nanostructures. In the nanostructure on left-hand side there is no diffusion of the shell material, while in the nanostructure on right-hand side diffusion of shell material into the core material takes place. Figure 4.5 presents material distribution in quantum dots.

Such systems are often modeled by parabolic potential. Pölsh-Teller type potential is also used for modeling. These two potentials will be considered below.

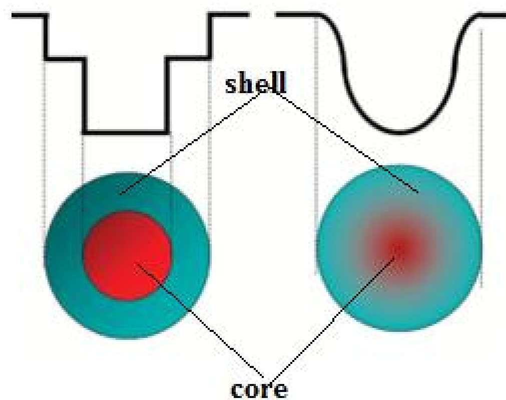


Figure 4.4 core-shell nanostructure with different material distribution profile

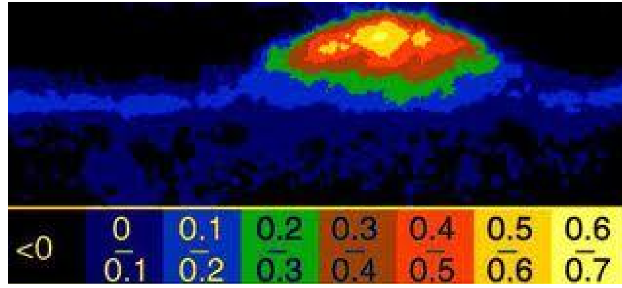


Figure 4.5 Material distribution in quantum dots, different colors correspond to different composition

Parabolic potential

Parabolic potential can be written as

$$V(x) = \frac{1}{2}kx^2 \tag{4.36}$$

This potential describes well confining potential formed by material diffusion in quantum structures.

For this potential the Hamiltonian has the form

$$\hat{H} = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2}kx^2 \tag{4.37}$$

Thus, we have to solve Schrödinger equation

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + \frac{1}{2}kx^2 = E_n\psi(x) \tag{4.38}$$

The equation is substantially simplified if make following notations

$$\xi \equiv \left(\frac{mk}{\hbar^2}\right)^{1/4} x \tag{4.39}$$

$$\text{y} = \frac{2E_n}{\hbar \text{Š}}, \quad \text{Š} = \sqrt{\frac{k}{m}} \tag{4.40}$$

In these notations equation (4.38) gets the following form

$$\frac{d^2\{ (<)}{dx^2} + (\text{y} - <^2)\{ (<) = 0 \tag{4.41}$$

Let’s first consider the solution for large $<$. It will be

$$\{ (<) \sim e^{\pm\frac{1}{2}<^2} \tag{4.42}$$

Proceeding from the fact that the wavefunction must remain finite when $< \to \infty$, only the solution with “-” sigh will be used.

For arbitrary $<$ the wave function can be written as a product of finite polynomial on (4.42)

$$\{ (<) \sim H(<) e^{-\frac{1}{2}<^2} \tag{4.43}$$

Substituting it into (4.41 we arrive at

$$\frac{d^2 H(<)}{d<^2} - 2< \frac{dH(<)}{d<} + (\text{y} - 1) H(<) = 0 \tag{4.44}$$

As mentioned above $H(<)$ is a finite polynomial

$$H(<) = A_0 + A_1< + A_2<^2 + ...A_n<^n \tag{4.45}$$

If substitute the polynomial in (4.44) one can obtain the recurrent relation

$$A_{n+2} = \frac{(2n+1-\lambda)}{(n+2)(n+1)} A_n \tag{4.46}$$

If we want to make polynomial finite, it should be cut at some value of n, for this A_{n+2} Should become zero while $A_n \neq 0$. This will be guaranteed if

$$(2n+1-\lambda) = 0 \tag{4.47}$$

Taking into account that $\lambda = \frac{2E_n}{\hbar\omega}$, the following expression is obtained for discrete energy levels in parabolic well:

$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega \tag{4.48}$$

The order of polynomial defines the number of levels. The higher n the higher the level, and higher the order of the polynomial, and consequently corresponding wavefunction has more nodes, as it should be.

Functions those are the solution of equation (4.44) and satify (4.46) recurent relation are Hermitian polynomials, which is expressed by

$$H(\xi) = (-1)^n e^{\xi^2} \frac{d^n}{dx^n}\left(e^{-\xi^2}\right) \tag{4.49}$$

The first four lower order Hermitian polynomials are

$$\begin{aligned} H_0 &= 1 \\ H_1 &= 2x \\ H_2 &= 4x^2 - 2 \\ H_3 &= 8x^3 - 12x \\ H_4 &= 16x^4 - 48x^2 + 12 \end{aligned} \tag{4.50}$$

Figure 4.5 shows the first five states of electron in parabolic well. Energies and corresponding wave functions are presented.

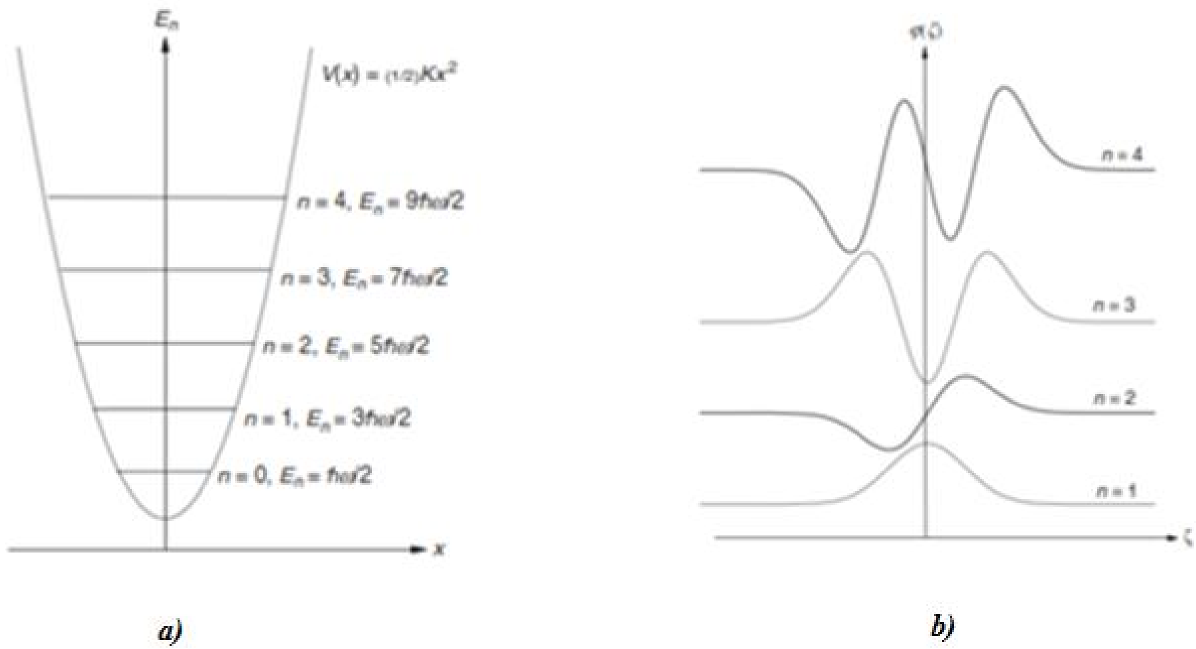


Figure 4.5 a) energy levels of electron in parabolic well, b) corresponding wave functions

Pöschl–Teller Type Potential

Pöschl–Teller type potential also can be used for describtion material diffusion in quantum structures (Figure 4.6). This potential is expressed by

$$U = -\frac{U_0}{\cosh^2 \alpha x} \tag{4.51}$$

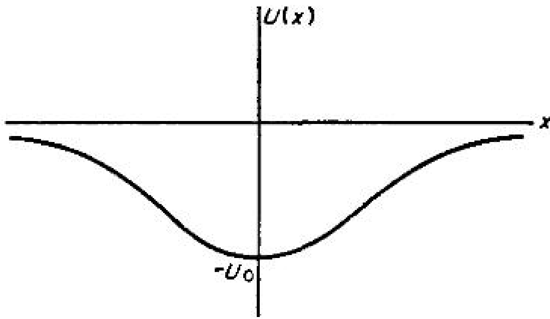


Figure 4.6 Pöschl–Teller potential

We are interested in bound states, therefore the energies are negative and discrete. For one dimensional case the Schrödinger equation has the form

$$\frac{d^2\psi(x)}{dx^2} + \frac{2m}{\hbar^2}\left(E + \frac{U_0}{\cosh^2 \alpha x}\right) = 0 \tag{4.52}$$

Let’s make the substitutions

$$\begin{gathered} \xi = \tanh \alpha x \\ \varepsilon = \frac{\sqrt{-2mE}}{\hbar\alpha}, \ \frac{2mU_0}{\alpha^2\hbar^2} = s(s+1) \end{gathered} \tag{4.53}$$

With these notations (4.52) equation gets the following form

$$\frac{d}{d\xi}\left[\left(1-\xi^2\right)\frac{d\psi}{d\xi}\right] + \left[s(s+1) - \frac{\varepsilon^2}{1-\xi^2}\right]\psi = 0 \tag{4.54}$$

This equation can be deduced to Hypergeometric form by

$$\psi = \left(1-\xi^2\right)^{\nu/2} \varphi(\xi) \tag{4.55}$$

If change the variable

$$u = (1/2)(1-\xi) \tag{4.56}$$

we arrive at

$$u(1-u)\varphi'' + (\nu+1)(1-2u)\varphi' - (\nu-s)(\nu+s+1)\varphi = 0 \tag{4.57}$$

Solution of this equation is the Hypergeometric function. Therefore

$$\psi = \left(1-\xi^2\right)^{\nu/2} F\left[\nu-s, \nu+s+1, \nu+1, \frac{1}{2}(1-\xi)\right] \tag{4.58}$$

For $\xi = -1$ (i.e $x = -\infty$) the Hypergeometri function is finite, if $(\nu - s) = -n$ is negative integer. In this case F is a n th order polynomial. Energy defined from the condition

$$(\nu - s) = -n \tag{4.59}$$

is expressed by

$$E = -\frac{\hbar^2\alpha^2}{8m}\left[-(1+2n) + \sqrt{1+\frac{8mU_0}{\hbar^2\alpha^2}}\right]^2 \tag{4.60}$$

The number of bound states is finite and is defined from the condition $\nu > 0$, that is $n < s$.

Chapter 5 Electrons in Quantum Semiconductor Structures: More Advanced Systems and Methods

In Chapter 4, low-dimensional systems were discussed in terms of a single-electron picture, and the behavior of an electron was examined in the case that it is acted on by various potentials in semiconductors. Those potentials have been supposed to be externally imposed, for instance by a discontinuity in the band gap at an interface between two materials. But an electron will also feel the effect of other electrons in the system in which it finds itself.

There are circumstances in which these many-electron effects can be ignored, for example, in an undoped semiconductor with very few free charges. But in many cases, effects due to the presence of other electrons can be extremely important. Some of the most interesting low-dimensional systems involve many charges: there can be many free electrons, and there will often be in addition some distribution of fixed charges (space charge). To study such systems properly, we must discuss how to take into account the presence of such charges. The problem is one of self-consistency because we are trying to predict the behavior of electrons (or holes), while that behavior will itself depend upon those charges whose behavior we are trying to predict: in other words, the problem itself depends upon the solution to the problem.

5.1 Many-body Effects, Hartree Approximation

Consider the reaction of conduction electrons to the presence of a potential well $V_0(z)$ (we suppose this to be an externally determined well, e.g. a finite square well, which restricts electrons into a two-dimensional region). Available electrons will be attracted to the well. Any one electron will react both to $V_0(z)$ and to the presence of all the other free electrons in the system. (If the system responds in such a way as to leave net fixed charges in some regions of space, the electrons will also interact with those charge distributions; we ignore this effect in the first instance, but it can easily be included later).

The simplest approximation which takes into account the presence of many electrons is to assume that the electrons as a whole produce an average electrostatic potential energy function V_{es} and that a given electron feels the resulting total potential, which is the sum of the original potential and this electrostatic potential

$$V = V_0 + V_{es} \tag{5.1}$$

This is the Hartree approximation. Since the external potential acts in the z-direction only, the electron gas will be confined in z but will be uniformly distributed in the x- and y-directions, so that $V = V(z)$. The electron wavefunction is then obtained from the Schrodinger-like envelope function equation

$$\left[\frac{\hat{p}^2}{2m^*} + V(z)\right]\psi(z) = E\psi(z) \tag{5.2}$$

with V given by equation (5.1).

The mobile electrons in the system all obey equation (5.2). They therefore form a static charge distribution $\rho(z)$ which is constructed from their wavefunctions. It is this distribution of charge which is responsible for V_{es}, the self-consistent part of the potential. Classically, the relation between a charge distribution and the electrostatic potential energy function V_{es} arising from that charge is given by Poisson's equation,

$$\nabla^2 V_{es}(\vec{r}) = \frac{e}{\varepsilon_0}\rho(\vec{r}) \tag{5.3}$$

where ε_0 is the dielectric constant of free space and e is the magnitude of the electronic charge. Note that, in semiconductor physics, by convention, an increase in the magnitude of the electron energy is taken as positive. In the present case, the r- dependence specializes to a z-dependence, while the fact that the electrons are in a semiconductor rather than in free space is taken into account by inclusion of the static dielectric constant $\varepsilon = \varepsilon_r \varepsilon_0$ of the medium (ε_r is the relative dielectric constant of the medium in question). The relevant equation is thus

$$\frac{d^2V(z)}{dz^2} = \frac{e}{\varepsilon_0 \varepsilon_r}\rho(z) \tag{5.4}$$

The self-consistent potential depends on the charge distribution, but that charge distribution depends on ψ:

$$\rho(z) = e\sum_{i} |\psi_i|^2 \tag{5.5}$$

where the sum is over all occupied states. This means that one must sum over each occupied level n of the quantized system, and then integrate over k_x and k_y up to the Fermi energy E_F, for the level in question (i.e., over the energy range $E_F - E_n$ for the level n). If there is also some distribution of fixed charges ρ, in the system then the potential depends on the distribution of all these charges,

$$\rho(z) = e\sum_{i} |\psi_i|^2 + \rho_d \tag{5.6}$$

The space (or depletion) charge density ρ_d will usually be the charge density of ionized donors or acceptors in the system.

Thus, ρ will determine V_{es} through equation (5.4), V_{es} will determine ψ through equation (5.2), ψ will determine ρ through (5.5) or (5.6), and so on. To solve this self-consistent problem, one should start with some reasonable guess. It is common to start with ψ_0, the solution to the problem of a single electron moving in the external potential V_0. The wavefunctions ψ_0 give a first approximation to the charge density using, say, (5.5). An approximate self-consistent potential function V_{es} is then obtained from Poisson's equation (5.3), new wavefunctions are calculated from (5.2) using the improved potential, and the process is repeated. The use of computers makes it a straightforward matter to continue this process until convergence is obtained, i.e. until the electrostatic potential generated from the wavefunctions is the same potential, to within a certain tolerance, as that appearing in the Schrodinger equation for which these wavefunctions are solutions.

5.2 Beyond the Hartree Approximation

The Hartree approximation treats the many-electron problem at the simplest level and is adequate for many purposes. The true solution to the problem, however, is given by the potential

$$V = V_0 + V_{es} + V_{xc} \tag{5.7}$$

Here, V_{xc} is the correction to the potential due to exchange and correlation effects. That such effects exist is evident simply from the Pauli Exclusion Principle — no two electrons can exist in the same quantum state. The presence of an electron in a given state automatically excludes the possibility of another electron being in the same state, and it can be thought of as exerting a sort of repulsion on any other electron. These exchange forces have not been included in the discussion above. In practice, V_{xc} stands for all those many-electron effects not included in the Hartree approximation. There is no exact theory from which V_{xc} can be derived. Exchange effects alone (those due to the Pauli principle) can be treated by the Hartree-Fock approximation, to be found in standard textbooks, but these corrections are rather cumbersome to calculate. A more productive approach seems to be that of the Thomas-Fermi approximation, or its modern extension, density functional theory. Since these corrections are often small in practice, we do not consider them further here.

5.3 The 2DEG at a Heterojunction Interface

An interface between two different semiconductor materials can result in a naturally occurring quantum well and one that is extremely important in practice. Consider a single plane interface between GaAs and AlGaAs of the type previously discussed. We suppose that the AlGaAs is n-doped, while the GaAs is undoped (Fig. 5.1). This is known as modulation doping.

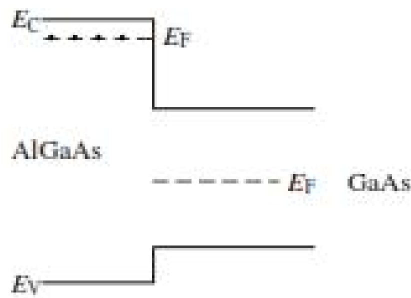


Figure 5.1 Band-edge diagram for a GaAs-AlGaAs interface before redistribution of charge has taken place. Mobile electrons can easily be excited into the conduction band of the n-doped AlGaAs, while the GaAs is undoped. E_v and E_c refer to valence and conduction band edges, respectively

The situation shown in Fig. 5.1 is physically impossible in equilibrium, because in equilibrium the system as a whole must have a common chemical potential which, in this case, is the Fermi energy E_F. Thus, if such a system is created, it must be unstable. What will happen is that electrons will be thermally excited into the conduction band of AlGaAs. They will then migrate into the adjoining GaAs, since there they can achieve states of lower energy (they will lose energy, e.g. by collisions with phonons). These electrons will thus leave behind in the AlGaAs a lack of electrons, i.e. the ionized donors will no longer be screened by an equal number of electrons, and the AlGaAs will acquire a net positive charge, which will build up as more electrons move into the GaAs. The mobile electrons which are now in the GaAs will be attracted by the fixed positive charge in the AlGaAs, but will no longer have enough energy to recombine with their ionized donors. They will thus be trapped in the vicinity of the interface: the band-edge discontinuity will prevent them from moving to the left, while the Coulomb attraction of the net positive charge in the barrier material will keep them from moving very far to the right. The process continues until the system reaches equilibrium and the electrons have been trapped in a quantum well, forming a two-dimensional electron gas at the interface between the two materials.

The charge distribution at such a heterojunction would thus be that shown in Fig. 5.2. The forces resulting from this redistribution of charge are conventionally represented as a bending of

conduction and valence bands (Fig. 5.3). The slope of the bands in the neighbourhood of the interface is proportional to the electric field there. Donor states higher than the Fermi level must be unfilled (the region of positive space-charge in the AlGaAs). Conduction-band states below the Fermi energy must be filled (electrons in these states, in the GaAs, form the 2DEG). The electric field must be continuous across the interface, as shown by the equal band-edge slopes in the two materials there (we will ignore the small change in electric field caused by the slight difference between the dielectric constants of the two materials). However, the electric field must be zero far from the interface, since there the materials are required to have their original bulk properties. (We assume that no external potential has been imposed on this system.) We see that self- consistency is vital for a correct description of this system, since only if the charge due to the electrons themselves is included in the potential will the potential correctly go to zero at infinity.

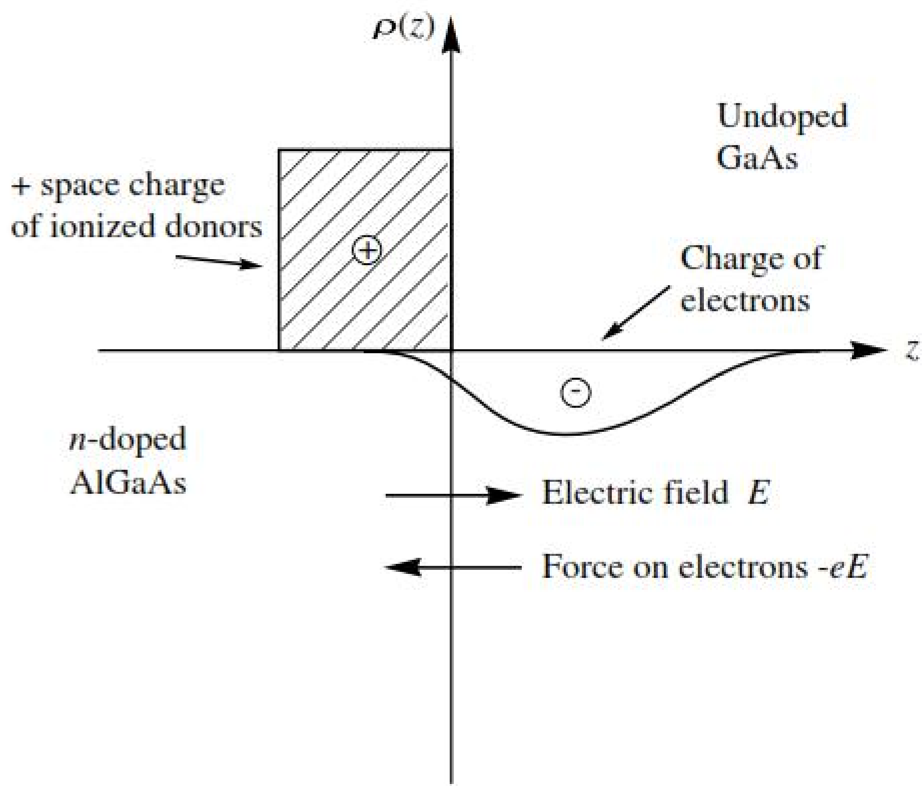


Figure 5.2 Charge distributions in a simple heterostructures

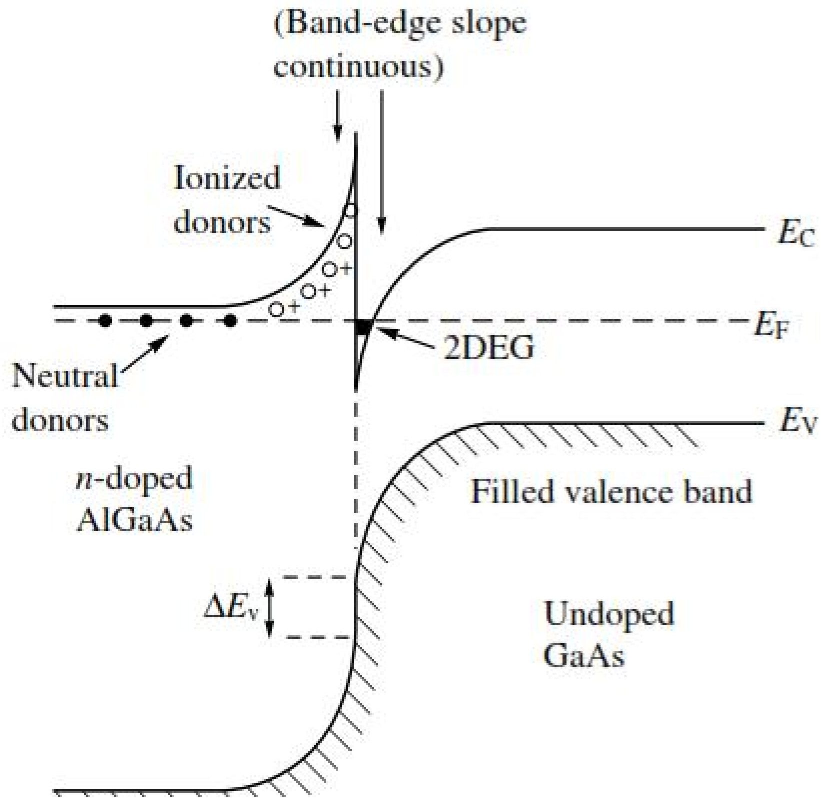


Figure 5.3 A 2DEG in a heterostructure

This type of quantum well is among the most widely used and studied, occur ring in slightly different forms with different names but with broadly similar characteristics. Two of these are the MOSFET (metal-oxide-semiconductor field effect transistor) and the MISFET (metal-insulator-semiconductor FET). These systems (Fig.5.4) have the useful property that the number of electrons in the 2DEG can be controlled experimentally. The metal layer forms a gate, the potential of which can be varied so as to attract electrons to the surface of the p-doped Si. When the attraction is strong enough that the conduction band is bent well below the Fermi energy, a quantum well is again formed and electrons will be trapped in a 2DEG, as shown in Fig. 5.5. Here, the depth of the well, and thus the number of electrons that can be trapped by

it, is regulated externally by the potential on the gate. The electrons form an inversion layer, so called since the normal state of affairs in a p-doped semiconductor is for current to be carried by mobile holes, rather than electrons.

Other similar systems go by the names HEMT (high-electron-mobility transistor) and MODFET (modulation-doped field-effect transistor), and have been used to obtain 2DEGs of very high mobility. Such systems are particularly good for the investigation of phenomena such as the (fractional and integer) quantum Hall effect. One way in which the mobility of two-dimensional electrons in these systems has been improved is by inserting a spacer layer between the doped region and the 2DEG, as illustrated in Fig. 5.6.

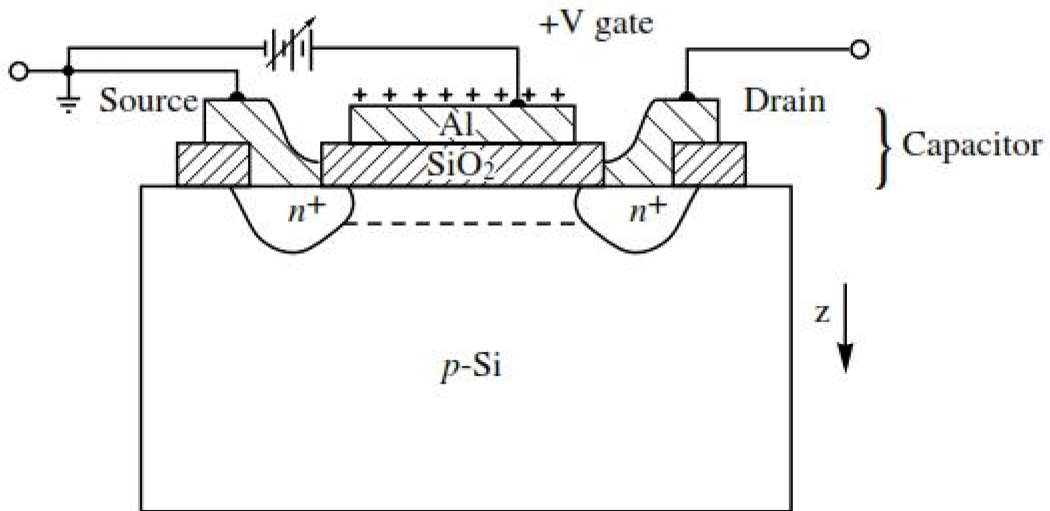


Figure 5.4 Schematic diagram of a MOSFET.

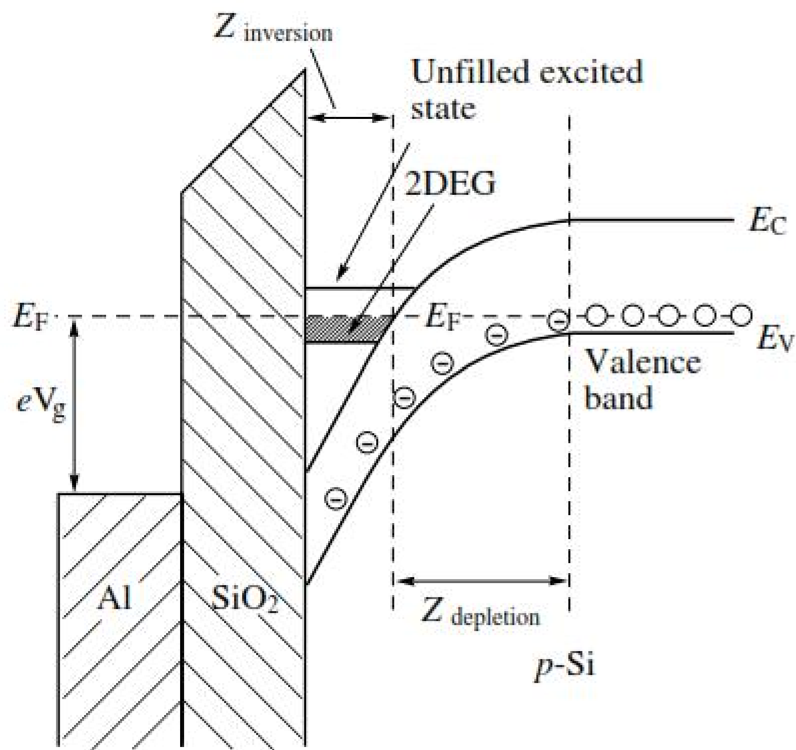


Figure 5.5 Band-edge diagram for the MOSFET system of Fig. 5.4

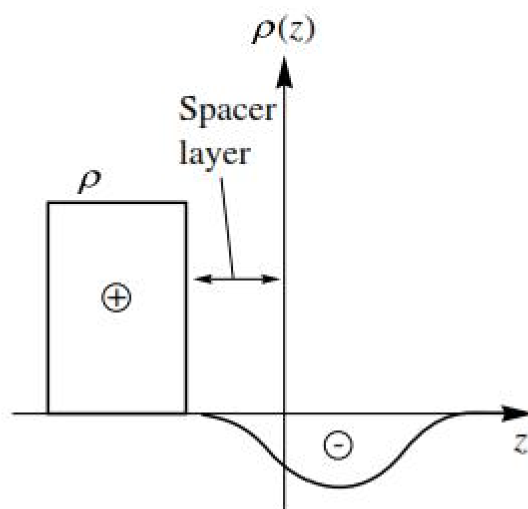


Figure 5.6 Charge distribution in a MODFET (modulation doped field effect transistor).

5.4 The Ideal Heterojunction

Predicting the properties of a 2DEG at a heterojunction is usually a process that involves full-scale computation. Nevertheless, the appropriate starting point is in the simplest possible description of the system. The starting point usually chosen is that of an infinite triangular well (Fig. 5.7). Here the band-gap discontinuity at the interface, to the left, is approximated by an infinite potential step, while the conduction band edge is assumed to have a constant slope corresponding to the value of the confining electric field which is present at the interface.

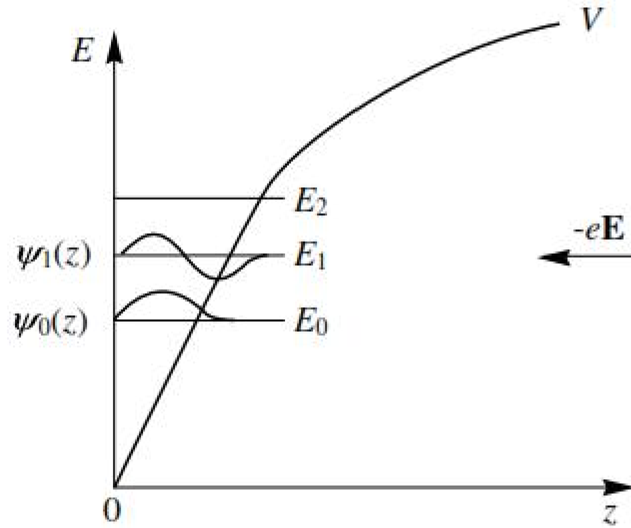


Figure 5.7 The infinite triangular well. The rounding of the band edge to the right of the diagram is drawn to correspond to physical expectations, while electron energies and wavefunctions are calculated assuming a strictly linear potential, to this approximation; E is the electric field at the interface.

This approximation is simple enough to allow closed-form solutions, which are Airy functions. As one would expect, the electron energies are quantized, and the electrons themselves are confined to a narrow region to the right of the interface. The predicted electron states spread

out progressively in each higher-energy quantum level. All these properties are physically correct at a qualitative level, but they can also be taken to be a reasonable physical approximation if there is a relatively low electron density in the 2DEG.

Since even Airy functions involve the use of tables, or calculations, it is useful to have on hand an approximation technique which can often be used in conjunction with simple physical approximations for the system of interest, to obtain quick, approximate, and often simple numerical predictions. The following section discusses two such techniques.

5.5 Some Calculational Methods

In this section we consider two methods which can give surprisingly good approximate predictions for low-dimensional structures in which quantum mechanical effects are important: the Wentel-Kramers-Brillouin (WKB) approximation and the Thomas-Fermi approximation. They are simple enough to be used either separately or together, and can sometimes give analytic answers for quantum energies and wavefunctions in small structures.

The WKB approach is an approximate way of solving the Schrodinger equation and is appropriate for systems in which many-electron effects are either weak or absent. (It can also be useful, as a first approximation, even when many-electron effects are important.) The Thomas-Fermi approach (in the version presented here) is an approximate way of taking many-electron effects into account, giving what can be a very good approximation to the full self-consistent potential felt by an electron in the presence of an external potential together with the band-bending effects resulting from the presence of all the other electrons. An even better approximation can be obtained when these methods can be combined, using the WKB approximation to solve the Schrodinger equation for an electron in a self-consistent potential which has itself been obtained from the Thomas- Fermi approximation.

The WKB Approximation

We consider here a method for calculating approximate wavefunctions and energy levels which is "semiclassical", but can nevertheless be quite powerful. This is the WKB approximation. If many-body effects are neglected, the WKB approximation often leads to analytic results, or at least to closed-form expressions (in which the answer can be obtained numerically by doing a simple integral). Even in systems where many-electron effects are important, the WKB approximation can yield useful information about the way in which wavefunctions and energies depend upon the parameters of the system. Here, we do not justify the method in detail, but merely present the resulting approximation procedure in a way that can be applied to systems of interest.

The WKB approximation is semiclassical: in some sense it can be thought of as an expansion that is good when quantum effects are small. The procedure can be stated in terms of a classical picture of the motion of a particle in a potential (Fig. 5.8). Classically, if a particle of energy E moves in a potential V , the particle can be found only in regions where $E > V$. The turning points of its motion (a, b in the diagram) are the points at which $E = V$ (the particle must turn around, since it cannot proceed into a forbidden region). If V were constant in each region (e.g. a finite square well), the solution of the Schrodinger equation, W , would be simply

$$W = \begin{cases} e^{\pm ikz}, \text{ for } E > V \\ e^{\pm i| z}, \text{ for } E < V \end{cases} \quad (5.8)$$

i.e. a travelling wave in the classically allowed region and a decaying exponential in the classically forbidden regions. If V is a slowly varying function of position, then one can try to approximate W by a form similar to that of equation (5.8):

$$W = e^{iu(z)} \quad (5.9)$$

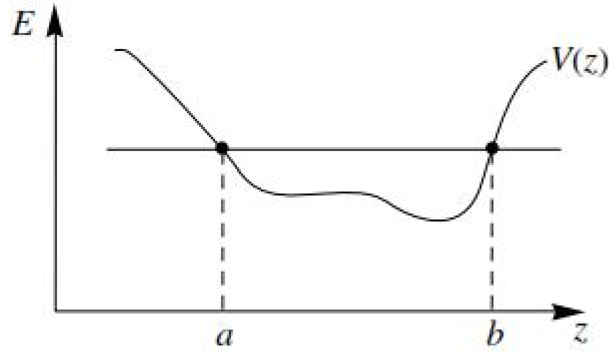


Figure 5.8 Turning points a and b for a particle of energy E in a potential V

where

$$u = \pm \int k(z)\,dz \tag{5.10}$$

with

$$k = \begin{cases} \frac{1}{\hbar}\left[2m(E-V)\right]^{1/2}, \text{ for } E > V \\ -i|\, z, \text{ for } E < V \end{cases} \tag{5.11}$$

so k is real in the classically allowed regions and purely imaginary in the classically forbidden regions, as in equation (5.8). The WKB approximation consists in expressing W as

$$\mathrm{W}(x) = \frac{c}{\sqrt{k(x)}} e^{\pm i \int k(x)\,dx} \tag{5.12}$$

According to this equation, we can write W, for example, in terms of *sines* and *cosines* in the allowed region, and real exponentials in the forbidden regions. It is then necessary to connect these two sorts of solutions at the turning points a and b. The appropriate connection formulae at the left-hand turning point (a, in Fig. 5.8) are

$$
\begin{array}{ccc}
\text{Forbiden region} & & \text{Allowed region} \\
-\frac{1}{\sqrt{|}}\exp\left(\int_z^a | \, dz\right) & \Leftarrow\!\rightarrow & \frac{1}{\sqrt{k}}\sin\left(\int_a^z k dz - \frac{1}{4}f\right) \\
\frac{1}{\sqrt{|}}\exp\left(-\int_z^a | \, dz\right) & \leftarrow\!\Rightarrow & \frac{1}{\sqrt{k}}\cos\left(\int_a^z k dz - \frac{1}{4}f\right)
\end{array}
\tag{5.13}
$$

and, for turning points like b (right-hand turning points),

$$
\begin{array}{ccc}
\text{Forbiden region} & & \text{Allowed region} \\
\frac{2}{\sqrt{k}}\cos\left(\int_z^b k dz - \frac{1}{4}f\right) & \Leftarrow\!\rightarrow & \frac{1}{\sqrt{|}}\exp\left(-\int_b^z | \, dz\right) \\
\frac{1}{\sqrt{k}}\sin\left(\int_z^b k dz - \frac{1}{4}f\right) & \Leftarrow\!\rightarrow & -\frac{1}{\sqrt{|}}\exp\left(\int_b^z | \, dz\right)
\end{array}
\tag{5.14}
$$

These connection formulae apply to turning points at which the potential is in some sense slowly varying. Other places at which one wants to apply the usual boundary conditions are at $z = \pm\infty$, where one specifies that W cannot correspond to a state that is exponentially increasing, or at an infinite potential barrier (a “solid wall”), where the wavefunction is required to vanish.

The arrows in equations (5.13) and (5.14) are not symmetric. This indicates that the connection between the indicated expressions is rigorous in one direction (double arrow) but not in the other. Thus, for example, an exponential with a negative coefficient to the left of the left-hand turning point (a) will always imply the existence of a cosine-type solution to the right of that turning point (the second of equations (5.13)), but the existence of a cosine-type component to the wavefunction to the right of a may not imply the existence of an exponential of negative coefficient to the left. Care is needed in the direction of the single arrows because the method is not sensitive to the existence of extremely small terms in forbidden regions (at any appreciable distance into the forbidden region, a growing exponential swamps a decaying

one, so that, if a growing one is present, it is not possible to say whether a decaying one is there or not).

It is the matching conditions at the turning points and other boundaries that lead to quantization of the energy: only for certain discrete values of E will it be possible to find a solution w which can be matched properly for all z. Since the matching conditions are approximate, however, the quantized values of E will also be approximate. Nevertheless, the approximation often proves to be a surprisingly good one. Note, too, that one has approximate wavefunctions (equations (5.13) and (5.14)) as well as approximate energies.

Examples

1. **Infinite well** (Figure 5.9) given by

$$V(x)=\begin{cases} \mathrm{f}(x), & 0<x<a \\ \infty, & x<0, x>a \end{cases} \tag{5.15}$$

where $\mathrm{f}(x)$ is slowly-varying function.

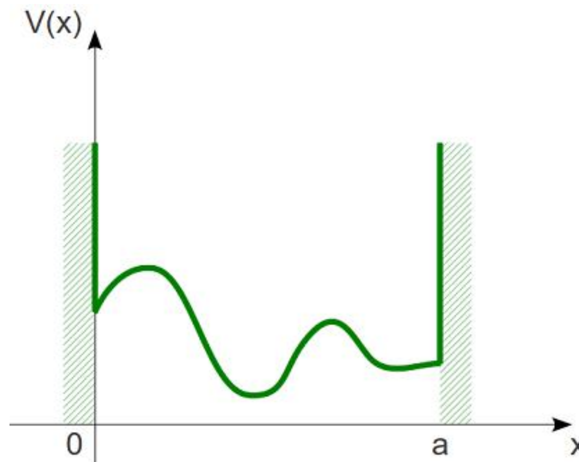


Figure 5.9 Infinite well with non-homogeneous bottom

When $E > V(x)$ the following expression is obtained for the wavefunction

$$Œ(x) = \frac{1}{\sqrt{k(x)}} \left(C_+ e^{i\int_0^x k(x')dx'} + C_- e^{-i\int_0^x k(x')dx'} \right) =$$
$$= \frac{1}{\sqrt{k(x)}} \left[C_1 \sin\left(\int_0^x k(x')dx' \right) + C_2 \cos\left(\int k(x')dx' \right) \right] \quad (5.16)$$

This wavefunction has to satisfy the

$$Œ(0) = Œ(a) = 0 \quad (5.17)$$

Which means cosine terms will disappear in (5.16), the equating the sine to zero gives

$$\int_0^a k(x)dx = nf \quad (5.18)$$

For $V(x) = 0$ exact solution can be obtained

$$E_n = \frac{n^2 f^2 \hbar^2}{2ma^2} \quad (5.19)$$

which coincides to the analytical results for infinite rectangular well.

2. **Arbitrary Shape Well**

We are considering the well given on Figure 5.8. In this case, W must be a decaying exponential as $z \to \pm\infty$. This means that we must use the second of equations (5.13) and the first of equations (5.14). These formulae give two different cosine expressions for the wavefunction; the condition that these expressions must be the same imposes a condition on the arguments of the cosines, which produces the quantization condition,

$$\int_a^x kdx + \int_x^b kdx = \frac{1}{\hbar}\int_a^b \left[2m\left(E-V\right)\right]^{1/2} dx = \left(n+1/2\right)f \qquad (5.20)$$

where n = 0, 1, 2, This condition is, in fact, the Bohr-Sommerfeld quantization condition, that a particle's orbit in phase space must equal a half-integer multiple of f . This is essentially a condition for a standing wave.

3. **Triangular well**

We are calculating a bound states in a triangular potential well as, for example, the simple picture of an inversion layer (Fig. 5.10), for which the potential energy function V is given by

$$V\left(z\right) = \begin{cases} \mathrm{e}\left|\mathrm{E}\right|\mathrm{z}, & z > 0 \\ \infty, & z < 0 \end{cases} \qquad (5.21)$$

where E denotes the magnitude of an electric field. The wavefunction must decay exponentially for positive z in the forbidden region and in this case must vanish identically at z = 0. In the classically forbidden region where z > 0, we must use the first of equations (5.14). Then, for the cosine function to vanish at the origin, we must require that

$$\int_a^b kdz = \left(n+3/4\right)f \qquad n = 0,1,2..... \qquad (5.22)$$

from which it is simple to show that the quantized energies E_n are, in this approximation, given by

$$E_n = \left(\frac{\hbar^2 e^2 \mathrm{E}^2}{2m}\right)^{1/3} \left[3/2(\mathrm{n}+3/4)\right]^{2/3} \qquad (5.23)$$

The WKB approximation has given us not only approximate values for the energy levels, equation, but also approximate wavefunctions,

$$\psi(z)=\frac{1}{\sqrt{k(z)}}\sin\left[\int_a^z k(z')dz'-\frac{1}{4}\pi\right] \tag{5.24}$$

example of a triangular potential well is often used as the starting point for a treatment of two-dimensional electrons in an inversion or accumulation layer.

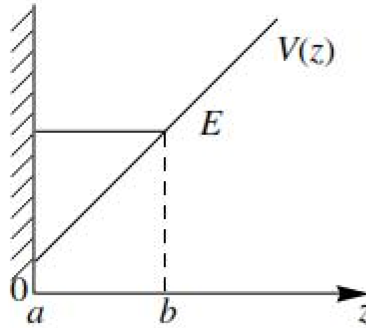


Figure 5.10 Triangular potential well at an infinite potential barrier.

Finally, we note that the WKB method as discussed here does not include many-electron effects, since it assumes that the confining potential is known in advance. To generalize to many-electron systems, one can use the WKB method as part of an iterative solution together with the Poisson equation. Alternatively, one can use the WKB method to solve a Schrodinger equation with an approximate but self-consistent potential obtained from a Thomas-Fermi approximation.

5.6 The 2DEG in Doping Wells

So far we have discussed compositional quantum wells and quantum wells at heterojunctions. Another important example is that of doping wells (Fig. 5.11). Consider an intrinsic semiconductor (e.g. GaAs), grown one plane atomic layer at a time, which has been uniformly n-doped in a slab of width *d* during growth. Suppose that all donors have become ionized. The resulting positive charge of these donors creates an attractive force which is then felt by the

mobile (donated) electrons, which are trapped by that force. The attraction can easily be strong enough, over a short enough distance *d*, that quantum effects are important. This is yet another kind of two-dimensional electron gas.

The charge density of the donors is assumed to be

$$\ldots\left(z\right) = en_d\left(z\right) \tag{5.25}$$

where the donor doping density n_d is constant for $0 \le |z| < 1/2\,d$ and the potential energy function V of these donors create is given by Poisson's equation:

$$\frac{d^2V}{dz^2} = \frac{e}{\mathrm{V}_r\mathrm{V}_0}\ldots\left(z\right) \tag{5.26}$$

This equation is easily solved: if the second derivative of V is constant, it must be quadratic in z,

$$V = a + bz + cz^2 \tag{5.27}$$

We can choose *a* to be zero since the choice of a zero of energy is always arbitrary. Moreover, if we choose the doping slab to be centred at $z = 0$, as in Fig. 5.11, then $b = 0$ from symmetry. Thus,

$$V\left(z\right) = \frac{e^2}{\mathrm{V}_r\mathrm{V}_0}n_D z^2, \quad 0 \le |z| < 1/2\,d \tag{5.28}$$

which yields a parabolic potential well within the region of donor doping. The potential in the undoped regions is even more straightforward: in a region of zero charge, the right-hand side of equation (5.26) must be zero, and the potential must therefore be linear in *z* (and symmetric in *z*):

$$V = a' + b'|z| \tag{5.29}$$

Here a' and b' are determined from equation (3.24) at $z = \pm\frac{1}{2}d$ using the fact that the potential must be continuous. One easily finds that

$$a' = -\frac{e^2}{2\varepsilon_r\varepsilon_0} n_D \left(\frac{1}{2}d\right)^2, \; b' = \frac{e^2}{2\varepsilon_r\varepsilon_0} n_D d \tag{5.30}$$

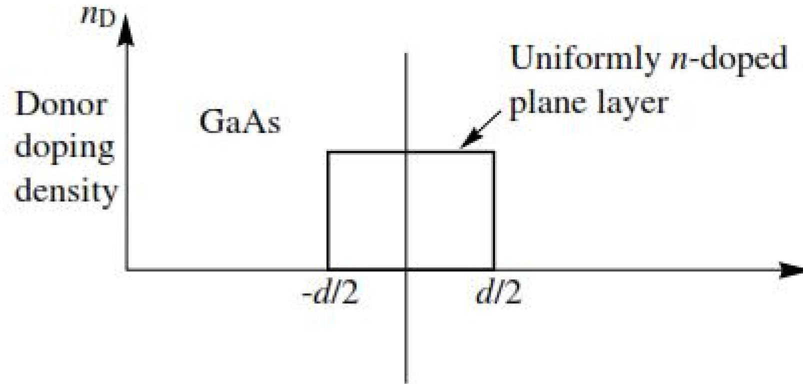


Figure 5. 11 Donor density profile: a uniformly n-doped slab creates a potential well within the host material.

The resulting The resulting quantum well is shown in Fig. 5.12

If electrons are actually trapped in the well, then their presence will modify this potential (self-consistent effects). Ignoring these corrections, the energies and wave functions for electrons in the well are easily obtained, e.g. from the WKB approximation. In the region $z \le \frac{1}{2}d$ the potential is that of a simple harmonic oscillator, for which the WKB approximation gives the exact answers. Thus, for instance, the energies are given by

$$E_n = \hbar\omega\left(n + \frac{1}{2}\right), \quad n = 1, 2, ... \tag{5.31}$$

where the “natural frequency” ω is

$$\omega = \sqrt{\frac{k}{m}} = \left(\frac{e^2 n_d}{2m\varepsilon_r\varepsilon_0}\right)^{1/2} \tag{5.32}$$

Self-consistent effects can modify this potential in important ways. First, $V(z)$ itself will be changed if electrons are present in the well. Secondly, the ionized donors themselves are more a random collection of point charges than a continuous uniform charge distribution. This randomness results in fluctuations in the potential $V(z)$. Electrons trapped in the well will tend to redistribute themselves in such a way as to reduce these potential fluctuations, thus screening out some of the effects of disorder.

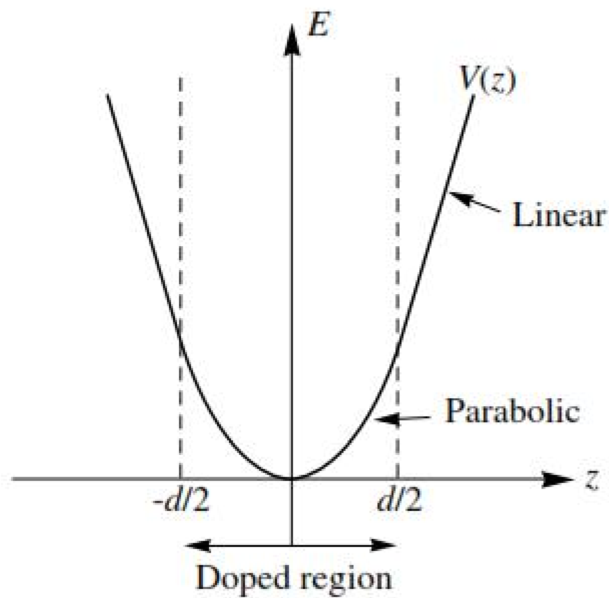


Figure 5.12. Quantum well created by the donor doping profile of Fig. 5.11

5.7 The Delta Well

An extreme example of a doping well is a delta well, formed when the layer of donors in a host material is as small as one atomic layer wide, approximately a delta-function distribution in z. (These are also called delta-doped, or spike-doped, systems.) One can therefore write

$$n_D(z) = D\delta(z) \tag{5.33}$$

where D is the number of ionized donors per unit area in the host material. Poisson's equation then gives the very simple potential energy function

$$V(z) = a + b|z| \tag{5.34}$$

with a chosen to be zero for convenience, and

$$b = \frac{e^2 D}{2\text{v}_r \text{v}_0} \tag{5.35}$$

This potential is given in Fig. 5.13.

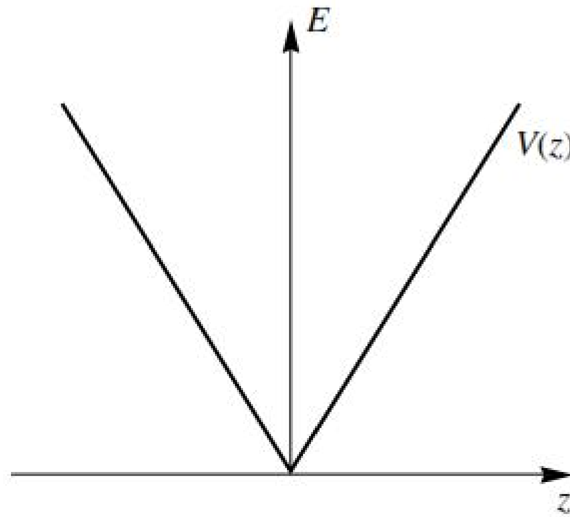


Figure 5.13 Bare potential for a delta well.

In this case the WKB approximation gives the energy levels

$$E_n = \left(\frac{\hbar^2 b^2}{2m}\right)^{1/3} \left[\frac{3}{4} f \left(n + \frac{1}{2}\right)\right]^{2/3} \tag{5.36}$$

As always, this can be the starting point for a better approximation, which must include self-consistent effects if electrons are in fact trapped in the well. These effects can be quite important, particularly for high dopin. Here, the number of electrons trapped by the well equals the number

of donors, and the potential (heavy black line) is seen to differ considerably from the linear, "bare" potential shown in Fig. 5.13.

We digress briefly here to note that delta-doping can be used, among other things, to create high electron mobility devices, starting from a basic accumula- tion or inversion 2DEG. Figures 5.1 and 5.2 show how electrons in such a 2DEG can come from nearby ionized donors. One is then often interested in the mobility of these electrons in the two-dimensional plane in which they are free to move.

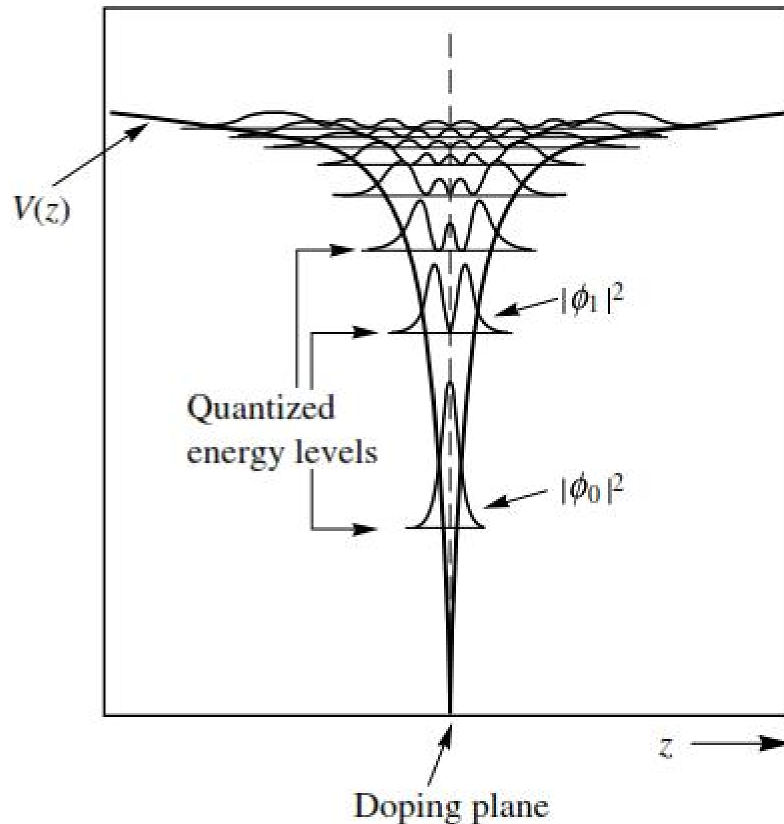


Figure 5.14 Energies and electron probability-densities for a delta well in InSb (self-consistent calculation; probability densities are each normalized to unity).

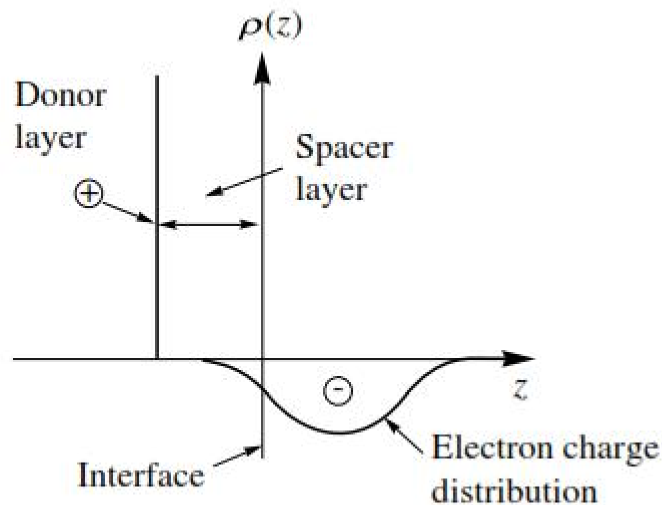


Figure 5.15 Charge distribution in a high mobility electron gas created at an interface with delta donor doping.

This mobility can be severely limited by the scattering of electrons from ionized impurities. In the situation illustrated in Figs. 5.1 and 5.2, a large number of these impurities lie very close to the plane of the 2DEG and will strongly limit its mobility. The mobility of the 2DEG can be increased by placing a spacer layer between the donors and the plane of the 2DEG, thus making the scattering centres more distant. If the donor region is made into a delta-layer, this scattering can be reduced still further (Fig. 5.15). Here, any free charges to the left of the interface will be tightly bound to the donor layer, and will tend to screen out disorder due to the random positions of the donors. This will help to produce an approximately uniform charge distribution, further reducing ionized impurity scattering.

5.8 The Thomas—Fermi Approximation for Two-dimensional Systems

The Thomas-Fermi approximation allows one to take account of many-electron effects on the potential felt by a single electron (in other words, it can give an approximate way of calculating

band-bending effects caused by the presence of many electrons or holes). This approximation can provide a very good description of the potential for the particular cases of heterojunctions, accumulation and inversion layers, and delta-doped systems in the presence of many-electron effects, and in some cases provides simple analytic formulae for the potential. Its use for other systems may be less straightforward.

When it can be used in this way, the Thomas-Fermi approach provides an approximate alternative to obtaining a self-consistent iterative solution of coupled Poisson and Schrodinger equations, as described in the preceding section. An early reference to this use of the method is given by Keyes (1976), whose approach we follow here.

We take the potential felt by an electron in a quantum well to be a sum of several parts:

$$V_{tot} = V_{ext} + V_{fixed} + V_e \tag{5.37}$$

where V_{ext} t is the external potential (the band-edge potential in the absence of free charges, and of any applied electric field), V_{fixed} is the potential due to any fixed space-charge distribution in the system, and V_e the self-consistent contribution to the potential (arising from the presence of mobile electrons in the system).

In terms of the density $n_e\left(\vec{r}\right)$ of mobile electrons, the self-consistent part of the potential obeys Poisson's equation

$$\frac{d^2V_{tot}}{dz^2} = \frac{e^2}{\mathrm{V}_r \mathrm{V}_0}\left[n_e\left(\vec{r}\right) + n_A\left(\vec{r}\right) + n_D\left(\vec{r}\right)\right] \tag{5.38}$$

where n_A and n_D are the densities of charged acceptors and donors, respectively. We next suppose that there are a number of mobile electrons in the system, and use the familiar relation between Fermi level (measured from the conduction band edge) and three-dimensional electron density, n_e

$$E_F = \left(3f^{\,2}\right)^{2/3} \frac{\hbar^2}{2m^*} n_e^{\,2/3} \tag{5.39}$$

which follows from the three-dimensional density of states. Here, E_F is understood to mean the difference between the Fermi energy and the total band-edge potential V_{tot}. Since depends on position, we write equation (5.39) as

$$E_F - V_{tot}\left(z\right) = \left(3f^{\,2}\right)^{2/3} \frac{\hbar^2}{2m^*} n_e\left(z\right)^{2/3} \tag{5.40}$$

If we choose our zero of energy to lie at E_F, we then obtain

$$-V_{tot}\left(z\right) = \left(3f^{\,2}\right)^{2/3} \frac{\hbar^2}{2m^*} n_e\left(z\right)^{2/3} \tag{5.41}$$

which can be rearranged to

$$n_e = \frac{1}{3f^{\,2}}\left(\frac{2m^*}{\hbar^2}\right)^{3/2}\left(-V_{tot}\right)^{3/2} \tag{5.42}$$

We can now obtain an approximation for V_{tot} if we restrict ourselves to external potentials which are either linear or constant in growth direction z. This includes inversion and accumulation layers and delta-doped systems when no space charge is present. (It can also be a reasonable first approximation even in the presence of space charge, since such charge is often distributed over lengthscales much larger than that of the quantum well in question, giving a relatively weak contribution to the band bending).

We thus consider Poisson's equation (5.38) and note that, since V_{ext} is assumed to be a straight line and V_{fixed} is neglected here,

$$\frac{d^2V_{sc}}{dz^2} = \frac{d^2V_{tot}}{dz^2} \tag{5.43}$$

Thus we can combine equations (5.38) and (3.39) to eliminate n_e (which is not yet known), to obtain

$$\frac{d^2V_{tot}}{dz^2} = \frac{1}{3f^{\,2}}\left(\frac{2m^*}{\hbar^2}\right)^{3/2}\left(-V_{tot}\right)^{3/2} \tag{5.44}$$

This is a differential equation which can be solved for the self-consistent band-edge potential V_{tot}. Note that V_{tot} itself is negative: the approximation refers to systems which contain mobile electrons, and does not apply when there are none (i.e. when $V_{tot} > 0$).

5.9 The Thomas—Fermi Approximation for Heterojunctions and Delta Wells

The Thomas—Fermi approximation of equation (5.44) is found to have a simple analytic form when applied to heterojunctions or delta wells with no background doping. It is easy to verify that (5.44) is satisfied for a potential of the form

$$V(z) = -b \frac{1}{\left(|z| + z_0\right)^4} \tag{5.45}$$

where b and z_0 are positive constants, and the interface (or the position of delta-doped layer) is at $z = 0$. A few simple manipulations show that in this case

$$b = \left(\frac{60f^{2} \mathrm{V}_{r} \mathrm{V}_{2}}{e^2}\right)^2 \left(\frac{\hbar^2}{2m^*}\right)^3 \tag{5.46}$$

and z_0 is determined by a boundary condition (e.g. at $z = 0$) for the particular system of interest. Since V_{tot} is given explicitly by (5.45) and (5.46), one also has an explicit form for the density n_e of mobile electrons from equation (5.42),

$$n_e = \frac{a}{\left(z + z_0\right)^6}, \quad a = \left(\frac{2m^* b}{\hbar^2}\right)^{3/2} \tag{5.47}$$

This charge density can be a very good approximation to that obtained from a full self-consistent solution of Schrodinger's and Poisson's equations.

One can now obtain (approximate) energy levels and subband occupations by solving the Schrodinger equation, using the approximate potential V_{tot} obtained by this procedure. The exact

self-consistent solution can also be obtained from this starting point, thus saving much computing time.

When the potential approaches zero as $z \to 0$, one boundary condition is needed to fix the single unknown z_0 . This may be, for instance, the electric field at the interface (i.e. the slope of the potential there), or the value of the band-edge potential at $z = 0$, or else the total two-dimensional density of mobile electrons in the system. In the more general case in which the system is not charge-compensated, for instance by removing mobile electrons from the system by some external mechanism, the situation is slightly more complicated, but an explicit analytic solution can still be found.

5.10 Excitons in Hartree Approximation

Exciton is a electron-hole pair bound by Coulomb interaction. In bulk crystals in the frame of effective mass approximation the effect of ions of crustal lattice on the motion of electrons and holes is included in their effective masses. Therefore the Hamilton operator of electron-hole pair consists of terms corresponding to their energy and the term that describes their Coulomb interaction.

$$\widehat{H} = -\frac{\hbar^2}{2m_e}\nabla_e^2 - \frac{\hbar^2}{2m_h}\nabla_p^2 - \frac{e^2}{\left|\vec{r_e} - \vec{r_h}\right|} \tag{5.48}$$

By means of the usual procedure we can separate motion of exciton as whole form electron-hole relative motion. For this we change variables

$$\vec{R} = \frac{m_e\vec{r_e} + m_h\vec{r_h}}{m_e + m_h}, \; \vec{r} = \vec{r_e} - \vec{r_h} \tag{5.49}$$

In these variables (5.48) operator gets the following form

$$\widehat{H} = -\frac{\hbar^2}{2\left(m_e + m_h\right)}\nabla_R^2 - \frac{\hbar^2}{2\sim}\nabla_r^2 - \frac{e^2}{\left|\vec{r_e} - \vec{r_h}\right|}, \quad \sim = \frac{m_e m_h}{m_e + m_h} \tag{5.50}$$

The eigen functiona and eigen values of this operator are

$$\Psi_{K,n} = e^{i\vec{K}\vec{R}} F_n\left(\vec{r}\right)$$
$$E_{K,n} = E_n + \frac{K^2}{2\left(m_e + m_h\right)}, \ E_n = -\frac{1}{n^2}\left(\frac{\sim e^4}{\mathrm{v}_r{}^2 2\hbar^2}\right) \tag{5.51}$$

In quantum structures new terms those describe confinement of electrons and holes are added to (5.50) Hamiltonian. In most semiconductor materials, these terms influence the motion of electron and holes much stronger than Coulomb interaction between them. Therefore, usually first are calculated single-particle states of electrons and holes, and than interaction between them is accounted in different approximation. That is why in total Hamiltonian single particle terms are separated:

$$\widehat{H} = -\frac{\hbar^2}{2m_e}\nabla_e^2 - \frac{\hbar^2}{2m_h}\nabla_p^2 - \frac{e^2}{\left|\overrightarrow{r_e} - \overrightarrow{r_h}\right|} + V_{oe} + V_{oh} = H_e + H_h - \frac{e^2}{\left|\overrightarrow{r_e} - \overrightarrow{r_h}\right|} \tag{5.52}$$

The procedure in Hartree approximation is as follows:
First we solve single particle problem for holes

$$H_h \{_h^o = E_h^0 \{_h^o \tag{5.53}$$

Than Coulomb term is averaged by $\{_h^o$ functions, and obtain effective potential for electrons

$$V_{eff}^e\left(\overrightarrow{r_e}\right) = \left\langle \{_h^o \right| - \frac{e^2}{\left|\overrightarrow{r_e} - \overrightarrow{r_h}\right|} \left| \{_h^o \right\rangle \tag{5.54}$$

and solving single particle problem with Hamiltonian with this effective potential

$$H_e + V_{eff}^e\left(\overrightarrow{r_e}\right)\{_e^o = E_e^0 \{_e^o \tag{5.55}$$

We obtain electron wave functions in zero order of approximation.

After this we again average Coulomb term over obtained at the previous stage electron wave functions:

$$V_{eff}^{h}\left(\overrightarrow{r_h}\right)=\left\langle \{_{e}^{o}\right|-\frac{e^{2}}{\left|\overrightarrow{r_e}-\overrightarrow{r_h}\right|}\left|\{_{e}^{o}\right\rangle \tag{5.56}$$

then solve single particle problem for (5.56)

$$H_h+V_{eff}^{h}\left(\overrightarrow{r_h}\right)\{_{h}^{1}=E_h^1\{_{h}^{1} \tag{5.57}$$

This procedure is continuing until desired convergence is obtained.

Finally energy is expressed as

$$E=E_e+E_h-\left\langle \{_{e}^{n}\{_{h}^{n}\right|\frac{e^{2}}{\left|\overrightarrow{r_e}-\overrightarrow{r_h}\right|}\left|\{_{e}^{n}\{_{h}^{n}\right\rangle \tag{5.58}$$

Where $\{_{e(\mathrm{h})}^{n}$ are single particle wavefunction obtained at the *n*th stage.

Chapter 6 Superlattices

A superlattice is a semiconductor structure created in such a way that periodicity is imposed on the system during growth. This periodicity typically ranges from tens to thousands of angstroms, so that it includes at least a few periods of the natural crystal structure, but is small enough so that quantum effects are important. These are thus mesoscopic structures. A simple example is a compositional superlattice, consisting of periodically alternating plane layers of, say, AlGaAs (A) and GaAs (B) (Fig. 6.1). The electronic periodicity is provided by the alternation of the conduction and valence band edges, as shown in Fig. 6.2. A superlattice creates a new kind of electronic raw material. The fact that it is a grown structure means that there is great freedom in creating a material with new sorts of electronic properties

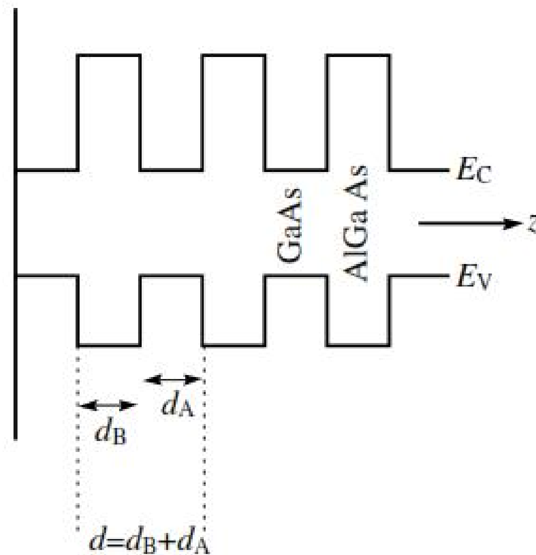


Figure 6.1 The alternating band edges provide a periodic array of quantum wells for electrons (E_c) and holes (E_v).

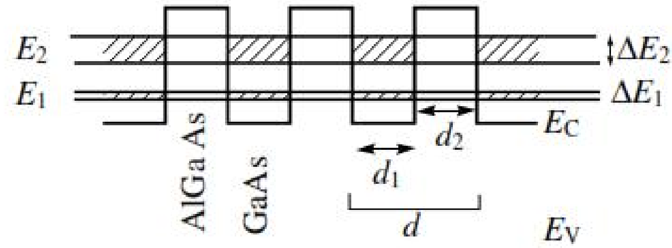


Figure 6.2 Band edges and miniband energies for a typical superlattice, with alternating layers of GaAs (wells) and $Al_{0.11}Ga_{0.89}As$ (barriers). For this case $d_1 = d_2 = 90$ A. Miniband energies are $E_1 = 26.6$ meV and $E_2 = 87$ meV and respective bandwidths are $E_1 = 2.3$ meV, $E_2 = 20.2$ meV.

6.1 Superlattices and Multi-quantum-wells

When a superlattice contains widely spaced quantum wells, so that electron tunneling from well to well is essentially prohibited, one can treat the array as a set of isolated quantum wells. The energies and wavefunctions of electrons in each well will then be determined just by the properties of an individual well. Such arrays (multi-quantum-wells) are often used to enhance the signal obtainable from a single well. A true superlattice is a similar system, but with thinner barriers (more closely-spaced wells), so that there is electron tunneling, and therefore good communication, from well to well. Some typical parameters are shown in Fig. 6.2.

There are two different ways of looking at a superlattice which are illuminating, and taken together give a good picture of the system. The first of these is the picture of a superlattice as a single bulk crystal with an additional modulation (periodicity) imposed on it. The second is the picture of a collection of equally- spaced quantum wells which are brought progressively closer together.

Crystal periodicity leads to the electron band structure observed in bulk crystals. Superlattice periodicity likewise gives a band structure, for the same reasons. However, since the superlattice spacing is greater than the crystal spacing, the superlattice k-space dimensions will be smaller than those of the crystal. This new band structure will be superimposed on the original bulk band structure, and will show up as a series of minibands and minigaps which will be superposed on the original band structure of the well material. These minibands and gaps result from zone folding: for a superlattice in

superlattice constant in this direction which is M times the atomic one. Associated with this new periodicity, there will be a new superlattice Brillouin zone with a size of $1/M$ times the crystal Brillouin zone. It is often the case that $1/M \ll 1$, so that many of the new mini-zones will fit into the original Brillouin zone. Thus, the band structure shows its new periodicity by breaking up into minibands and minigaps whose scale is determined by the size of the superlattice layers. Figure 6.3 illustrates this situation near the conduction band minimum of the well material, where the band structure can be taken to be parabolic to a good approximation.

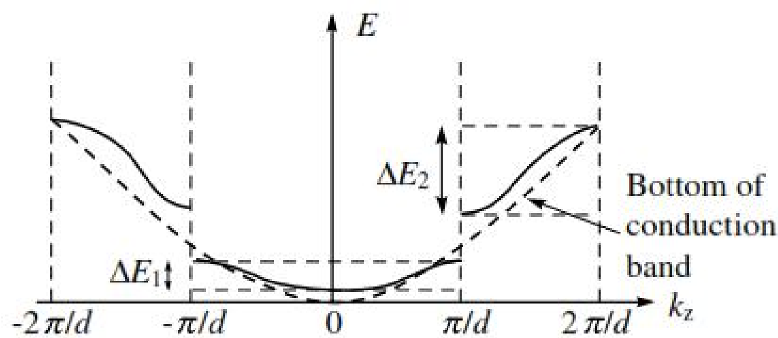


Figure 6.3 Formation of minibands and minigaps near the conduction-band minimum of a direct-gap semiconductor. The minizone width *2r/d* is typically much less than the Brillouin zone width *2r/a,* where a is a lattice constant.

One can also consider superlattices as collections of identical, isolated quantum wells which are brought closer together in such a way that they remain separated by equal distances. Each single well has its own set of discrete energy levels from electron confinement in the z-direction. If, for instance, one had started with only two such wells separated by a very large distance, there would be some common energy E_1, say, which an electron could have by being in one well or in the other (a two-fold degeneracy). As these quantum wells are brought closer together, interaction between the wells becomes possible, so that the levels are no longer degenerate, but have energies (E_1 +) and (E_1 —), where increases from zero as the barrier width decreases. In place of a single degenerate level, one now has two levels, slightly split. The communication between wells causing this splitting comes about because of the fact that an electron confined in one well can really be present in the barrier region as well, with a small probability. If the second well is near enough, the electron can also penetrate

eigenstates of a two-well problem. In each, the electron has equal probability of being in either well, and the wavefunction is symmetric or anti- symmetric, in the state with energy (E_1 +) or (E_1 —), respectively.

Since the energy splitting is bigger if the amount of communication between wells is greater, one finds that the splitting of a higher-energy level $E_2 > E_l$ will be greater than that of a lower level. Higher-energy states have a higher probability of being present in the barrier regions: they have longer tails and thus can 'see' the presence of other wells more effectively. This in turn follows from the fact that, since the energy of such states is higher, the effective barrier $V — E$ through which they have to tunnel is lower.

Bringing many (N) identical wells together has a similar effect. In this case, a single-well level E_l will be N-fold degenerate when the wells are far apart. As they are brought closer together in a uniform fashion, this degenerate level will split up into a set of closely spaced levels (N of them). This set of levels can be thought of as a continuum - a miniband. One will have other minibands, corresponding to each of the original levels of the single well. As in the two-well case, one expects higher-energy minibands to have a greater bandwidth than lower-energy ones. This can be a dramatic effect, as indicated in Fig. 6.2. One can say that level broadening increases as tunneling becomes more effective.

Miniband broadening is also indicated in Fig. 6.4, which shows the effect on the density of states, for a set of square wells (as in Fig. 6.2) brought close together into a superlattice structure. Note the minibands (*a*—*b*), (*c*—*d*) and mini- gaps (*b*—*c*), and the increase of the bandwidth with miniband energy in Fig. 6.4.

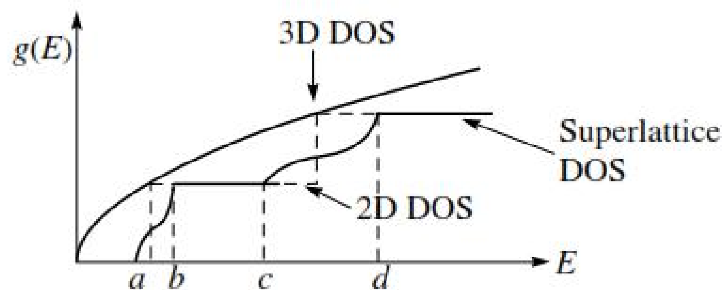


Figure 6.4 Superlattice density of states (DOS) in relation to that of a 3DEG and of a2DEG in a square quantum well.

Miniband Properties: The WKB Approximation

The WKB model can give good physical estimates for superlattice properties, if the system has relatively thick barriers. It also tends to give better than expected results for more general systems.

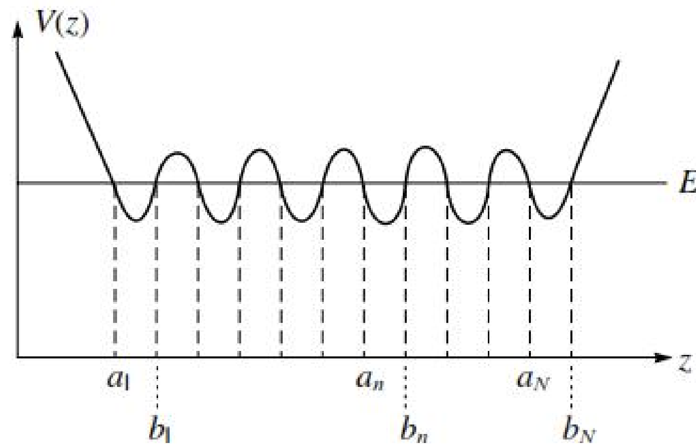


Figure 6.5 Model of a superlattice potential, with N equally-spaced, equivalent quantum wells

Consider the *N*-well system shown in Fig. 6.5. When one treats this problem in the WKB approximation, certain basic ingredients emerge:

$$\dagger = \int_{a_n}^{b_n} k(z)\,dz \qquad (6.1)$$

(an integral over the allowed energy region in, say, the nth well), with

$$k = \frac{1}{\hbar}\left\{2m^*\left[E - V(z)\right]\right\}^{1/2} \qquad (6.2)$$

and

$$\ddagger = \int_{b_n}^{a_{n+1}} |\,(z)\,dz \qquad (6.3)$$

(an integral over a single forbidden barrier region), with

$$\frac{1}{\ }\{2 \ \ *[V(\ )\ \ E]\}^{1/2} \qquad (6.4)$$

Here, we want to describe the motion of an electron with energy E in a system with, say, the classical turning points a_n and b_n. As always, the confinement, and the requirement that the wavefunctions match at the classical turning points, results in quantization conditions, which may be written

$$† = \left(n + \frac{1}{2}\right)ƒ + e^{-‡} \cos\left(\frac{mƒ}{N+1}\right) \tag{6.5}$$

where $n = 1, 2, \ldots$ and $m = 1, 2, \ldots \mathrm{N}$ The first term on the right-hand side of (6.5) is the WKB approximation to the energy levels (labeled n) of an isolated well. The second term describes how these single levels split into N sublevels (labeled m), thus forming a miniband. This term is small because of the exponential in $-‡$, so that the sublevels are closely spaced. Assuming such a small splitting of levels, one obtains from (6.5) the quantized energies

$$E_n = E_n^0 + \frac{\hbar Š}{ƒ} e^{-‡^{(0)}} \cos\left(\frac{mƒ}{N+1}\right) \tag{6.6}$$

Here, E_n^0 is the zero-order energy (coming from the first term on the right in (6.5)) and $‡^{(0)}$ is calculated using the approximate energy E_n^0 in the definition (6.3). Finally, S, the frequency of classical motion in a single quantum well, is defined by

$$\frac{ƒ}{Š} = \frac{m^*}{\hbar} \int_{a_n}^{b_n} \frac{dz}{k^{(0)}(z)} \tag{6.7}$$

where $k^{(0)}$ is calculated from (6.2) using the approximate energy E_n^0.

These results give, among other things, the WKB approximation for the mini-band width, which may be read off from (6.6). Since $2\cos(ƒ/N+1)$ is the maximum difference between the highest and lowest values of $\cos(mƒ/N+1)$, when $m = 1, 2, \ldots \mathrm{N}$ we obtain the bandwidth

$$\Delta E_n = \frac{\hbar Š}{ƒ} e^{-‡} \left[2\cos(ƒ/N+1)\right] \tag{6.8}$$

From this equation one can see how the bandwidth depends upon the single-well energy level n from which that miniband arises. This dependence enters into the parameters S and ‡ of equation (6.8). The latter is by far the most important dependence, because it appears as an argument of an exponential. Explicitly, one has that

$$\ddagger^{(0)} = \frac{1}{\hbar}\int_{b}^{b}\left\{2m^{*}\left[V\left(z\right)-E\right]\right\}^{1/2} dz \qquad (6.9)$$

It is evident that $\ddagger^{(0)}$ will be smaller (and the bandwidth greater) when E_n is greater (i.e. nearer to V). In this case the effective barrier $V — E$ is lower, and thus the integrand in (6.9) is smaller. As shown in Fig. 6.6, a second effect usually operates to increase the bandwidth with increasing *n.* The higher-energy states see a barrier which is also thinner, since at higher energies the limits *b* and *a* are often closer together. Communication between wells is of course aided by thinner barriers.

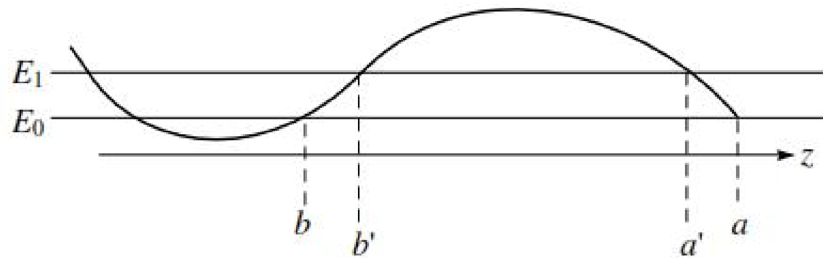


Figure 6. 6 A single well and barrier, for the superlattice of Fig. 6.5. E_O and E_1 are unperturbed WKB energies for a single well with infinitely thick barriers.

6.2 Doping Superlattices

From what has already been said, it should be evident that a superlattice can be created by imposing a new periodicity of any sort on a semiconductor. One such example is the doping superlattice, first proposed by Dohler (1972) (see also Ruden and Dohler, 1983). The idea is to introduce a new periodicity into a semiconductor by doping it selectively, first by acceptors, then by donors, in repeated plane layers, as it is grown. The ionized impurities create repeated layers of negative (*n*) and positive (*p*) space charge in the conductor (Fig. 6.7), often separated by intrinsic (*i*), or undoped, layers. The periodically repeating electric fields which result create a superlattice in the semiconductor which, for obvious reasons, is called a *n-i-p-i* superlattice.

Figure 6.8 shows how such fields create a superlattice. The electric field of the ionized

previous chapter, while a parabolic well for holes (an inverted parabola) is created in the p-type regions.

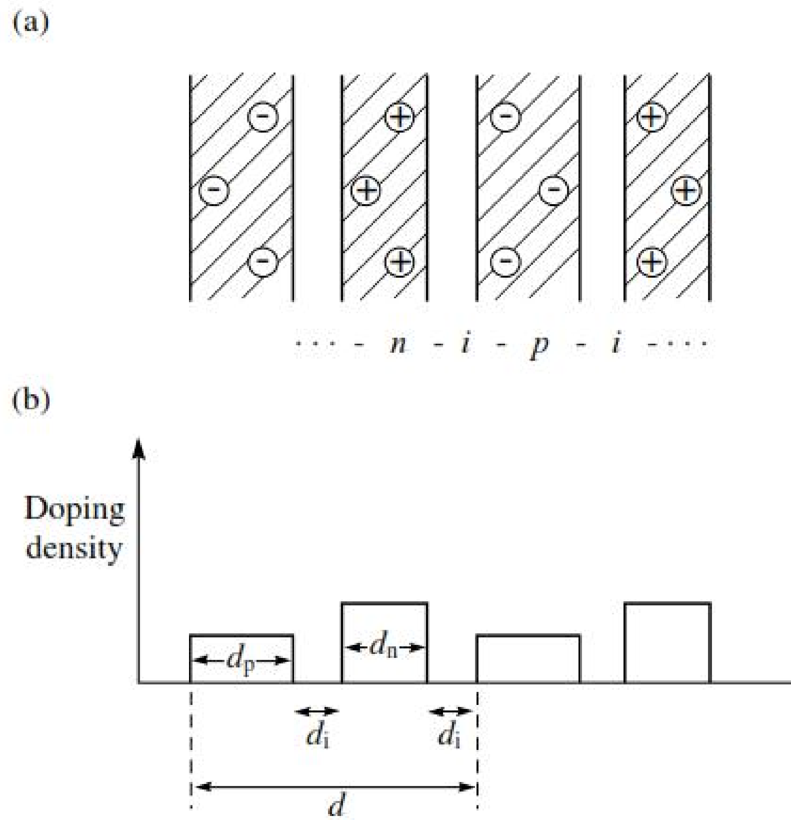


Figure 6.7 (a) Ionized impurities introduced in alternate layers in a semiconductor. (b) Possible doping profile leading to the space charge shown in (a)

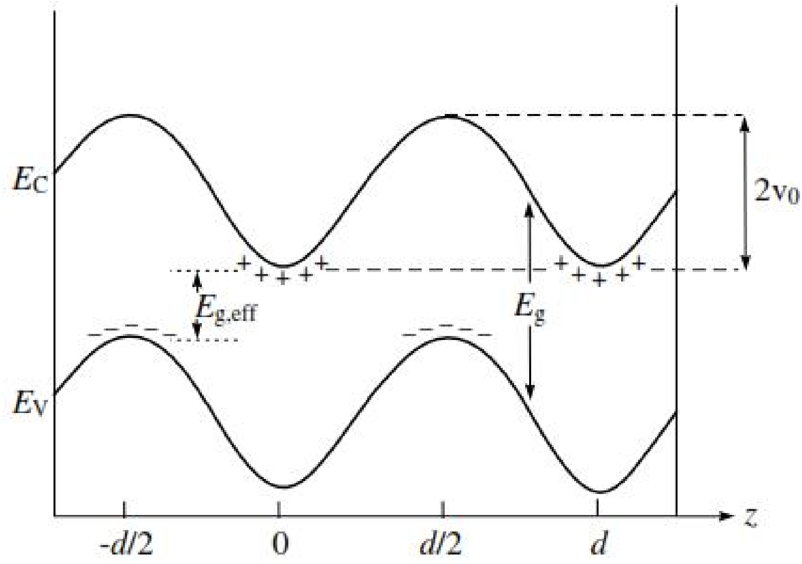


Figure 6.8 Conduction and valence band edges in a semiconductor with n-i-p-i-type doping. $E_{g,eff}$ is the effective band gap.

The charge density of the impurities may be written as

$$\rho_0(z) = e\left[n_D(z) - n_A(z)\right] \tag{6.10}$$

where n_D is the number of donors per unit volume, n_A is the number of acceptors per unit volume and z is the growth direction. The band-edge potential $V_0(z)$ is easily calculated from Poisson's equation:

$$V_0(z) = \left(\frac{e^2}{2\varepsilon_r\varepsilon_0} n_D\right) z^2 \tag{6.11}$$

for $0 \le |z| \le \frac{1}{2} d_n$, and

$$V_0(z) = 2v_0 - \left(\frac{e^2}{2\varepsilon_r\varepsilon_0} n_A\right)\left(\frac{1}{2} d - |z|\right)^2 \tag{6.12}$$

for $\frac{1}{2} d - |z| \le |z| \le \frac{1}{2} d_p$, where we have set $V = 0$ at the origin, which is taken to be the middle of an *n*-type layer, and spacings d_p etc. are defined in Fig. 6.7. In the intrinsic regions, where there are no ionized donors, the potential must be a linear function of z. We assume that this basic pattern is repeated indefinitely. The quantity $2v_0$ in equation

conduction band edge must be continuous (as shown), the parabolic regions in (6.7) and (6.8) must be joined by straight lines. This determines the depth of the modulation:

$$2v_0 = \frac{e^2}{2\mathrm{V}_r\mathrm{V}_0}\left[n_D\left(\frac{1}{2}d_n\right)^2 + n_A\left(\frac{1}{2}d_p\right)^2 + n_D d_n d_i\right] \tag{6.13}$$

Note that if the total number of ionized donors in one *n*-layer equals the number of ionized acceptors in a *p*-layer, the electrons that have been released by the donors will all reside on acceptors if the system is in its ground state. Thus, the potential that a free electron would feel is indeed given correctly by equations (6.11)—(6.13). There will be no free electrons in the wells to modify this potential with self-consistent effects.

The WKB approximation can be applied to the parabolic parts of *n-i-p-i* quantum wells to predict the superlattice energies. In particular, the miniband splittings and miniband widths can be calculated, as described above. The unperturbed energies (those for isolated wells), in particular, are given by

$$E_n^{(0)} = \left(n + \frac{1}{2}\right)\hbar \check{S} \tag{6.14}$$

where, as before, S given by (6.7), the miniband splittings by (6.6), and the miniband widths by (6.8). The calculation of these quantities is straightforward, but since the resulting expressions are rather unwieldy, they will not be reproduced here.

6.3 Delta-doped n-i-p-i s

A delta-doped n-i-p-i is essentially one in which the dopant layers are each only one atom wide. The resulting band-edge diagram is shown in Fig. 6.9, for *n* and *p* layers with equal numbers of ionized impurities. In this case, the conduction band edge is given by

$$V_0 = \frac{e^2 D}{2\mathrm{V}_r\mathrm{V}_0}|z|, \quad |z| \le \frac{1}{2}d, \text{....and repeating} \tag{6.15}$$

where *D* is the number of ionized donors (or acceptors) per unit are, and the superlattice amplitude is given by

$$2v_0 = \frac{1}{2} d \left(\frac{e^2}{2\text{V}_r \text{V}_0} \right) \tag{6.16}$$

The WKB approximation can be used here, too, to obtain predictions for the properties of this superlattice, in the way already described; in particular the miniband parameters $\text{Š}^{(0)}$ and $‡^{(0)}$ are obtained in an analogous manner.

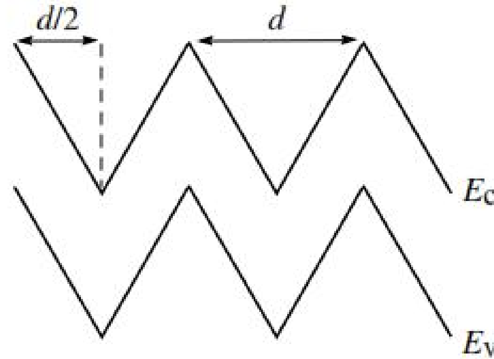


Figure 6.9 Conduction and valence band edges for a *n-i-p-i* superlattice with equally- *spaced n* and *p* delta layers of equal strength.

6.4 Compositional and Doping Superlattices

For electronic devices, one of the most important semiconductor parameters is the value of the energy gap between the valence band maximum and the conduction band minimum. Many applications, particularly those that depend on optical properties, depend crucially on the value of the fundamental gap. One would like a wide range of effective band gaps, $E_{g,eff}$, for device use. The fundamental gap E_g is indeed modified in compositional quantum wells and superlattices, as is shown in Fig. 6.10. Once such a system is grown, its effective gap will be greater than E_g , and will moreover be fixed.

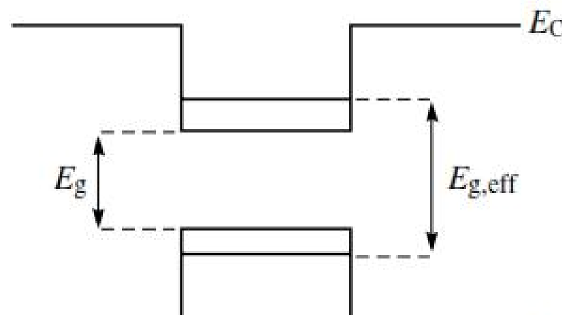


Figure 6.10 Mo dification of the fundamental gap in an AlGaAs-GaAs quantum well.

To obtain effective band gaps smaller than E_g, one must turn to something like *a n-i-p-i* system. Fig. 6.11 shows how different effective band gaps can be obtained using the same host material, with the same periodicity, but with different doping densities. From equation (6.16), the superlattice modulation is proportional to the product D_d. A small value of D results in a weak modulation of the band edge (Fig. 3.35(a)), and gives an effective gap not much smaller than that for the bulk semiconductor. The effective gap decreases, however, as doping strength D increases (b); and at high doping, as in (c), a semi-metal can result. The same principles hold for other doping superlattices.

Figure 6.11 shows the case of a doping superlattice in which there is complete compensation: the numbers of donors and acceptors are equal. Since the system is assumed to be in its ground state, all donated electrons reside on acceptors. Once such a system is grown, however, there is also freedom to tune it. Figure 6.12 shows such a

many-subband doping superlattice in its ground state, with no free carriers.

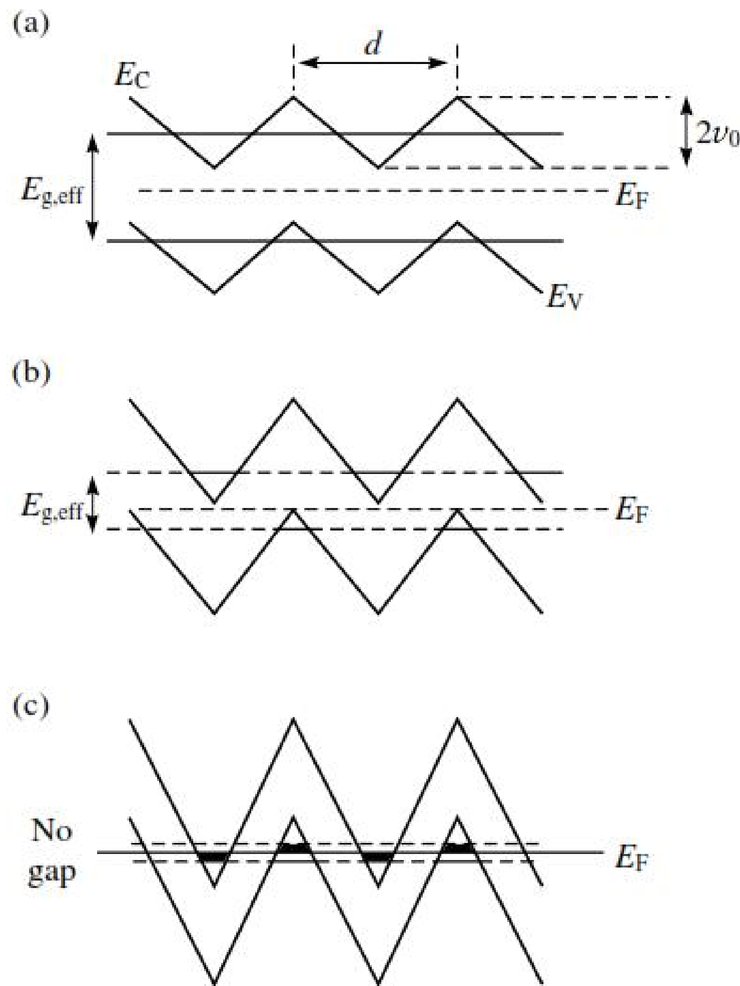


Figure 6. 11 Delta doping superlattices with (a) weak, (b) moderate, and (c) strong dop- ing densities.

Electrons and holes are confined to separate regions of the superlattice (this is sometimes said to be a system with an "indirect gap in real space"). One can now excite this system, for instance with light of the right frequency, to create electron-hole pairs. Free carriers have now been introduced into the system. Electrons will be attracted to the quantum wells in the *n*-layers, and holes to the *p*-layers. Since these regions are well separated in space, however, the electron and hole wavefunctions will be well separated. This small overlap means that electron—hole recombination will be very slow. Thus there will be an appreciable time during which these free carriers will reside in their respective wells, where they will screen the space charge already present. The net effect will be to reduce the superlattice modulation, and thus increase the effective band gap (Fig. 6.13).

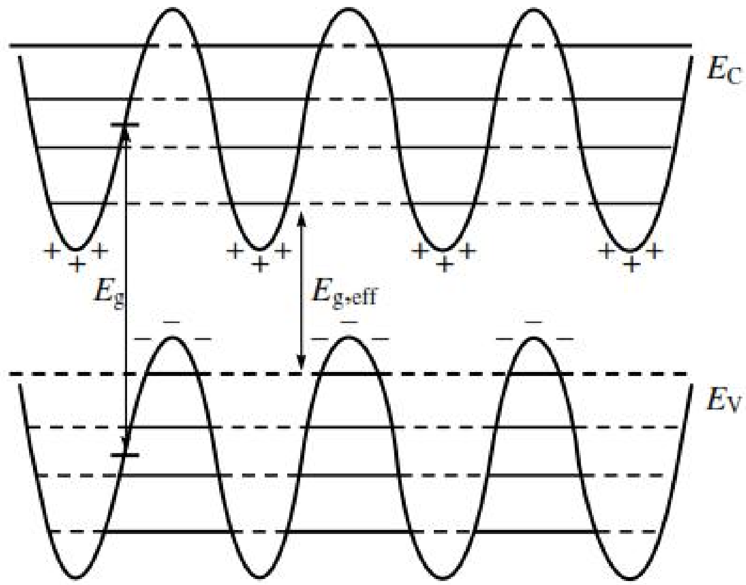


Figure 6. 12 A doping superlattice in its ground state.

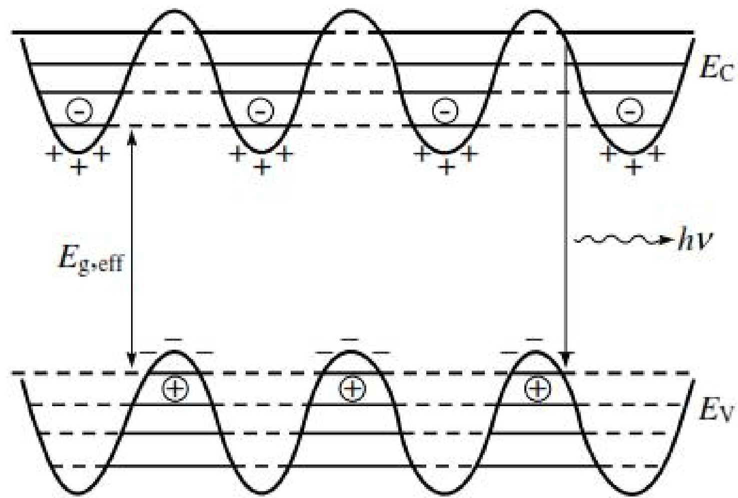


Figure 6. 13 A doping superlattice in its excited state.

6.5 Other Types ofSuperlattices

Compositional superlattices of the GaAs/AlGaAs type, with low Al composi- tions, are called Type I superlattices. Materials such as the InAs-GaSb system offer a

different possibility: the Type II superlattice. In Type II superlattices, electrons are confined to one material and holes to the other (Fig. 6.14).

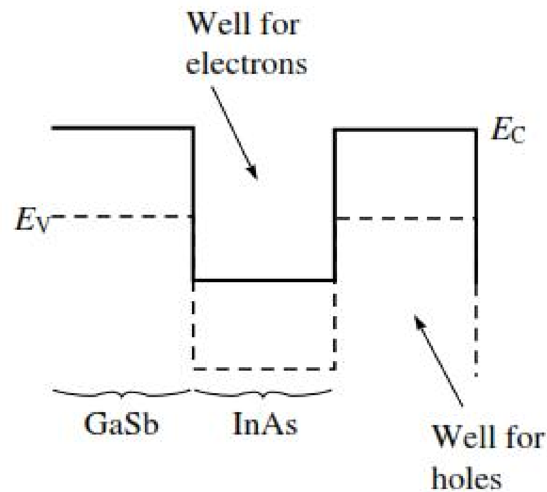


Figure 6. 14 Schematic band-edge diagram for a Type II superlattice.

Strained-layer superlattices are compositional superlattices in which the constituent materials are not perfectly lattice-matched to each other. If the materials are not too different, and if the layers of each material are not too thick, good growth of one material on another, in layers, is still possible. A given layer will then be compressed, or extended, in the plane perpendicular to the growth direction, by atomic forces arising from the layer onto which it is trying to grow. The effects of strain can also be useful. It breaks the degeneracy of hole states, and changes the band structure in other ways, all of which offer new possibilities for device development.

Any way of imposing an artificial periodicity on a semiconductor system can in principle make a superlattice. The periodicity does not need to be imposed on the growth direction of a low-dimensional system. One superlattice of interest, for instance, is made by shining a laser interference pattern onto an already-created 2DEG. Extra electrons are generated (for instance in the surrounding medium) by the light, in a periodic way, thus creating a periodic electric field which modulates the 2DEG in its own plane. A similar way of modulating an existing 2DEG is with the help of acoustic waves. Here, the modulation is caused by the periodic electric field which is generated by the piezoelectric effect.

Still another kind of superlattice can be created using a regular array of closely spaced quantum dots. In this case, an artificial periodicity will be created in at least two different directions.

Chapter 7 Phonons

One of the central themes of this book is how reducing the size of semiconductor structures down to mesoscopic and smaller scales brings the quantum wave nature of electrons into play, resulting in electronic and optical properties which are markedly different from those of bulk semiconductors. One of the key challenges facing physicists and engineers is how to make devices operate at room temperature. The main obstacle to achieving this goal is the unavoidable presence of phonons, the quantum vibrations of atoms making up a solid, and their ability to scatter electrons.

Phonons would thus appear to occupy an uncomfortable position in the study of low-dimensional structures, with an understanding of their properties required solely for the purpose of finding ways to reduce their interaction with electrons. As we shall see in this chapter, however, the physics of phonons in low-dimensional structures is sufficiently fundamental and non-trivial to be of interest in its own right. Just as for electrons, phonons can be confined within heterostructures and we would like to know how the dynamics of low-dimensional phonons differs from that of bulk phonons. We would also like to understand the effects of dimensionality on the electron-phonon interaction and hence such electron transport properties as the phonon-scattering-limited mobility.

In fact, finding ways to reduce the electron-phonon interaction is not the only reason for investigating phonons in low-dimensional structures. Being the vibrations of atoms making up a structure and given their interaction with electrons, phonons have proved to be an effective probe of the electronic and structural properties of low-dimensional semiconductors. Two such probe techniques are acoustic phonon pulse spectroscopy and Raman spectroscopy. Phonon pulse spectroscopy relies on the fact that acoustic phonons with wavelengths on the order of several hundred angstroms can travel for relatively large distances in semiconductor crystals without scattering. By detecting the flux of acoustic phonons emitted from an electron gas heated above the lattice temperature or, alternatively, measuring the response (e.g. conductance) of an electron gas due to the interaction with an incident beam of acoustic phonons, information about the electron gas can be obtained.

For example, phonon pulse spectroscopy can be used to study the electron current distribution of a two-dimensional electron gas and more complicated electron states. Acoustic pulse methods have also been used to study the materials properties of semiconductors, such as surface and interface roughness. Raman spectroscopy has proven to be an extremely powerful probe of the dynamics of phonons in heterostructures. This method takes advantage of the fact that energy conservation requires a photon which absorbs or emits a phonon to suffer a change in its frequency. This frequency change is called the Raman shift. By measuring the intensity versus Raman shift of laser light which is inelastically scattered from a solid, the energies and, hence, the frequencies of the allowed phonon modes of the solid can be measured. Since the phonon frequencies are determined by the interatomic forces and atom masses that comprise a material, Raman spectroscopy can be used to infer various materials properties of heterostructures.

However, before going on to study the various phonon spectroscopies and their interpretation in terms of heterostructure properties, we should first learn about the basic physics of phonons and electron-phonon interactions in heterostructures. This is the purpose of the present chapter.

7.1 Phonons in Heterostructures

In the classical approximation, the vibrations of the atoms making up a crystalline solid are most conveniently described by the function $\vec{u}_j\left(\vec{R},t\right)$ which gives the displacement from the equilibrium position at time instant t of the j-th basis atom in the unit cell located at $\vec{R} = n_1\vec{a}_1 + n_2\vec{a}_2 + n_3\vec{a}_3$. The displacement is specified by its magnitude and direction and thus is a vector. The displacement of a given atom as a function of time will be governed by its interaction with all the other atoms making up the solid. Because of the non-linear nature of the full equations of motion and the very large number of atoms involved, it is in practice impossible to find solutions for the atom displacements without first making some simplifying approximations. The most common approximation is to expand the interaction potential energy in the displacements and keep only terms to quadratic order. (Note that linear order terms vanish since the displacements are defined with respect to the equilibrium positions of the atoms.) The

equations of motion are then linear in the displacements and we have the important result that any solution can be expressed as a sum of solutions with the property that all atoms vibrate at the same frequency:

$$\vec{u}_j\left(\vec{R},t\right) = \mathrm{Re}\left[\sum_{\alpha} c_{\alpha}\, e^{i\omega_{\alpha} t}\, \vec{u}_{j\alpha}\left(\vec{R}\right)\right] \tag{7.1}$$

Because of the latter property, this is called the harmonic approximation.

Thus, to understand the classical dynamics of atomic vibrations in a crystalline solid, it is sufficient to find the single frequency solutions $e^{i\omega_{\alpha} t}\vec{u}_{j\alpha}\left(\vec{R}\right)$, the so-called normal modes. The quantum dynamics and interpretation in terms of vibration quanta — phonons — then follow directly. For example, the total energy of a crystal in a given quantum state is

$$E = \sum_{\alpha} \hbar\omega_{\alpha}\left(N_{\alpha} + \frac{1}{2}\right) \tag{7.2}$$

where N_{α} is the number of phonons present of mode-type α .

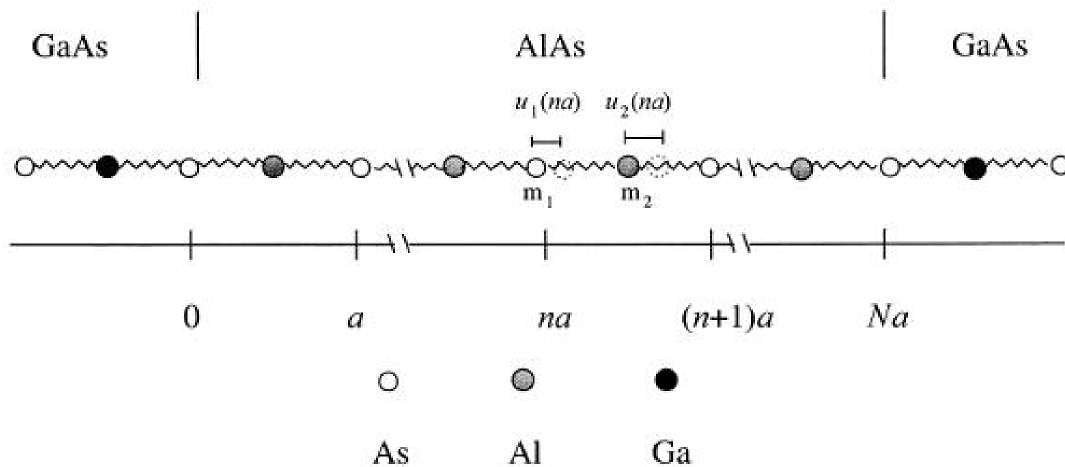
This picture of the dynamics, involving the concepts of classical normal modes and quantum phonons, is valid provided the higher order non-quadratic terms in the potential energy expansion - the anharmonic terms - are much smaller than the quadratic terms. Physically, this means that changes in the distances between neighbouring atoms due to their vibrations are small compared with their equilibrium separations. This is the case as long as we are well below melting temperatures, as we indeed are for the low-dimensional structure physics and applications discussed in this book. However, that the presence of even small anharmonic terms will mean that these phonon modes have only a finite lifetime.

7.2 Superlattices

To gain some idea of the nature of phonon modes occurring in heterostructures, we shall focus our discussion on the modes of a special type of heterostructure - the superlattice.

As an illustrative example, we consider a superlattice formed from alternating GaAs and AlAs slabs. Both GaAs and AlAs crystals have zincblende structure and thus a superlattice grown on an (001) face of a GaAs crystal substrate will comprise alternating layers of atoms of a single species with a monolayer spacing equal to half the lattice constant. The lattice constant of GaAs is 5.65 A, while that of AlAs is 5.62 A, so that the respective monolayer spacings are 2.83 A and 2.81 A.

One class of modes which occurs in such a superlattice involves the vibration of entire (001) planes of atoms in the direction normal to the planes, i.e. the [001] direction. These are called longitudinal modes. Because the atoms in a given monolayer are all moving in unison, we can model the superlattice by using a one-dimensional system. Further simplifications follow if we make the harmonic approximation and include only nearest-neighbour interactions between atoms.



. Figure 7.1. Linear chain model for [001] longitudinal modes in a GaAs/AlAs superlattice. The atoms are shown in their equilibrium positions.

The resulting dynamics is identical to that of a mechanical system consisting of a chain of masses connected with springs (Fig. 7.1). As we shall now see, this linear chain model gives a reasonably good description of the [001] longitudinal modes.

In the GaAs or AlAs slabs, the equations of motion take the form

$$m_1 \frac{d^2u_1}{dt^2}(na,t) = -| \left\{2u_1(na,t) - u_2(na,t) - u_2\left[(n-1)at\right]\right\} \quad (7.3)$$

$$m_2 \frac{d^2u_2}{dt^2}(na,t) = -| \left\{2u_2(na,t) - u_1(na,t) - u_1\left[(n+1)at\right]\right\} \quad (7.4)$$

where a is the monolayer spacing (i.e. the unit cell length), | is the force constant and $u_j(na,t)$ denotes the displacement from equilibrium along the line of the chain at time t of one of the two atoms of mass m_1 and m_2 in the unit cell located at na. To obtain solutions for the entire chain, these equations must be supplemented by certain conditions on the displacements at the boundaries between GaAs and AlAs slabs. However, before we consider these full solutions, let us first address the simpler problem of a single slab with periodic boundary conditions. Substituting the trial mode solution $u_j(na,t) = A_j \exp\left[-i(\text{Š}t - qna)\right]$, $q = 2f/\}$, into equations (7.3) and (7.4), we find that we can have a non-trivial solution (i.e. $A_j \neq 0$) only if the determinant of the coefficients of the two unknowns A_1 and A_2 vanishes. This results in the following relation between the angular frequency S and wavevector q:

$$\text{Š}^2 = \frac{|}{m_1 m_2}\left\{m_1 + m_2 \pm \left[m_1^2 + m_2^2 + 2m_1m_2\cos(qa)\right]^{1/2}\right\} \quad (7.5)$$

The mode angular frequency dependence on wavevector is called the dispersion relation and the two roots in (6.5) are the branches of the dispersion relation. Before we can display the dispersion relations for GaAs and AlAs, we must determine their force constants. This can be done, for example, by measuring S at $q = 0$ using Raman spectroscopy and then fitting the positive root at $q = 0$ in (7.5) to the measured value of S. For GaAs, this gives the value | = 90.7 N/m, while for AlAs we obtain | = 95.4 N/m. In Fig. 7.2 we show the resulting dispersion curves for GaAs and AlAs. In fact, they are a set of discrete points at $q = \pm 2f n/Na$, for n = 0, 1, 2, ... , where N is the number of monolayers. Altogether there are $2N$ distinct modes, N for each branch. Following the usual convention, the angular frequency has been divided by the factor $2fc$, where c is the speed of light in vacuum (3 x 10_{10} cm s^{-1}). This has the advantage that the Raman shift wave-number corresponding to a given phonon

mode can be inferred directly from the plot in Fig. 7.2. To convert to electron-Volts (1eV = 1.6 x 10^{-19} J), the wave-number value is multiplied by the factor 1.24 x 10^{-4} eV cm. For a mode in a GaAs or AlAs positive root branch, we see that the Raman shift will be in the infra-red range. For this reason, the positive root branch is called the optical branch. The negative root branch extends down to zero frequency as q goes to zero and thus is called the acoustic branch.

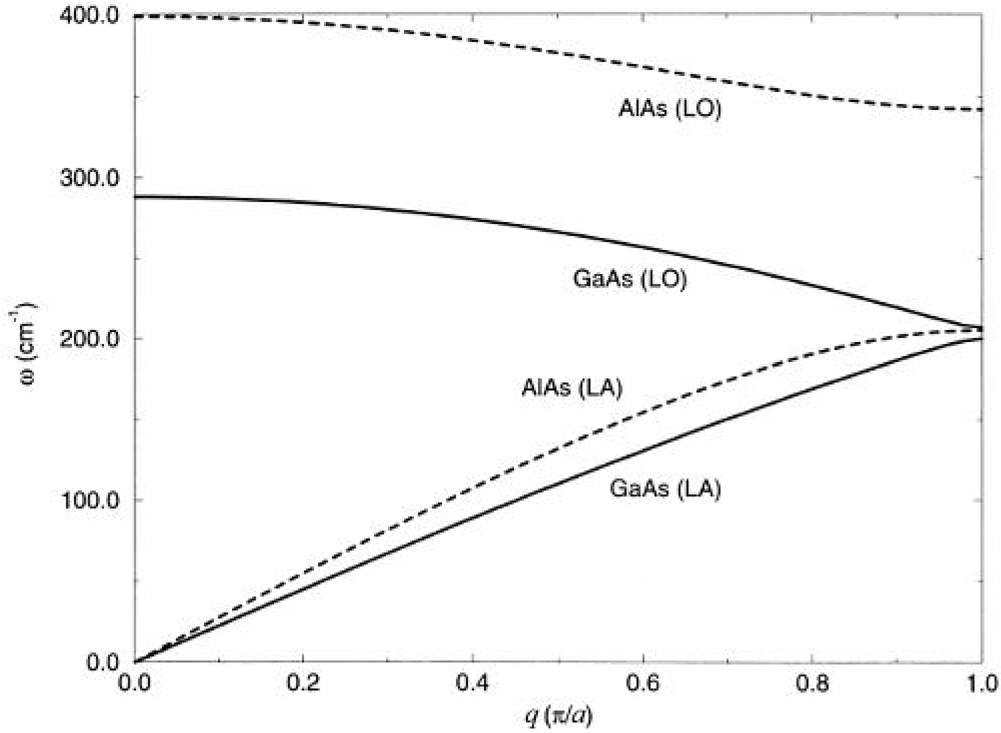


Figure 7.2 Longitudinal [001] acoustic (LA) and optical (LO) phonon dispersion curves for GaAs (solid line) and AlAs (dashed line).

If we expand the dispersion relation (7.5) with respect to q and keep only the lowest-order non-vanishing terms, we find for the optical branch:

$$\check{S} = \left[\frac{2| \left(m_1 + m_2 \right)}{m_1 m_2} \right]^{1/2} \tag{7.6}$$

and for the acoustic branch:

$$\check{S} = qa\left[\frac{|}{2\left(m_1 + m_2\right)}\right]^{1/2} \tag{7.7}$$

Thus, as q approaches zero, the optical group velocity $v_g = d\check{S}/dq$ vanishes, while the acoustic group velocity becomes constant and non-zero, coinciding with the acoustic phase velocity $s = \check{S}/q$. To determine how the atoms move for small q, we require the remaining part of the solution to equations of motion (4.3) and (4.4), namely the relation between the constants A_1 and A_2:

$$A_2 \Big/ A_1 = -\frac{|\left(1 + e^{iqa}\right)}{m_2\check{S}^2 - 2|} \tag{7.8}$$

Substituting expression (7.6) for S into (7.8), we find that for q tending to zero, the optical modes approach a standing wave with the Ga and As (or Al and As) atoms vibrating 180° out of phase: $A_2/A_1 = - m_1/m_2$. For the acoustic modes all the atoms vibrate in phase.

Substituting the force constant values given above into (7.7), we have a prediction for the small-q group velocity of longitudinal acoustic (LA) phonons propagating in the [001] direction. For example, for GaAs we obtain the value 4040 m s^{-1}, whereas the actual value is about 4770 m s^{-1} and therefore the linear chain model prediction is out by about 15%. This is quite reasonable given the simplicity of the model. A more accurate linear chain model will take into account interactions between next-nearest neighbor atoms as well, with the new unknown force constants determined by fitting to additional measured frequencies. The existence of a non-negligible acoustic group velocity gives rise in heterostructures to a range of acoustic phonon transport phenomena and, as mentioned, makes possible the field of acoustic phonon pulse spectroscopy.

Our discussion so far has been about the longitudinal [001] modes of a single GaAs or AlAs slab. Let us now consider the longitudinal [001] modes of the full GaAs/AlAs superlattice structure. An initial, basic understanding of the superlattice modes can be gained just by comparing the single-slab GaAs and AlAs dispersion relations (Fig. 7.2). The acoustic branches of AlAs and GaAs overlap in frequency and therefore the acoustic modes will extend throughout the entire superlattice. On the other hand, the AlAs and GaAs optical branches do not overlap in frequency - a consequence of the large difference between the Ga and Al atom masses -

and hence we expect the superlattice optical modes will be confined either to the AlAs or GaAs slabs. An estimate of the extent to which an optical mode is confined to a given slab can be obtained by considering the imaginary wavevector solutions to the single-slab equations of motion (7.3) and (7.4). We find that optical modes confined to the GaAs slabs extend between about 1 to about 0.3 to 0.4 monolayers into the GaAs slabs, where the lower and upper limits are for zone centre (q=0) and zone boundary ($q = \pm f/a$) modes, respectively. Thus the superlattice optical modes are well-confined.

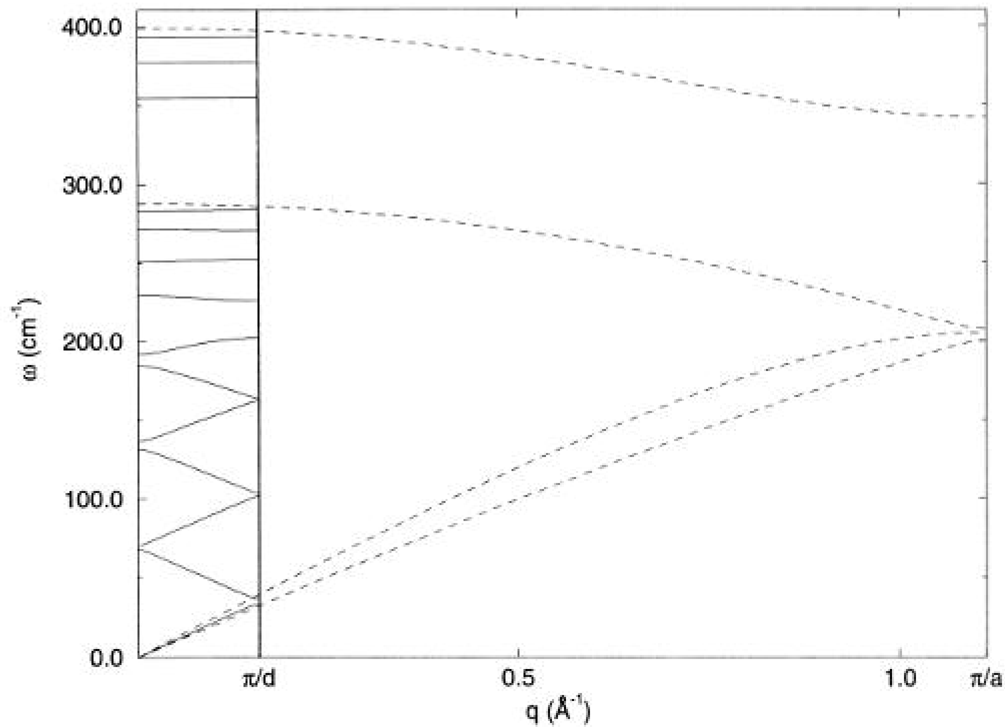


Figure 4.3. Dispersion curve for [001] longitudinal phonons of a superlattice comprising alternating 4-monolayer slabs of GaAs and 3-monolayer slabs of AlAs. Also shown for comparison are the dispersion curves for bulk GaAs and AlAs (dashed lines).

The superlattice dispersion relation and mode solutions are determined by first constructing solutions in a single superlattice unit cell (which comprises an AlAs and GaAs slab) and then using Bloch's theorem to extend the solutions throughout the whole superlattice. In Fig. 7.3,

we show the longitudinal [001] dispersion relation for a superlattice with 4-monolayer GaAs slabs and 3-monolayer AlAs slabs — a (4, 3) superlattice. Also shown for comparison are the single-slab dispersion relations. A single slab with periodic boundary conditions has period a, the monolayer spacing, while an (M, N) superlattice has a larger period $d = (M + N)\, a$ (where we regard the AlAs and GaAs monolayer spacings to be the same). Thus, the superlattice Brillouin zone (boundaries at $q = \pm f\ /d$) is smaller than the slab Brillouin zone (boundaries at $q = \pm f\ /$a). Notice that the part of the dispersion curve below about 200 cm^{-1} can be approximately obtained by "folding" the bulk acoustic phonon branches into the smaller superlattice zone. For this reason, this part of the dispersion curve is called the folded acoustic phonon branch. A Raman spectrum of a superlattice is shown in Fig. 7.4, which gives clear evidence of folded LA phonons in the form of double peaks occurring at the predicted frequencies.

.

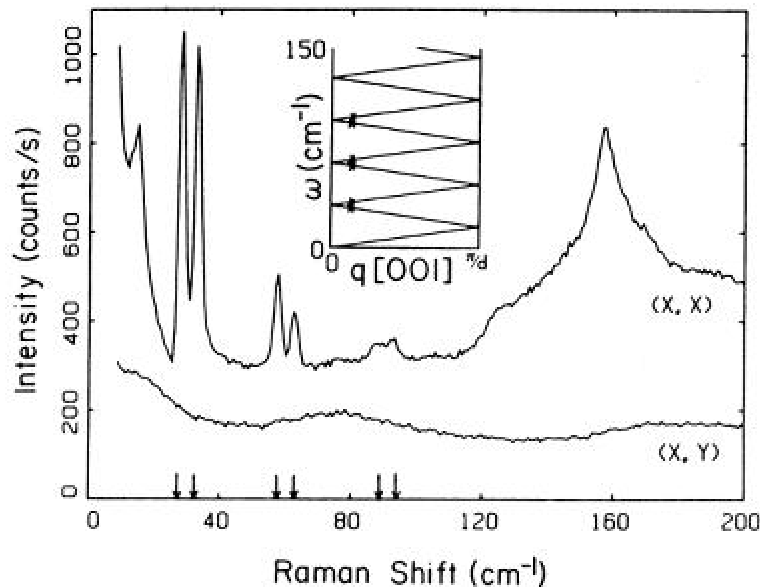


Figure 7.4. Raman spectrum of a superlattice comprising alternating 42 A GaAs slabs and 8 A Al0.3 Ga0.7 As slabs . The inset shows a model calculation of the folded acoustic branch as well as the location of the observed peaks.

Several other types of mode can occur in the GaAs/AlAs superlattice. For example, the (001) planes of atoms can also freely vibrate in the [110] direction, transverse to the direction of

propagation. Another type of mode which can arise are the interface modes which are confined to the regions of the interfaces between the GaAs and AlAs slabs. These various mode-types will also occur in GaAs/AlAs superlattices grown on other GaAs substrate faces and also in superlattices made from other alloy materials.

Although heterostructures such as quantum wells, wires and dots lack the layer periodicity of the superlattice, the basic features of the phonon modes are the same. For example, in a quantum well formed by sandwiching a GaAs slab between two AlAs slabs, the optical modes will be confined either to the GaAs or AlAs slabs, while the majority of the acoustic modes will extend throughout the whole structure. Interface modes will also occur.

Chapter 8. Optical Properties

Recent progess in epitaxial growth techniques has promoted the use of semiconductor heterostructures in optoelectronic devices. The physics of these materials relies upon the similarity between the electronic band structures of the different semiconductors. If the bulk band structures are sufficiently similar then changes in composition can be represented primarily as changes in the band splitting and other bulk parameters. In a direct band-gap semiconductor (where minimum of conduction band coincides to maximum of valence band in *k*-space (Figire 8.1)) an abrupt change in composition from wide to narrow band gap results in a discontinuity in the conduction and valence band profiles in the growth direction.

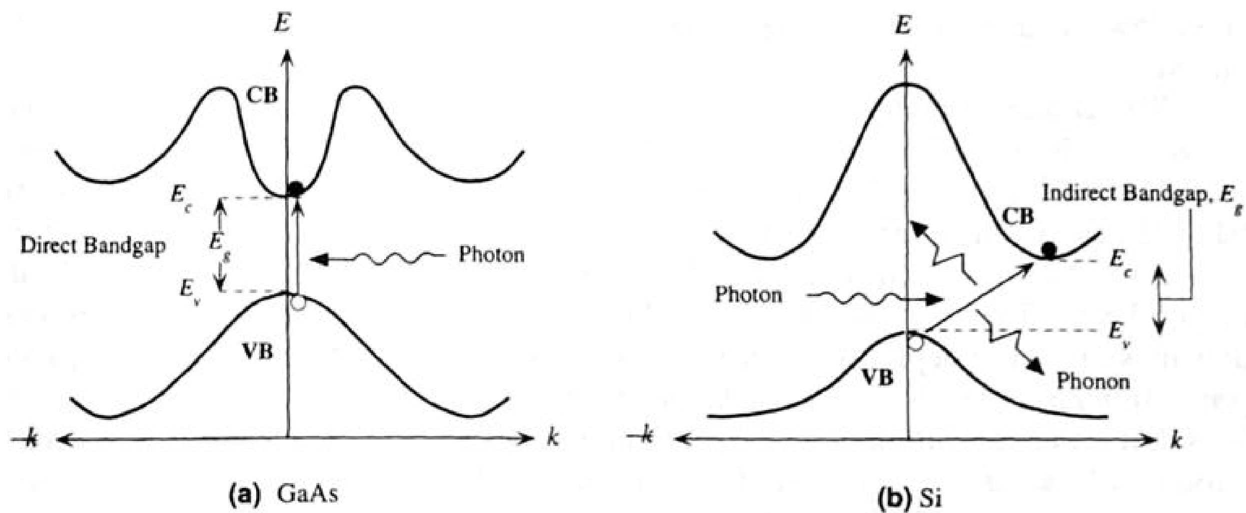


Figure 8.1 (a) direct bamd-gap semiconductor; (b) indirect band –gap semiconductor. In direct band-gap semiconductor $E(k)$ dispersion relation for conduction band has the minimum at the same value of k, where valence band dispersion relation has the maximum. In indirect band-gap semiconductor optical transition between the conduction and valence band is much less probable because it is simultaneous conservation of energy and momentum is impossible if only electron and photon participate in transition process

The heterointerface so formed is Type I or Type II, depending on the band-gap alignments, determined by the conduction band offset (Fig. 8.2).

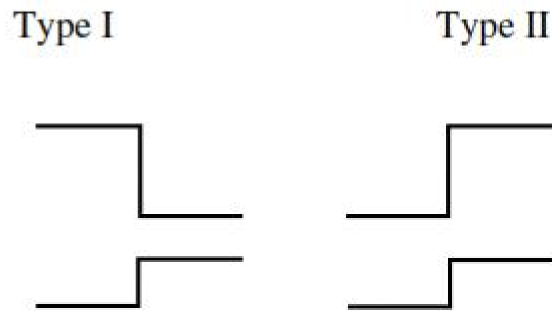


Figure 8.2 Type I and Type II heterointerfaces.

A quantum well (QW) is made by growing a thin layer - typically a few nanometres (nm) or 10s of nm - of narrower gap material within a wider-gap semiconductor, where the inserted layer is thin enough to cause quantum confinement of the carriers. QWs are similarly classified as Type I or II in direct-gap materials (Fig. 8.3).

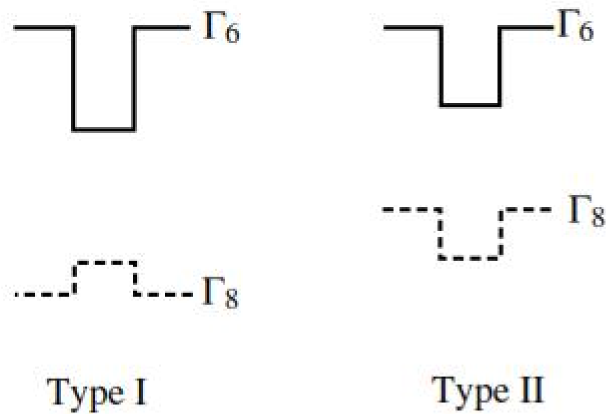


Figure 8.3 Type I and Type II quantum wells.

In indirect gap materials (where minimum of conduction band does not coincide to maximum of valence band in *k*-space) we need to consider the band-edge discontinuities at different points in the band structure. The AlAs/GaAs heterointerface, for example, is Type I at the point but Type II at the X point (Fig. 8.4). The overall band structure of an AlAs/GaAs QW system thus depends on the relative well and barrier widths.

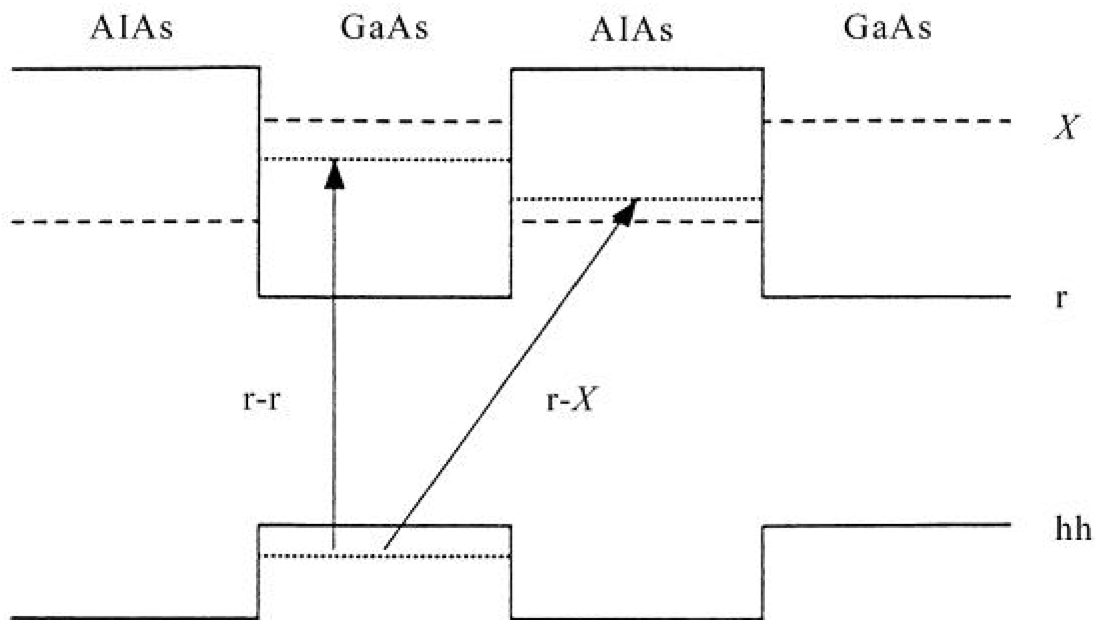


Figure 8.4Conduction and valence band profiles in AlAs/GaAs at the point and the X

In a multi-quantum well (MQW) system a series of QWs is separated by layers of wider-gap (barrier) material thick enough to isolate carriers in the wells (Fig. 8.5). A regular MQW with a barrier thin enough to allow carrier communication between the wells is known as a superlattice (SL).

The choice of materials for heterostructures is influenced by lattice spacings as well as by the symmetry of the crystal band structures and differences in the effective band gaps (Fig.

8.6). Strain forces at the interface between materials of very different lattice constant may produce imperfections which degrade the material and interface quality. However, in a small-period SL the thin layers may accommodate the alternating strain forces to produce useful heterostructures where the potential wells in different crystal bands are distorted by strain (Fig. 8.7).

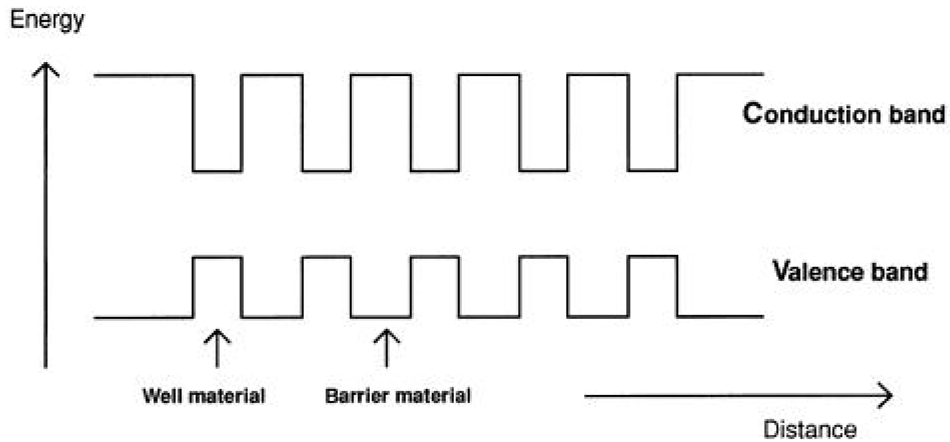


Figure 8.5 Multi-quantum well band profiles.

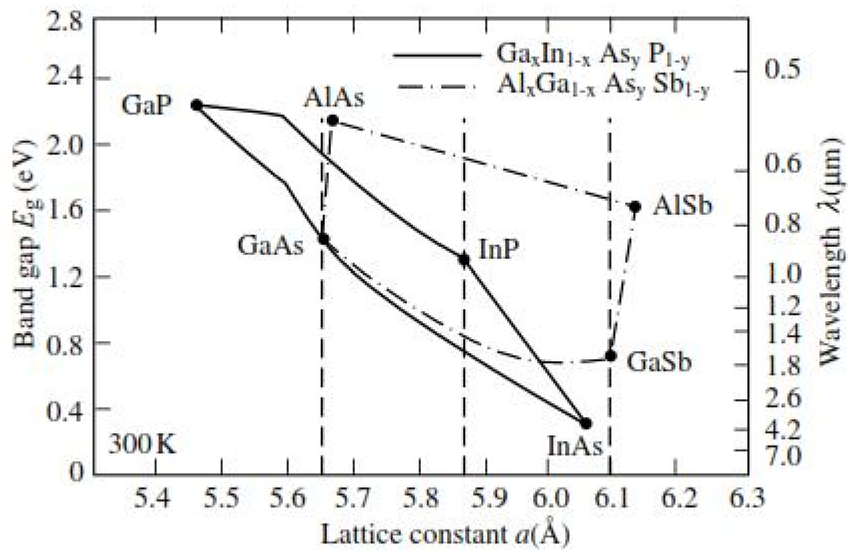
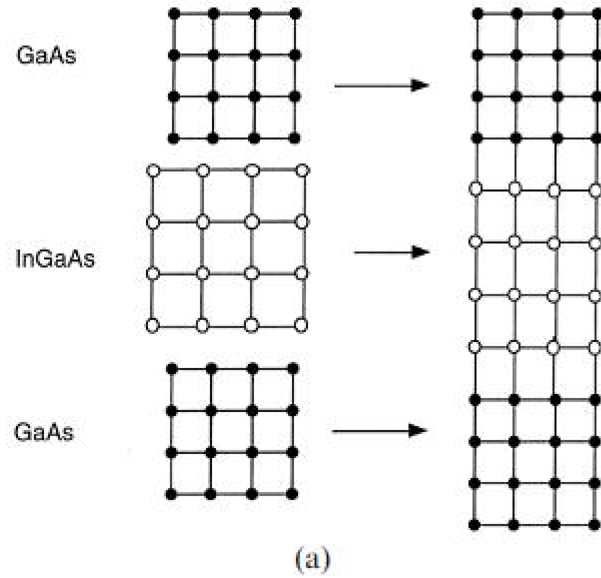


Figure 8.6 Room temperature band gap E_g versus lattice constant for the main binary III—V semiconductors. Tertiary alloys are represented by the lines joining relevant binaries.



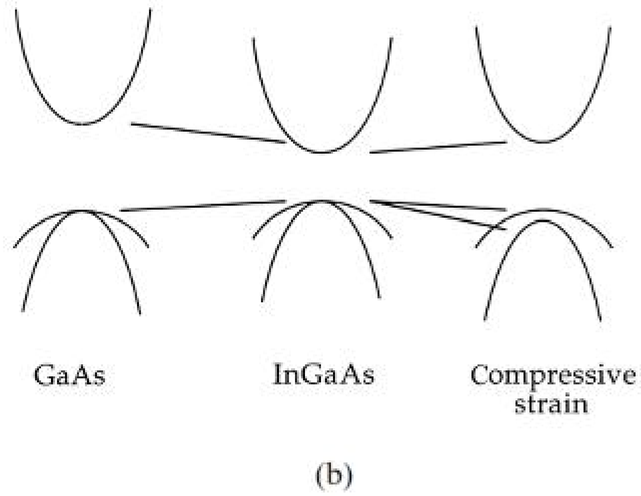


Figure 8.7 Dispersion relationships in bulk 1nGaAs and in strained GaAs/1nGaAs.

8.1 Optical Absorption

The absolute magnitude of the absorption for a single QW is low: approximately 1% of the light is absorbed on the step above the first light-hole and first heavy- hole transitions. This means that direct absorption measurements are difficult to perform except for systems with many QWs. Alternative methods of probing the absorption spectrum are photocurrent (PC) and photoluminescence excitation (PLE) spectroscopy.

In a photoconductivity measurement the QW sample must be contacted (typically as a *p-i-n* device) so that a bias may be applied. The QW is then illuminated with monochromatic light of frequency S and a small bias is applied to remove carriers photogenerated in the QW (Fig. 8.8). The photocurrent collected at the contacts, when measured as a function of S, reflects the QW absorption, but only to within a factor representing the quantum efficiency for escape from the well. At low temperatures, or in deep wells, this factor may vary significantly with the excitation energy and produce an unreliable picture of the relative absorption strength, though not of the positions of features. Another difficulty with this technique, particularly when dealing with wide wells, arises from the Stark shift- ing of the subband edges by the applied bias.

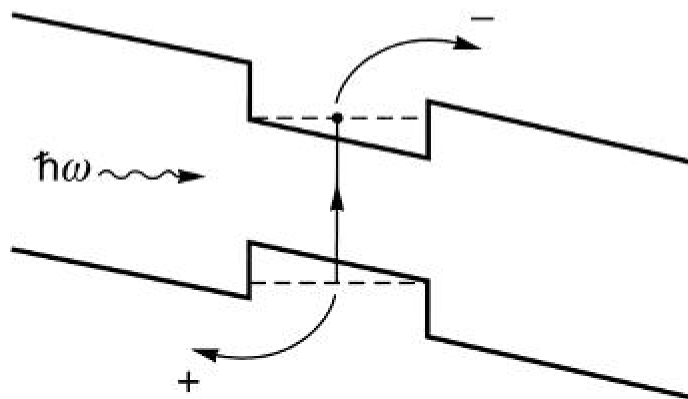


Figure 8.8 Measurement of photoconductivity.

Photoluminescence excitation spectroscopy may be used to study uncontested samples. In a PLE measurement, the QW is illuminated by light of frequency S and a detector is fixed at an energy just below the effective absorption edge (the first electron-heavy-hole exciton). As the energy of the incident light is varied, electrons are promoted from the valence band to the conduction band according to the absorption strength. Carriers first thermalize rapidly to their respective band edges and then recombine slowly by radiative recombination across the band gap (Fig. 8.9). The PL signal will thus vary with w in harmony with the strength of the absorption (Fig. 8.7). PLE will be more reliable for states close to the absorption edge, where carriers are less vulnerable to removal by other mechanisms.

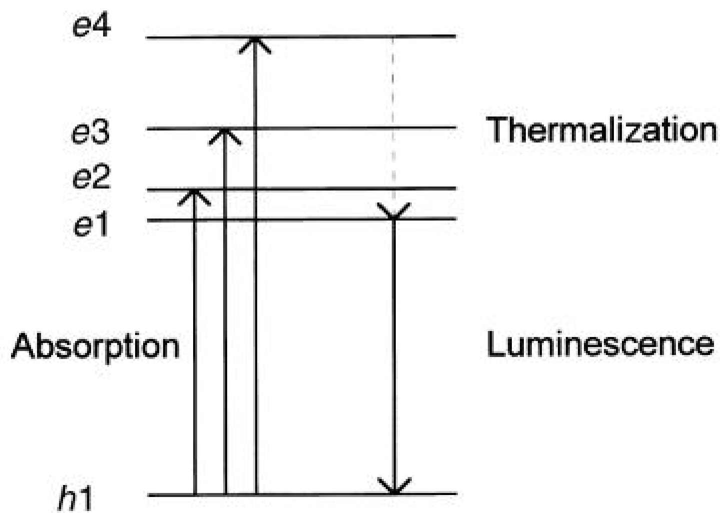


Figure 8.9. Photogeneration and recombination of carriers in a quantum well in photo- luminescence excitation.

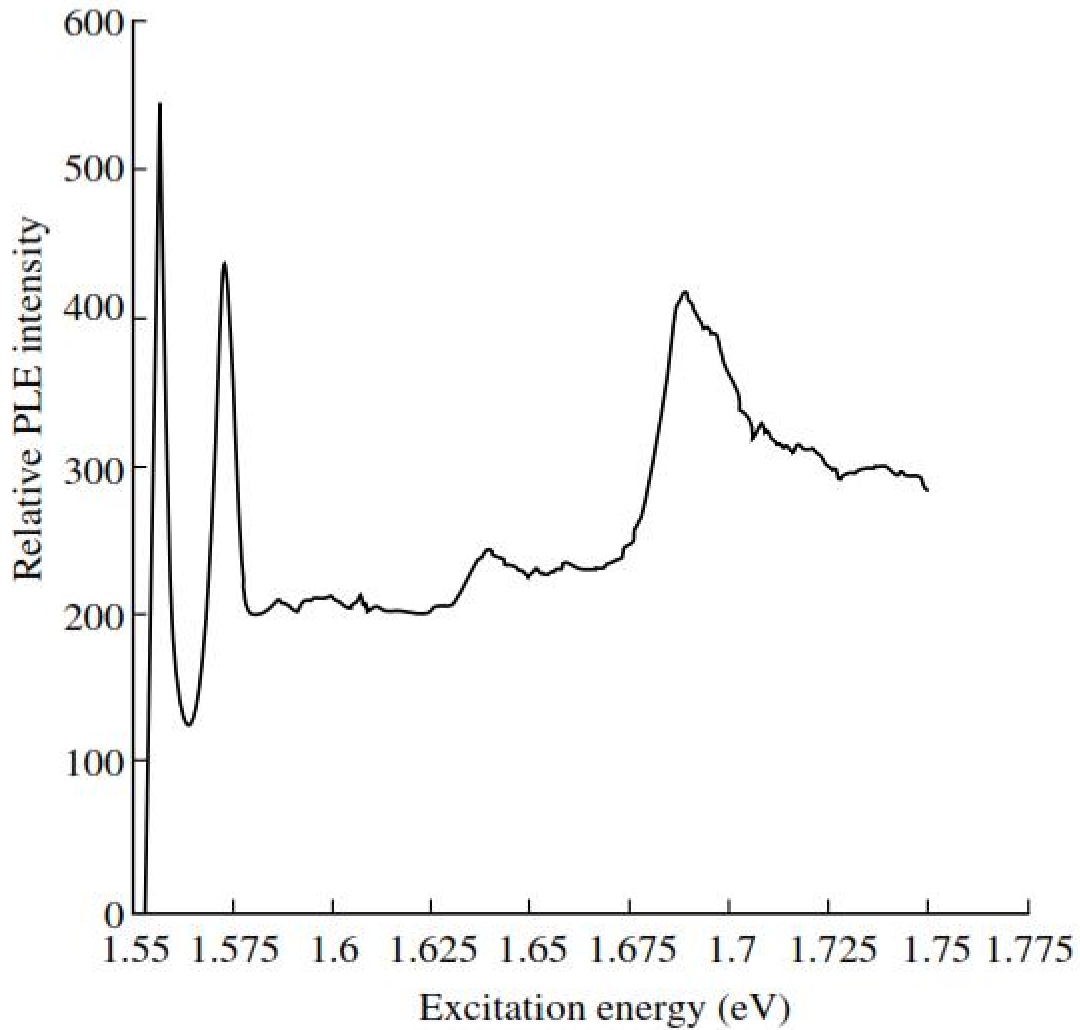


Figure 8.10 Photoluminescence excitation spectrum of an 8.4 nm $Al_{O.33}$ $Ga_{O.67}$ As/GaAs multi-quantum well at 11 K. As well as the main transitions between subbands of the same order (el—hhl, el—lhl and e2—hh2) the exciton for the optically weak el—hh3 transition can be seen at l.65 eV.

8.2 Features of Optical Spectra

The peculiarities of optical spectrum of nanostructures will be discussed at the example of colloidal quantum dots. Colloidal II -VI semiconductor nanocrystals or quantum dots (QDs) have at aimed a great research focus due to their advantages in optical properties including tunable

emission spectra, high photostability, resistance to photobleaching and control able surface characteristics. CdSe Qds find a wide range of applications in optoelectronic devices, photo catalysis, solar energy conversion and biological imaging and labeling. The main property of CdSe QDs is optical property. The nanoparticles are a bridge between bulk materials and atomic or molecular structures. A bulk material should have constant physical properties regardless of its size, but at the nano-scale this is not the case. The properties of materials change as their size ap roaches the nanoscale and the percentage of atoms at the surface of a material becomes significant. The size of the nanoparticles is finte, so the continuous energy band of the bulk crystal transforms into a series of discrete states. The nanoparticles frequently display photoluminescence and sometimes display electroluminescence. It is well known that the quantum confinement effect modifies the electronic structure of nanocrystals when their diameter is comparable to or smaller than the diameter of the bulk exciton.

As mentiond, quantum dots are nanoparticles of semiconductors materials ranging from 2 to 10nm in diameter, like CdSe and ZnS. **There electronic characteristics are closely related to the size and shape of the individual crystal.** If the size of crystal is small , then band gap between the higher valence band and the lowest conduction band becomes high and more energy is require for exciting the dot and consequently, more energy is released when the crystal returns to its resting state. A principal advantage with quantum dots is that by controlling the size of crystals, the conductive properties of the material is controlled. Because of their small size, quantum dots displays unique optical and electrical properties. The most immediately apparent of these is the emission of photons under excitation, which are visible to human eyes as light. The wavelength of these photon emissions depends not on the material from which the quantum dot is made, but its size. The ability to control the size of quantum dot enables the manufacturer to determine the wavelength of emission, which in turn determines the color of light the human eye perceives. The smaller the dot, closer it is to the blue end of the spectrum and the larger the dot, closer to the red end as shown in Fig 8.11. When the size of the quantum dot is smaller than the critical characteristic length called the exciton. In Bohr radius, the electrons crowding lead to the spliting of the original energy levels into smaller ones with smaller gaps between each successive level. The quantum dots that have radii larger than the exciton Bohr radius are said to be in the “weak confinement regime” and the ones that have radii smaller than the exciton Bohr radius are said to be in the “strong confinement regime”.

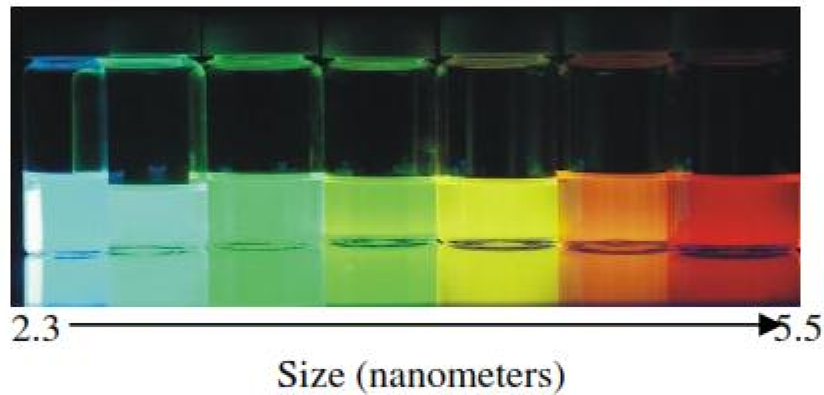


Figure 8.11 Different size quantum dots emitting light at different frequencies.

The fluorescence of the quantum dots is a generated when valence electron excite with a certain energy and they emits energy in the form of photons as the excited electron returns to the ground state, combining with the hole. The energy of the emitted photon is determined by the size of the quantum dot due to quantum confinement effects. The energy of the emitted photon is sum of the band gap energy between occupied level and unoccupied energy level, the confinement energies of the hole and the excited electron, and the bound energy of the exciton as shown in Fig 8.12.

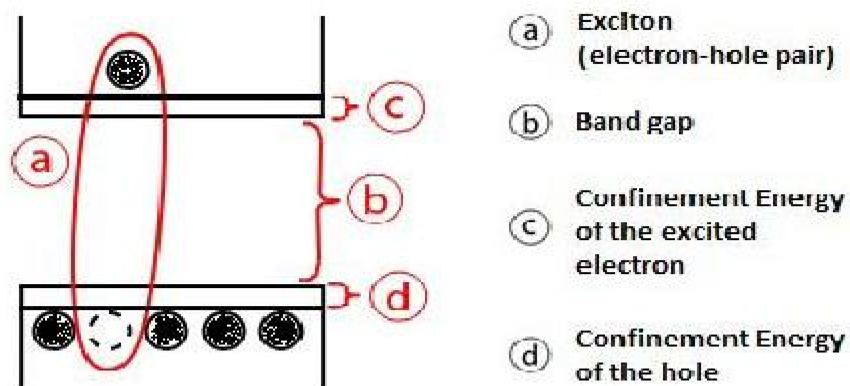


Figure 8.12 bang gap diagram

The electronic state is an important property and can be described in terms of valence and conductivity bands and a gap between these bands. However, as the particles become smaller, the wavelength of the electrons is closer to the range of the particle sizes and the laws of classical physics have to be substituted by quantum confinement or quantum size effect. The UV-visible spectrum showed that the absorption peak of obtained CdSe QDs in aqueous solution is 543 nm (2.28 eV), relative blue-shift to the band gap of bulk cubic CdSe (1.78 eV, 698 nm).

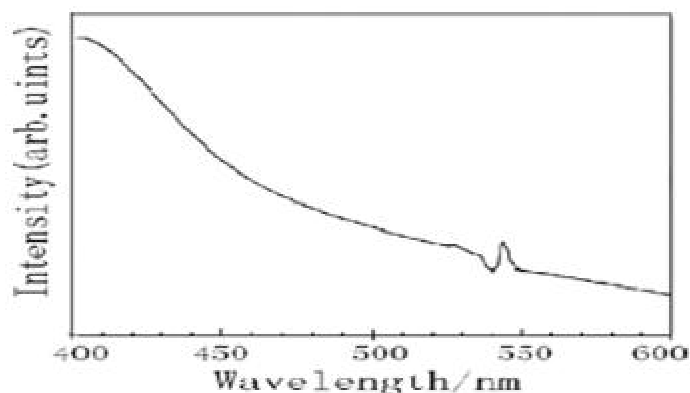


Figure 8.13 UV-Visible absorption spectra of CdSe QDs.

The quantum dots of the same material, but with different sizes, can emit light of different colors. The physical reason is the quantum confinement effect. The larger dot gives low energy fluorescence spectrum. Conversely, smaller dots emit bluer light (Fig.8.14). The coloration is directly related to the energy levels of the quantum dot. The band gap energy that determines the energy of the fluorescent light is inversely proportional to the size of the quantum dot. Larger quantum dots have more energy levels which are also more closely spaced. This allows the quantum dot to absorb photons containing less energy, i.e., those closer to the red end of the spectrum. The lifetime of fluorescence is determined by the size of the quantum dot. Larger dots have more closely spaced energy levels in which the electron-hole pair can be trapped. Therefore, electron-hole pairs in larger dots live longer causing larger dots to show a longer lifetime.

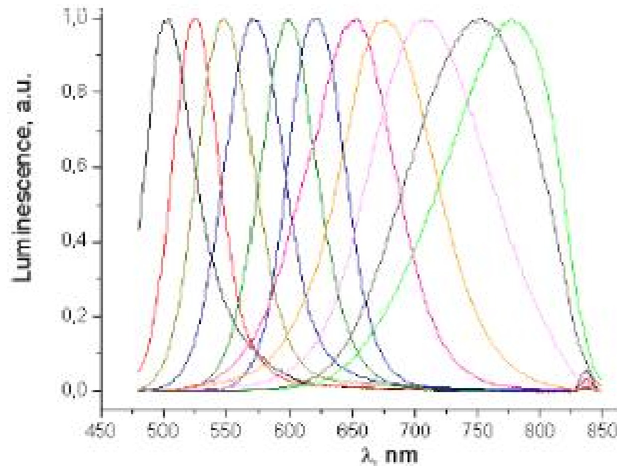


Figure 8.14 Fluorescence spectra of CdSe quantum dots of various sizes.

8.3 Quantum-well Solar Cells

Quantum-well structures are important in optoelectronics because they offer the joint benefits of highly confined electron and hole populations and a tunable band gap. One rather less obvious application is to high efficiency photovoltaics: the quantum-well solar cell (QWSC).

A simple, single band gap photoconverter works by absorbing incident photons of energy greater than the band gap and separating the charges so produced to deliver an electric current to an external circuit. In a semiconductor solar cell the charges are separated by the built-in electric field of a p-n (or p-i-n) junction. Not all the light energy absorbed can be converted into electrical energy since photogenerated carriers quickly decay to their ground state, losing any excess energy as heat. A photon of energy ħŠ absorbed in a semiconductor of band gap E_g can thus deliver no more than E_g of electrical potential energy to the external circuit. Increasing the band gap increases the potential - and hence the cell voltage; but decreases the photocurrent since only high energy photons can be absorbed. Therefore, for any broad band incident spectrum, optimum power conversion is achieved at some intermediate band gap E_{opt}.

These considerations place a limit on the efficiency available from a perfect single threshhold photoconverter. In the standard air mass 1.5 solar spectrum a limit of 31% is reached at a band gap of 1.35 eV. In principle, such limits may be surpassed in a multi-band-gap or tandem system, where different parts of the spectrum are preferentially absorbed in materials of different band gap. Independent optimization of band gaps and photocurrents increases the theoretical efficiency to about 45% for the standard solar spectrum or 50% under a concentration of 1000 . Practical improvements fall far short of these limits, usually through electrical losses.

Quantum well structures are interesting for photoconversion firstly as an alternative multi-band-gap approach to the tandem cell. Quantum wells added to the space charge region of a single band gap p-i-n solar cell extend its spectral response to longer wavelengths and so increase photocurrent (Fig. 8.15). At the same time the QWs act as centers for enhanced recombination and reduce the cell's operating voltage. If the improvement in current outweighs the loss in voltage, a net increase in power conversion efficiency is achieved.

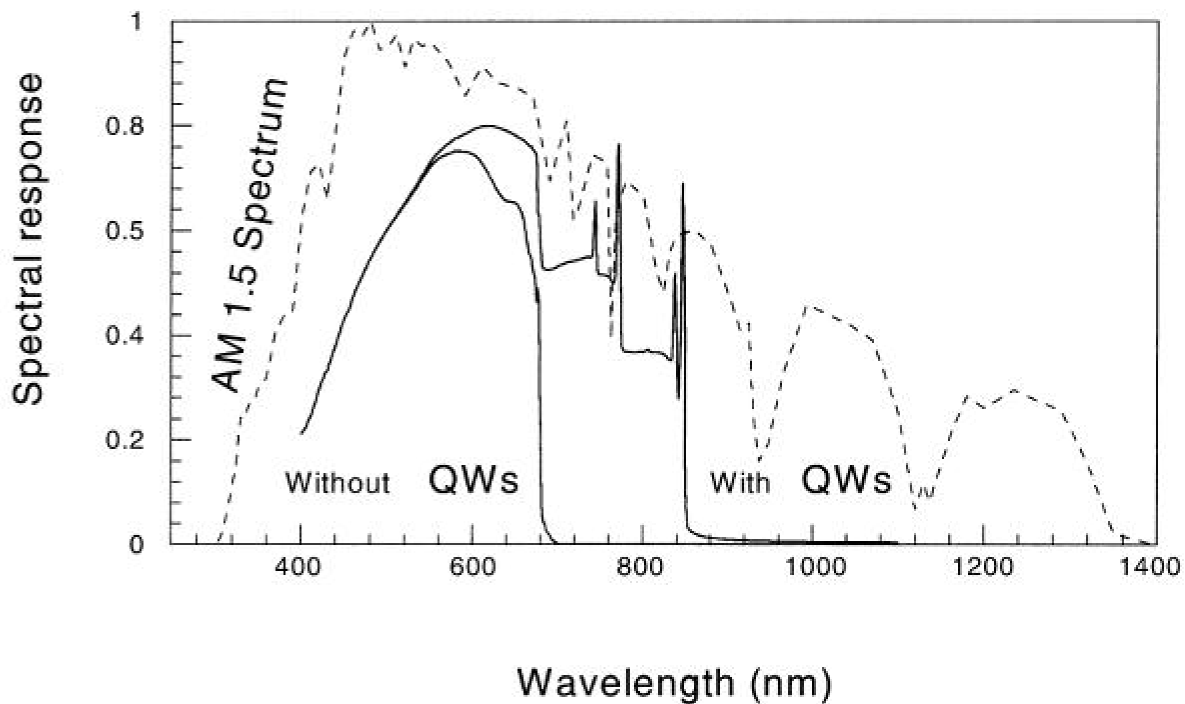


Figure 8.15 . Calculated spectral response of an $Al_{0.33}Ga_{0.67}As$ p-i-n solar cell with and without GaAs quantum wells in the 0.8 μm intrinsic region. Quantum wells extend the spectral response to longer wavelengths where the solar spectral irradiance is still high. (AM air mass)

An increase in the limiting efficiency of a monolithic solar cell requires that such enhancements be available even using a host cell of optimum band gap. This situation essentially requires that the quasi-thermal equilibrium between the QW and host material is broken, so that recombination proceeds more slowly than expected.

Less radically, QW structures are interesting in the possibility of superior practical performance to the homogenous cell of equivalent band gap. By careful choice of materials for well and barrier it may be possible to design a QWSC where the net recombination current is smaller than in the equivalent bulk alloy. This has been experimentally confirmed in InP/$In_{O.52}$ $Ga_{O.48}$ As QW structures.

QWSC behaviour depends upon the QW electronic structure in several ways. The joint density of states function determines the QW absorption spectrum which, in turn, determines the photocurrent enhancement. Density of electron states in the conduction band and hole states in the valence band determine the spatial distribution of electrons and holes and, hence, the recombination current. Finally, the mixing of states between QW and host material or between neighboring QWs can affect the efficiency of carrier transport.

8.4 Quantum-dot Solar Cells

As mentioned above, a solar cell is a device that converts photons from light into electricity. Fundamentally, the device needs to fulfill two functions: photogeneration of charge carriers (electrons and holes) in a light-absorbing material, and separation of the charge carriers to a conductive contact that will transmit the electricity.

Due to the nature of photovoltaics, the light-absorbing material will only absorb certain energy level from photon. For example, Silicon has a band gap of 1.14 eV, which means 1.14 eV can excite the electron into conduction band. If photon has energy more than 1.14 eV, excess energy will generate heat instead of generating more electron-hole pairs (Fig.8.16). This phenomenon limits the overall efficiency of the conventional semiconductor photovoltaic device.

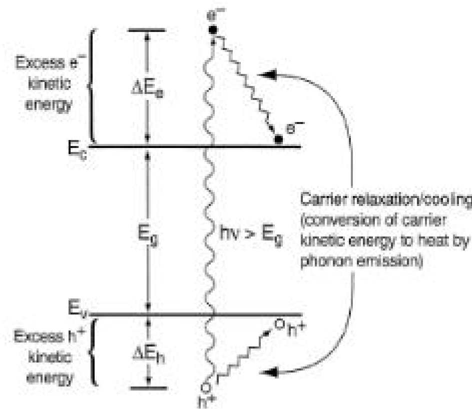


Figure 8.16 Thermalization losses in solar cells

One of the potential approaches to overcome this limit is carrier multiplication (CM), or multiple excition generation (MEG). The idea is illustrated in figure 8.17. In a traditional photon excitation in figure 8.17 a), one phonon can generate only one pair of exciton. The excess photon energy $\hbar\check{S} - E_g$ is dissipated as heat via phonon emission and therefore is wasted. In this case, the quantum efficiency (QE) of photon-to exciton conversion is zero below E_g , the energy gap, and is 100% above it. This corresponds to the Shockley Queisser limitation of the maxima quantum efficiency mentioned earlier. In figure 8.17 b), the number of excitons produced by a single photon is only limited by energy conservation. Photons with energies E_g , 2 E_g , and 3 E_g produce one, two, and three excitons, respectively. The QE is increased by 100% if photon energy is increased by E_g .

The MEG phenomenon has been known in bulk semiconductors since the 1950s. However, since the restrictions imposed by energy and momentum conservation, the power conversion efficiency improvement due to MEG was less than 1%. Recently, it was discovered that while being low efficiency in bulk semiconductors, MEG can become extremely efficient in ultrasmall semiconductor nanocrystals. This lead to the quantum dots (QDs) solar cells.

In semiconductor nanocrystals i.e. quantum dots, discrete quantized energy levels are formed that affect the relaxation dynamics of hot electrons and carriers, and allow other relaxation channels to compete with phonon emissions. As Schrödinger equation explains, the

states are quantized according to angular momentum in quantum dot, and translation momentum conservation is replaced with angular momentum conservation, which is less restrictive.

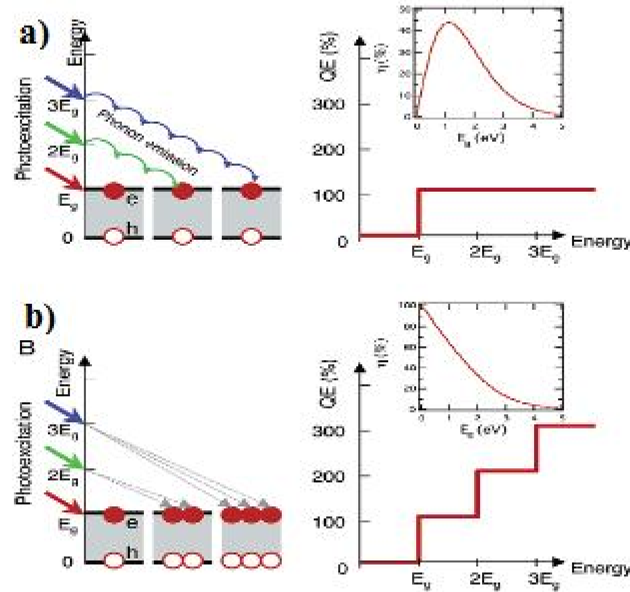


Figure 8.17 a) Traditional schematic of solar cell; b) Multiple excitons are generated with one absorption of a single photon

As mentioned at limitation of bulk semiconductor solar cell, high-energy electrons (hot electron-hole pair created by absorption of photons larger than band gap) convert their excess energy to heat through phonon emission. Normally the time scale of this relaxation in bulk semiconductor is in the order of sub picosecond scale. However, in quantum dot the generated electron-hole pairs become bound to each other due to strong quantum confinement. And the formation of discrete quantized energy level slows recombination time scale. So the slowed cooling of energetic excitons are able to enhance the photon conversion efficiency by allowing free energy to be extracted from the high energy excitons before they relax to their lowest state and produce heat.

Limitation of solar cell is also related with recombination of electron-hole pair. Recombination is part of a process to restore equilibrium to a semiconductor that has been perturbed, or disturbed out of equilibrium. Perturbations can be in the form of an applied electric field, a change in temperature or exposure to light. Recombination occurs when there is an excess of carriers and they are destroyed, by recombining. When electron-hole pairs are destroyed, a negatively charged electron is attracted to a positively charged hole, and as they get together, their charges are canceled and the electron is part of a bond once again. Auger recombination is a type of band-to-band recombination that occurs when two carriers collide (Fig.8.18). The collision transfers the energy released from the recombining carrier to the surviving carrier. In other words, one carrier loses energy and the other gains it. The one that loses it is recombined, and the one that gains it goes to a higher energy level. Eventually, this highly energized carrier "thermalizes" - loses energy in small steps through heat producing collisions with the semiconductor lattice, until it eventually recombines or gains energy once more. And larger hole effective mass leads to rapid thermalization.

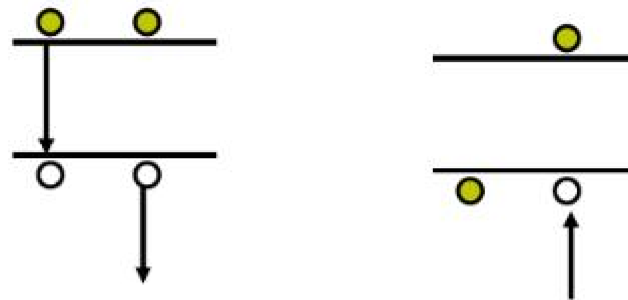


Figure 8.18 Auger recombination in quantum dots

One of the approaches to overcome those limitations (thermalization and Auger recombination) is MEG or CM. CM is very efficient in quantum-confined semiconductor nanocrystals, whereas it is inefficient in bulk semiconductors (the maximum CM-induced increase in the efficiencies of traditional solar cells is less than 1%). By analyzing dynamical signatures of excitons and multi-excitons in PbSe quantum dot, it is discovered that the absorption of a single photon can produce

two or even three electron–hole pairs (excitons), which results in internal quantum efficiencies greater than 200% for conversion of light quanta into charge carriers. The schematic of generation/relaxation kinetics in a quantum dot is described in Figure 8.19. At time t1, a photon with energy greater than the CM threshold results in the generation of is absorbed 'hot' bi-exciton(at t2) on the time scale of CM. After relaxing to its ground state (at t3) with the sub-picosecond time constant r, this bi-exciton recombines on the sub-nanosecond time scale (timeconstant A) by the non-radiative Auger process to produce a single exciton (at t4). Finally, the exciton recombines radiatively on a much slower timescale of tens to hundreds of nanoseconds. A significant difference in the relaxation behavior of bi-excitons (fast decay) and single excitons (slow decay) is the key property of nanocrystals. Therefore considering CM and Auger recombination kinetics, current should be extracted before Auger recombination occurs.

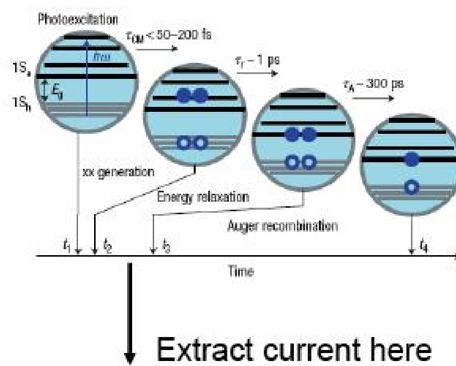


Figure 8.19 Evolution of electron –hole system when photon is absorbed

Chapter 9 Localization and Quantum Transport

Traditional solid-state physics is based on the concept of the perfect crystalline solid, sometimes with a relatively low density of defects. This perfect crystallinity has played a crucial role in the development of the subject, with Bloch's theorem stating that wave function of electron in crustal can be written by

$$Œ_{n\vec{k}}\left(\vec{r}\right)=u_n\left(\vec{r},\vec{k}\right)e^{i\vec{k}\vec{r}} \tag{9.1}$$

providing the central conceptual base. Concepts that arise from this theorem, such as bands, Brillouin zones, vertical transitions, effective mass and heavy and light holes, are really only well-defined in a perfect infinite crystal. In the absence of crystallinity none of these concepts is strictly valid, though in some cases it provides a useful starting point. In general, however, a new approach is required to characterize electrons and phonons in disordered solids.

When we consider low-dimensional structures Bloch's theorem may or may not be valid. There is nothing intrinsic to low dimensionality which invalidates it. Many of the simple examples in quantum mechanics and solid-state physics text-books are, in fact, one-dimensional (e.g. the particle in a box, the Kronig—Penney model). Indeed, in a quantum well prepared by any of the standard growth methods, much of the physics can be understood by using basics of quantum mechanics and the effective mass approximation. This is because a region of adjacent GaAs layers in $Al_x Ga_{1-x}$ As can, for many purposes, be regarded as a perfect potential well. By doping the AlGaAs, the electrons in the well can be spatially separated from the scattering due to the ionized donor atoms. Thus, in many respects, the electrons in this system can be treated as particles in a one-dimensional box.

The quantum well is, however, a very special quasi-two-dimensional system, albeit a very important one. As discussed, it is very difficult to prepare low-dimensional samples of high quality for other than lattice-matched planar heterostructures. Thus, most heterojunctions, such as those with a significant lattice mismatch, metal-oxide-semiconductor field-effect transistors

(MOSFETs), narrow quantum wells and quantum wires, etc., are in practice highly disordered with an effective density of scatterers which can approach the density of atoms. Clearly, in such systems, it cannot be valid to treat the effect of scatterers with perturbation theory using the perfect crystalline case as a starting point.

As we shall see later, there is one sense in which low-dimensional systems are intrinsically different from three-dimensional systems. The amount of scattering required to produce dramatic changes in the behavior can sometimes be so small that perturbation theory may never be valid.

9.1 Localization, Percolation

Let us start with a simple classical problem. How does a fluid flow through a random medium? This is a problem of considerable practical importance in its own right: the extraction of oil from porous rock strata. Consider a random landscape which is being slowly filled with water. At first there will be a continuous land mass with a few lakes (Fig. 9.1(a)). When the water level is very high we have islands in a sea (Fig. 9.1(b)). Let us now suppose there is a dam at the edge of the area which requires large quantities of water to drive a power station. When the water level is low only the lake next to the dam can be used and it will soon run out. As the level is raised this lake becomes larger but still finite. The power station will run longer but will still eventually drain the lake and have to stop. At a critical water level (Fig. 9.1(c)) the system changes from a lake district to an archipelago. This is analogous to the percolation transition, where the water first forms a continuous network through the landscape. After this the power station can run indefinitely without fear of running out of water.

This phenomenon has much in common with more conventional phase transitions. There is a characteristic length scale which diverges at the transition: the size of the lakes or islands. There is a well-defined critical water level, rather like the critical temperature of the freezing transition or the ferromagnetic-to-paramagnetic transition in iron.

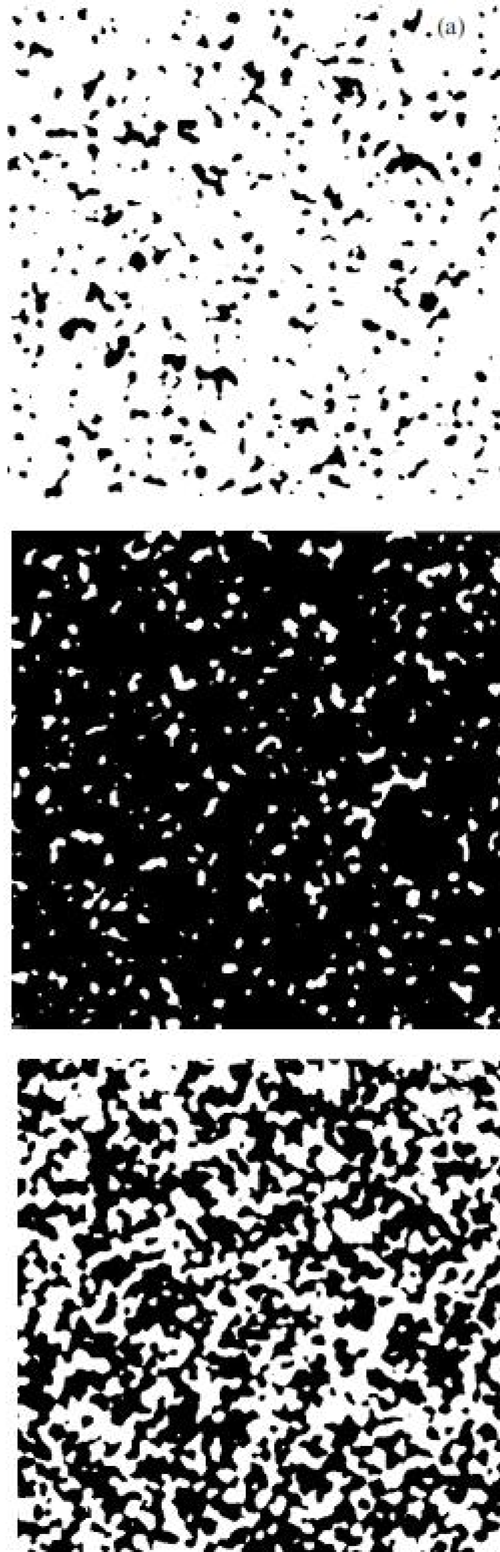


Figure 9.1. Percolation diagrams, with (a) low water level with a few lakes, (b) high water level with a few islands, (c) intermediate (critical) water

If we think in terms of the density of blockages rather than the water level we see that there is a critical density above which the flow of water stops.

The one-dimensional version of this problem is special. Any blockage of the channel is enough to prevent the flow of water. The critical density is zero. This is an example of a problem which cannot be solved by perturbation theory. There is a discontinuous jump in the behavior between a system with no blockages and one with a single blockage. In higher dimensions, in contrast, water can flow around the blockage.

9.2 The Anderson Transition and the Mobility Edge

The concept of the localization of electrons caused by disorder is due to Anderson. He argued that an electron which starts at a particular site cannot completely diffuse away from that site if the disorder is greater than some critical value. Anderson thus introduced the concept of localized and extended states. The characteristics of these states can be summarized as follows

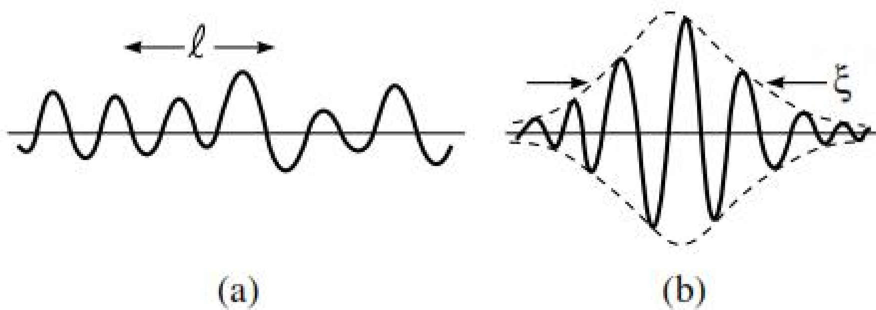


Figure 9.2. Schematic diagrams of (a) extended and (b) localized states, showing the correlation length l, and the localization length ‹.

(a) extended (i) spread over the entire sample
(ii) not normalizable
(iii) contributes to transport

(b) localized (i) confined to a finite region
(ii) normalizable
(iii) does not contribute to transport

It is worth noting at this point that the phenomenon of localization is not confined to electrons, but can also be observed in other wave phenomena in random media, such as acoustic and optical waves, as well as water waves (Fig. 9.3).

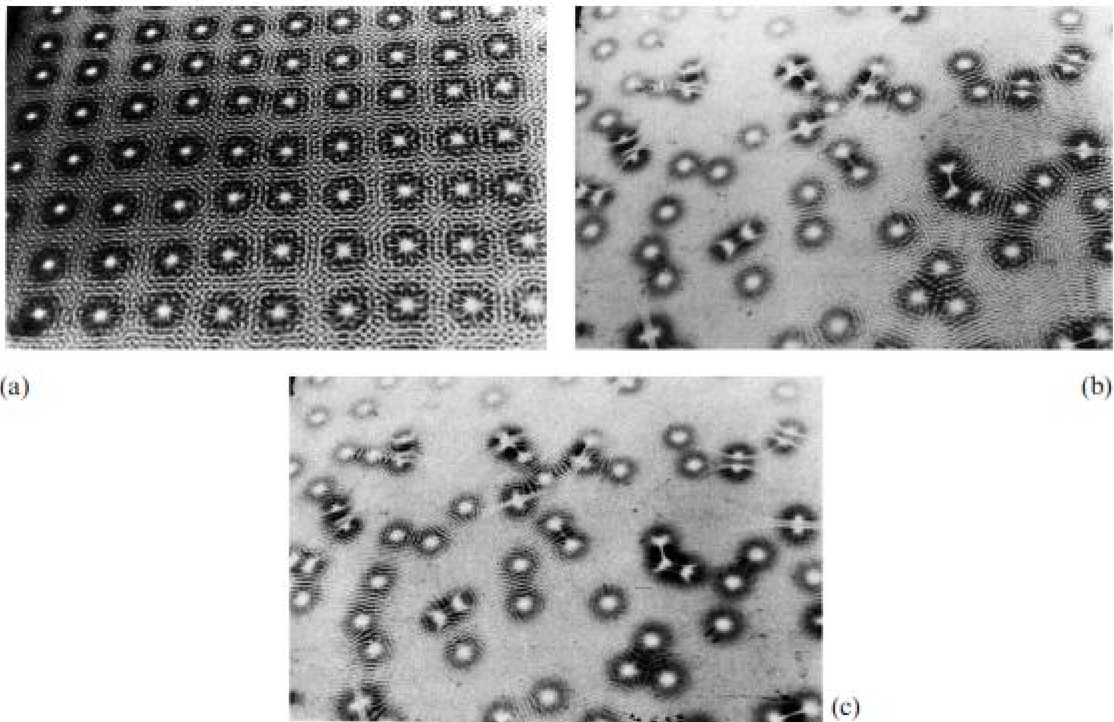


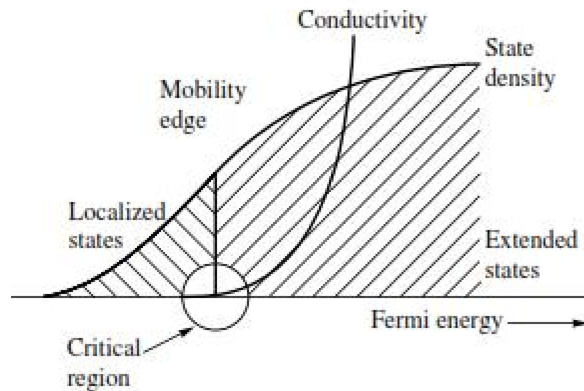
Figure 9.3 Three photographs of a water bath exposed to an audio-frequency oscillation. (a) Shows a situation where the obstacles sit in a regular quadratic lattice (frequency 76 Hz). We see strong Bragg reflection corresponding to standing waves. (b) and (c) show randomly spaced obstacles exposed to two different audio frequencies (105 Hz and 76 Hz). Both (b) and (c) show standing wave patterns, but localized in different areas.

Mott later introduced the concept of a mobility edge (Fig. 9.4). He argued that it is meaningless to consider localized and extended states which are degenerate since any linear combination of a localized and an extended state must be extended. Thus, the concept of localization can only be meaningful if there are separate energy regions of localized and extended states, rather like bands and gaps. These regions are separated by a mobility edge. Mott further argued that the states close to a band edge are more likely to be localized than those in the middle of a band. Since the localized states do not take part in conduction, electrons in a disordered semiconductor must be activated to beyond the mobility edge rather than simply to the band edge to contribute to the conductivity. This activated process would be manifested in a conductivity σ of the form

$$\sigma = \sigma_0 \exp\left(-\frac{E_\mu - E_F}{k_B T}\right) \tag{9.2}$$

where E_μ and E_F are the mobility edge and Fermi energy, respectively. This form should reveal itself as the slope in an Arrhenius plot of the conductivity, i.e. a plot of $\ln\sigma$ vs. $1/T$:

$$\ln\sigma = \ln\sigma_0 - \frac{E_\mu - E_F}{k_B T} \tag{9.3}$$



Metallic Conductivity

The semi-classical conductivity can be written in the form

$$\dagger = \frac{n\ddagger e^2}{m} = \frac{ne^2 l}{mv_F} = \frac{ne^2 l}{\hbar k_F} \quad (9.4)$$

where n is the density of conduction electrons, m and e are the electron mass and charge, respectively, ‡ is a scattering time, l is the mean free path, and v_F and k_F are the Fermi velocity and wave vector. The density n of electrons is proportional to temperature in semiconductors, however in metals it is constant value, and conductivity depends on *l,* which in its turn is determined by scattering processes.

Conductivity in nanowires

Several physical reasons predict that the conductivity of a nanowire will be much less than that of the corresponding bulk material. First, there is scattering from the wire boundaries, whose effect will be very significant whenever the wire width is below the free electron mean free path of the bulk material. In copper, for example, the mean free path is 40 nm. Copper nanowires less than 40 nm wide will shorten the mean free path to the wire width. The conductivity of a nanowire can be studied suspending it between two electrodes.

Nanowires show peculiar electrical properties due to their size. Unlike carbon nanotubes, whose motion of electrons can fall under the regime of ballistic transport (meaning the electrons can travel freely from one electrode to the other), nanowire conductivity is strongly influenced by edge effects. The edge effects come from atoms that lay at the nanowire surface and are not fully bonded to neighboring atoms like the atoms within the bulk of the nanowire. The unbonded atoms are often a source of defects within the nanowire, and may cause the nanowire to conduct electricity more poorly than the bulk material. As a nanowire shrinks in size, the surface atoms becomes more numerous compared to the atoms within the nanowire, and edge effects become more important.

Furthermore the conductivity can undergo a quantization in energy: i.e. the energy of the electrons going through a nanowire can assume only discrete values, multiple of the Landauer constant G = 2e2 / h (where e is the charge of the electron and h is Planck's constant).

The conductivity is hence described as the sum of the transport by separate channels of different quantized energy levels. The thinner the wire is, the smaller the number of channels available to the transport of electrons.

Chapter 10 Application

As described, 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.

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

10.1 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 asymetric nanowires using a number of approaches including vapor-liquid-solid (VLS) and solution- liquid-solid (SLS) growth. The move to applications has occured 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.

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.

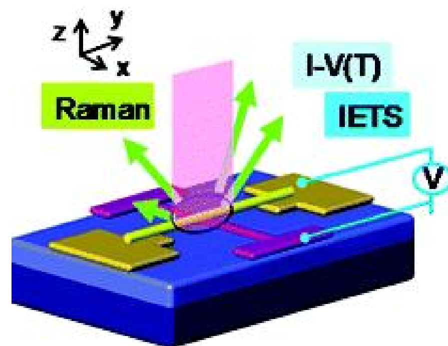


Figure 10.1: A crossed nanowire junction

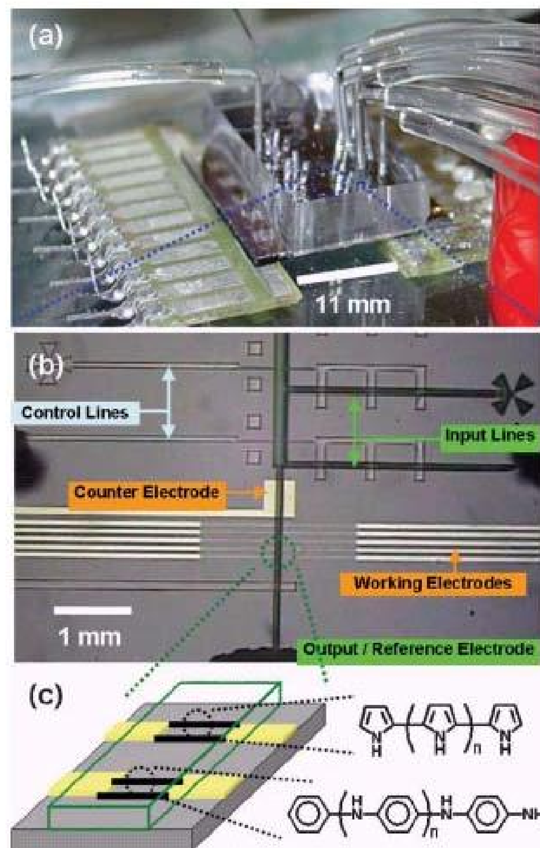


Figure 10.2 Conducting polymer nanowire sensor formed directly in microfluidics device showing (a) actual view of fabricated device, (b) optical micrograph of microfluidics device and (c) schematic with polyaniline and polypyrrole

10.2 Quantum Dots

In the realm of colloidal quantum dots the following applications have been proposed:

- Quantum dots for biological labeling
- 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 be- cause 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 106 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 fluorphores. 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 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 that organic dyes when it comes to photobleaching. Dots have been seen to absorb and emit over 108 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.

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 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.

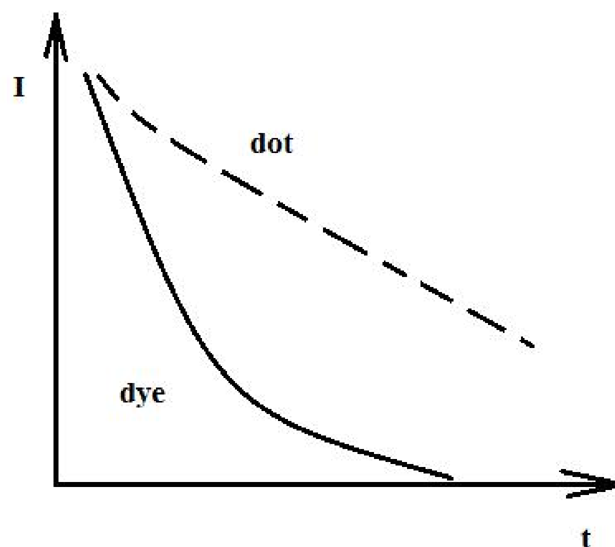


Figure 10.3 Comparison of a quantum dot to organic dye photobleaching rate

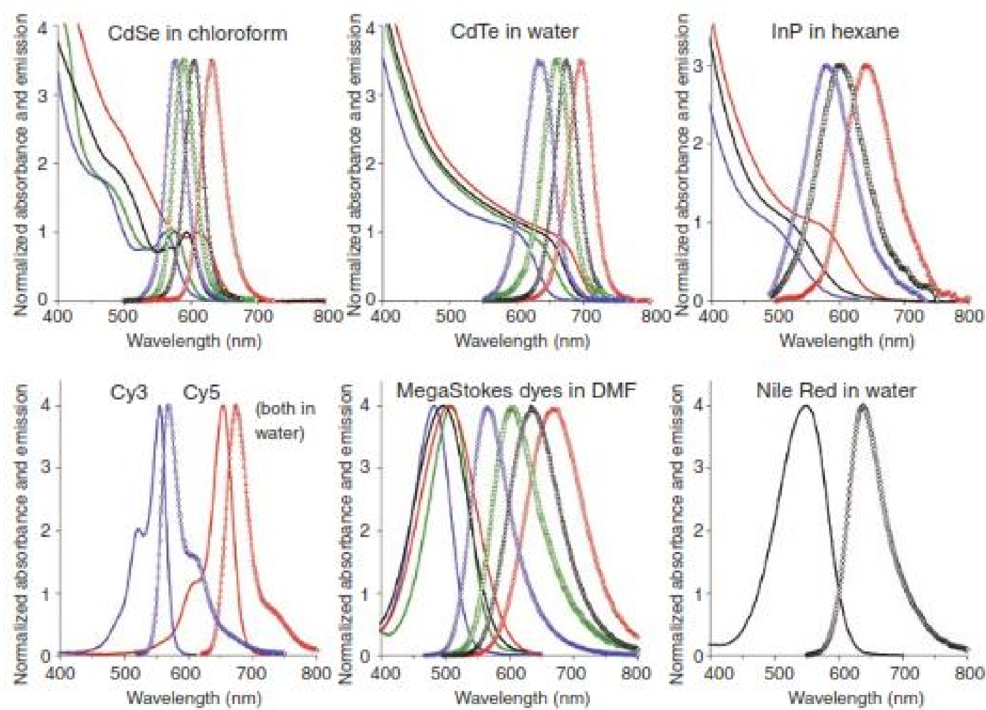


Figure 10.4: Comparison of quantum dot absorption/emission spectra to organic dye absorption emission spectra in light of multicolor labeling experiments

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 su- permarket 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 re- ceived 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 po- tentially 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,

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 chlorophyl 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. Further- more, 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 exctinction 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. Further- more, 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 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 revolutionalize 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 semicon- ductor 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 quan- tum 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 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 diffaction 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 refered to as single electron transistors.

Early on, researchers realized that if one has a very small metal nanopar- ticle, its capacitance might be large enough to store discrete charges. Low- ering 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 knows 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.

Appendix 1. Crystal structure and Materials

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 A1.1.

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 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.

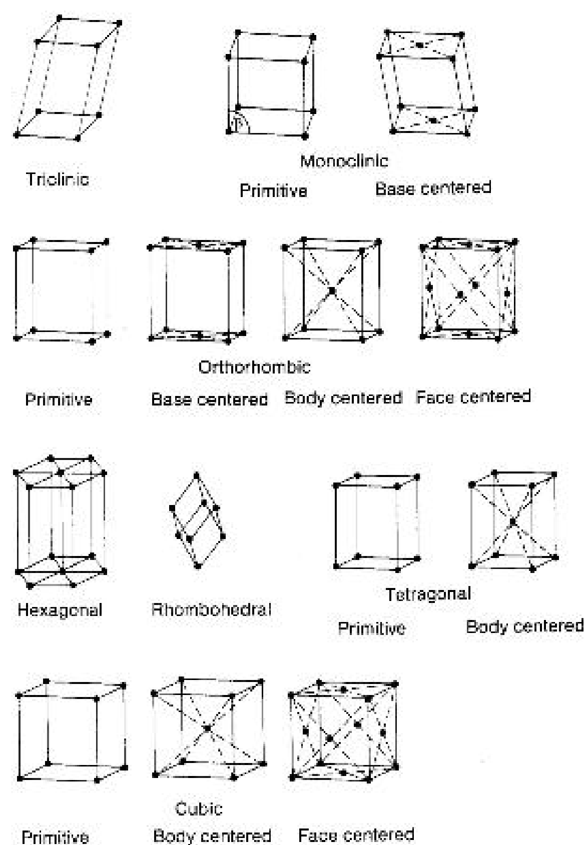


Figure A1.1: 14 3-dimensional Bravais lattices.

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 A1.2 and A1.3. 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.

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. The letters A, B, and C etc. represent different atom planes. The HCP structure has a conventional unit cell but also a primitive unit cell shown in Figure A1.4 It contains 2 atoms per unit cell (8 on the corners and 1 inside) as opposed to the conventional cell which has 12 per cell.

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 im mediately resemble any of the 14 Bravais lattices. It is adopted by elements that have a tendency to form strong covalent bonds, resulting in tetrahedral bonding arrangements (Figure A1.5). 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.

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. 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.

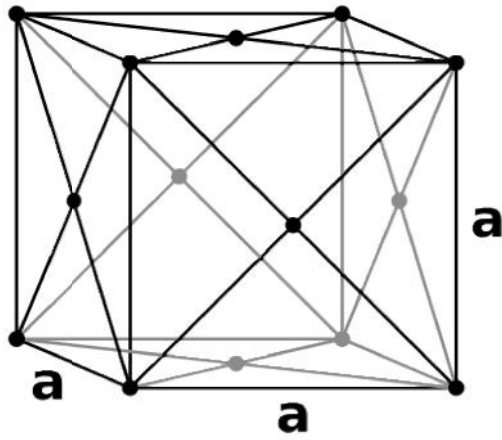


Figure A1.2 FCC unit cell

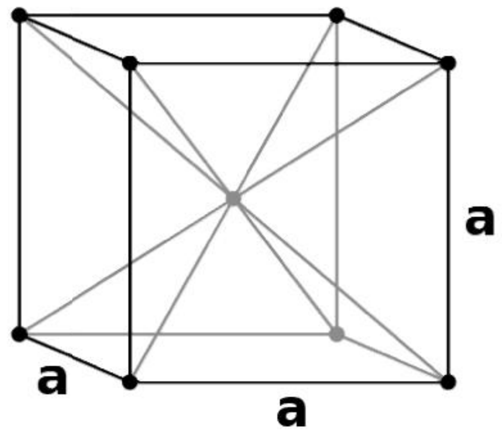


Figure A1.3 BCC unit cell

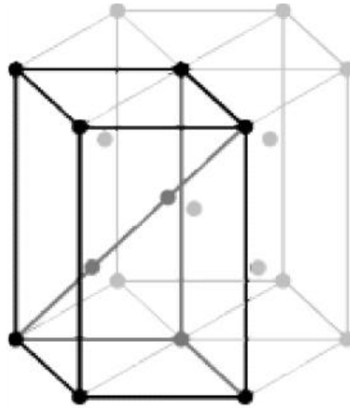


Figure A1.4 Primitive hexagonal unit cell.

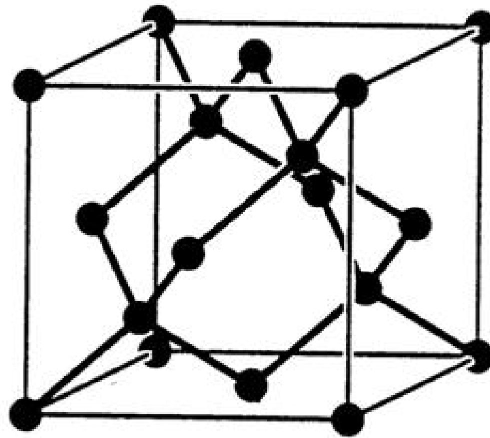


Figure A1.5 Diamond structure unit cell

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 (1/4 1/4 1/4) in Figure A1.6. 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. 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.

The NaCl structure can be visualized as 2 interpenetrating FCC lattices offset by (1/2 , 0, 0) in Figure A1.7. 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.

The CsCl structure is the compound material version of the single element BCC unit cell. It is shown in Figure A1.8 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 CsCl has two atoms per unit cell, 1 from each element.

The wurtzite crystal structure is the compound material version of the single element HCP structure. It has a multi atom basis. The primitive unit cell is shown in Figure A1.9 and contains 4 atoms per unit cell, 2 atoms from element 1 and 2 atoms from element 2.

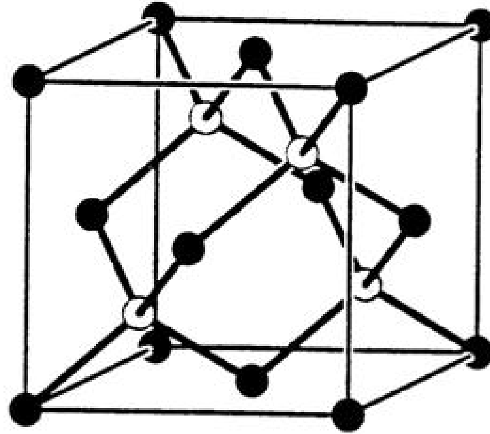


Figure A1.6. Zincblende or ZnS structure unit cell.

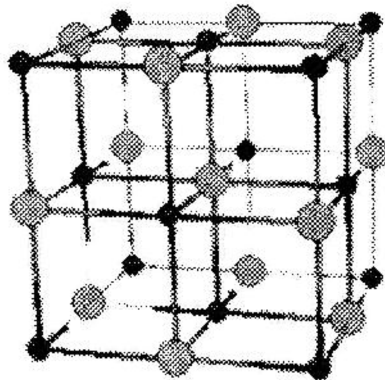


Figure A1.7 NaCl structure unit cell.

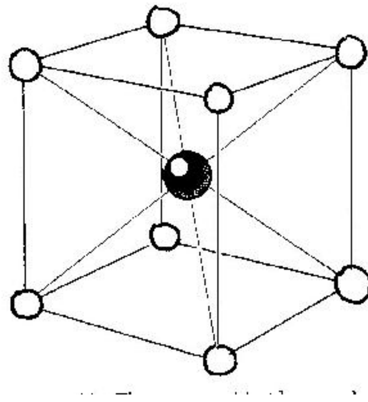


Figure A1.8 CsCl structure unit cell.

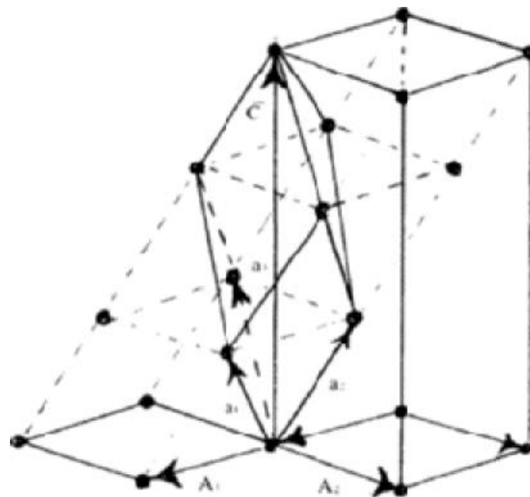


Figure A1.9 Primitive wurtzite unit cell.

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*)
- 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 $\left(a\bar{b}c\right)$.

Examples are illustrated in Figures A1.10 and A1.13.

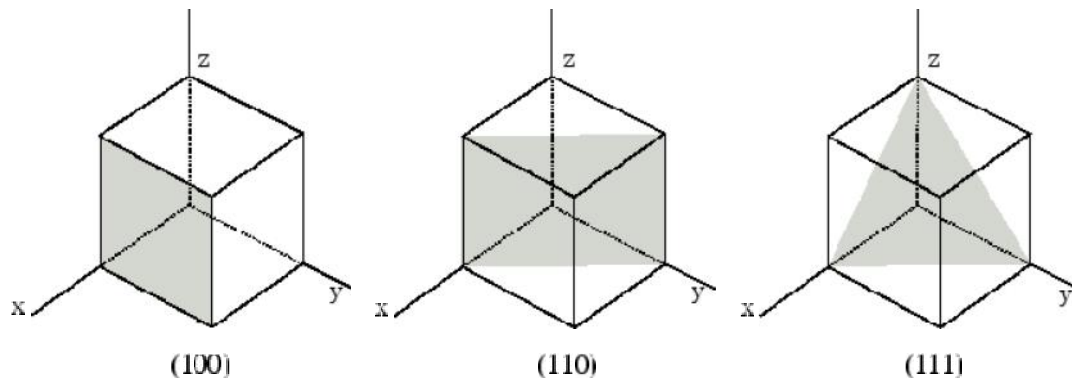


Figure A1.10: Examples of using Miller indices

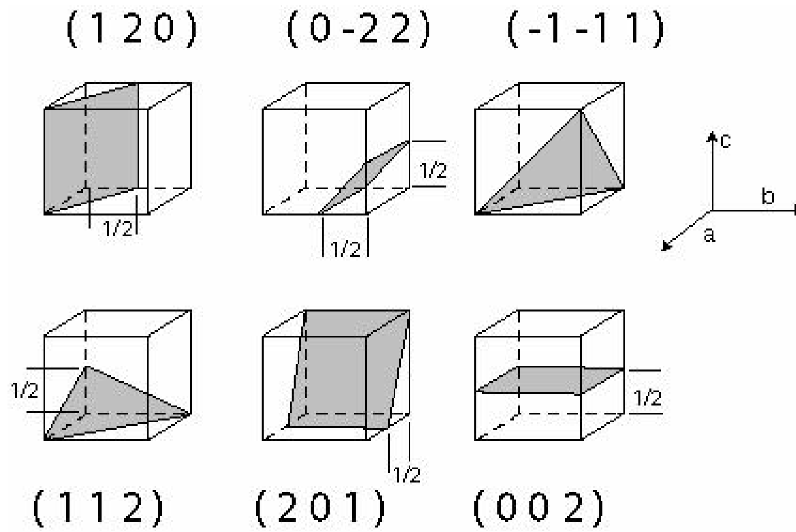


Figure 2.13: More examples of using Miller indices.

.

Common Metals

Table A1.1: 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 A1.2: 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

Group III-V

Table A1.3: 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

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

Group II-VI

Table A1.4: 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]

Group IV-VI

Table A1.5: Group IV-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

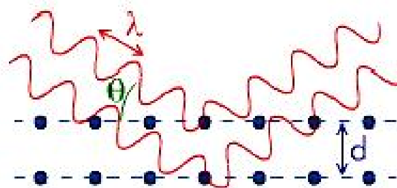
- PbS=NaCl structure
- PbSe=NaCl structure
- PbTe=NaCl structure

Appendix 2. Bragg's Law of Diffraction

The diffraction of X-ray from a crystal can also be easily understood in terms of which was first proposed by Bragg. In this model, an X-ray which reflects from the surface of a substance has traveled less distance than an X-ray which reflects from a plane of atoms inside the crystal. The penetrating X-ray travels down to the internal layer, reflects, and travels back over the same distance before being back at the surface. The distance traveled depends on the separation of the layers and the angle at which the X-ray entered the material. For this wave to be in phase with the wave which reflected from the surface it needs to have traveled a whole number of wavelengths while inside the material. Bragg expressed this in an equation now known as Bragg's Law:

Bragg's Law:

$$n \lambda = 2 d \sin(\theta)$$



where

λ is the wavelength of the rays

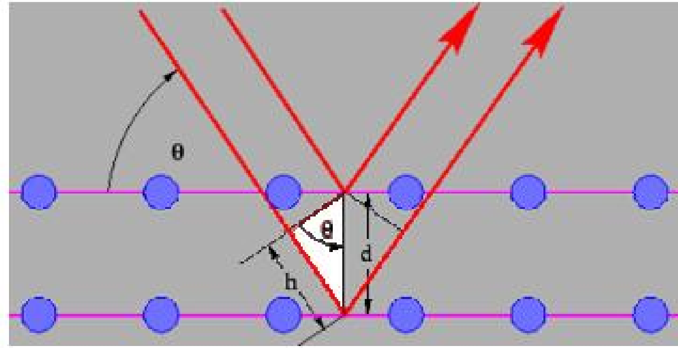
θ is the angle between the incident rays and the surface of the crystal

d is the spacing between layers of atoms

and constructive interference occurs when n is an integer (whole number)

When n is an integer (1, 2, 3 etc.) the reflected waves from different layers are perfectly in phase with each other and produce a bright point on a piece of photographic film. Otherwise the waves are not in phase, and will either be missing or faint.

Derivation of Bragg's Law:



Note that

$$h = d \sin \theta \qquad (A2.1)$$

so the path length difference between these two beams are

$$\Delta l = 2d \sin \theta \qquad (A2 .2)$$

The phase difference will be:

$$\Delta \phi = \left(2d \sin \theta\right)\left(2\pi / \lambda\right) \qquad (A2 .3)$$

When this phase difference is equal to $2\pi n$ (with n being an integer) we will have constructive interference. This leads to the Bragg equation:

$$\left(2d \sin \theta\right)\left(2\pi / \lambda\right) = 2\pi n \qquad (A2 .4)$$

or

$$2d_{hkl} \sin \theta = n\lambda \qquad (A2 .5)$$

In general, we only need to consider the first order diffraction (n=1) as diffractions of high orders can be considered as the first order diffractions from planes with an interplaner spacing of d_{hkl}/n. For example, the second order diffraction from (100) planes can be considered as the first order diffraction from (200) planes.

Appendix 3. Review of Quantum Mechanic

In Classical Mechanics, a state of a particle's motion is specified by its position and momentum. In Quantum Mechanics, the state of motion for a particle is not specified by its position and momentum. In fact, the position and momentum cannot be precisely determined simultaneously. Instead, the state of motion for a quantum particle is described by a wavefunction (or state function) which extends to a large region of space, and can be a complex function. Typically, a (time independent) wavefunction of a stationary state for a quantum particle is written as

$$\Psi_k\left(\vec{r}\right)$$

where k represents one set of so-called quantum numbers, usually discrete. Examples of quantum numbers are linear momentum, angular momentum, etc. Different set of quantum numbers, say, k1, k2, …, represent different wavefunction

$$\Psi_{k_1}\left(\vec{r}\right), \Psi_{k_2}\left(\vec{r}\right)...$$

which correspond to different states of the particle's motion. Therefore, one can use these discrete set of quantum numbers to characterize state of the particle's motion. A state function suc as $\Psi_k\left(\vec{r}\right)$ cannot be measured directly. It has a meaning

of probability: its modular $\left|\Psi_k\left(\vec{r}\right)\right|^2$ gives the spatial distribution of a particle's position in a state with quantum number k. Hence, in this representation of a quantum state, the quantum number (e.g., momentum) is known precisely, but particle's position is unknown (known by a probability distribution).

Wavefunction should satisfy the following conditions:

1. Since $\left|\Psi_k\left(\vec{r}\right)\right|^2$ has the meaning of distribution, it must be normalizable

$$\int\left|\Psi_k\left(\vec{r}\right)\right|^2 d^3r = 1 \tag{A3.1}$$

2. Linear superposition of wavefunctions also describes the particle possible states. If $\Psi_{k_1}\left(\vec{r}\right), \Psi_{k_2}\left(\vec{r}\right)$ are two possible states of a particle, their linear summation

$$\Psi\left(\vec{r}\right) = C_1\Psi_{k_1}\left(\vec{r}\right) + C_2\Psi_{k_2}\left(\vec{r}\right) \tag{A3.2}$$

is also a state of the particle. In the above (A3.2) C_1 and C_2 are any two complex numbers.

In quantum mechanics, all observables become operator, represented by headed notation, e.g., $\hat{A}$. An operator is meant to act the function describing particle states, and the result is another function also corresponding to particle state.

The experimentally measurable quantity is the so-called expectation value of $\hat{A}$ is a mean value

$$\left\langle \hat{A} \right\rangle = \int \Psi^* \hat{A} \Psi \, d^3r \tag{A3.3}$$

In so called coordinate representation the following correspondence takes place between classical quantities and quantum operators (correspondence principle):

$$\begin{aligned} \vec{r} &\rightarrow \hat{r} = r \\ \vec{p} &\rightarrow \hat{p} = -i\hbar\vec{\nabla} \end{aligned} \tag{A3.4}$$

And in general

$$F\left(\vec{r}, \vec{p}\right) \rightarrow \hat{F} = F\left(\vec{r}, -i\hbar\vec{\nabla}\right) \tag{A3.5}$$

For example, the expectation value of position and momentum of a particle at in state $\Psi\left(\vec{r}\right)$ are

$$\left\langle \hat{r} \right\rangle = \int \Psi^* \hat{r} \Psi \, d^3r = \int \vec{r} \, d^3r \left|\Psi\right|^2$$

$$\left\langle \hat{p} \right\rangle = \int \Psi^* \hat{p} \Psi \, d^3r = -i\hbar \int \Psi^* \vec{\nabla} \Psi \, d^3r \tag{A3.6}$$

According to a correspondence principle, the Hamiltonian operator of a particle is given by

$$\hat{H} = \frac{1}{2m}\hat{p}^2 + V\left(\vec{r}\right) = -\frac{\hbar^2}{2m}\nabla^2 + V\left(\vec{r}\right)$$

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{A3.7}$$

Hamiltonian operator corresponds to total energy of a system. In stationary condition Hamilton operator does not depend on time explicitly, and wave function and corresponding energies can be found my means of Schrödinger equation

$$\hat{H}\Psi = E\Psi \tag{A3.9}$$

i.e

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V\left(\vec{r}\right)\right]\Psi = E\Psi \tag{A3.10}$$

This is a problem of finding operator eigen vectors and eigen values.

As it is known form linear algebra, if two operators $\hat{A}$ and $\hat{B}$ commute with each another,

$$\hat{A}\hat{B} - \hat{B}\hat{A} = 0 \tag{A3.11}$$

they have got the same eigen vectors. In quantum mechanics we say that, if two operator commute with each another, the values of physical quantities corresponding to them, can be measured simultaneously.

In stationary condition physical quantity is measurable, if corresponding operator commutes with Hamilton operator. Wave function depends on quantum numbers, each of which corresponds to one of the commuting with Hamiltonian operator.

Appendix A: k.p Method

A.1 Band Theory: Bloch Functions

A solid consists of many atoms and electrons. The total energy of the system is therefore the sum of the kinetic energies of all the nuclei and electrons, the potential energy due to nuclear forces, the potential energies of electrons in the field of nuclei, the potential energy due to electron–electron interactions, and the magnetic energy associated with the spin and the orbit. The total Hamiltonian of the system may be constructed accordingly. The formidable problem of solving the resultant Schrödinger equation is bypassed by introducing several approximations. Since the motion of nuclei is sluggish, the electrons instantaneously adjust their motion to that of the ions. The total wavefunction is then written as a wavefunction for ions $\phi(R)$ and that for all electrons $\psi(r, R)$ instantaneously dependent on all ionic positions $\mathbf{R}$. An approximation, known as the *adiabatic approximation*, is introduced to decouple the Schrödinger equation into a purely ionic and a purely electronic equation, which are expressed, respectively, as

$$H_L\phi(\mathbf{R}) = E_L\phi(\mathbf{R}) \tag{A.1}$$

and

$$H_e\psi(\mathbf{r}, \mathbf{R}) = E_e\psi(r, R) \tag{A.2}$$

where $\mathbf{r}$ denotes the electronic coordinates [1, 2].

The electron potential energy is due to electron–electron and electron–ion interactions. If a suitable average is found for the first, a constant repulsive contribution can be added to the electron energy and then each electron becomes independent. The one-electron Schrödinger equation then takes the form

$$H_{ei}\psi_i(\mathbf{r_i}, \mathbf{R}) = E_{ei}\psi_i(\mathbf{r_i}, \mathbf{R}) \tag{A.3}$$

where

$$H_{ei} = \frac{\mathbf{p}_i^2}{2m_0} + \sum_i V(\mathbf{r_i}, \mathbf{R}_i) \tag{A.4}$$

Silicon Photonics: Fundamentals and Devices, First Edition. M. Jamal Deen and P. K. Basu.
© 2012 John Wiley & Sons, Ltd. Published 2012 by John Wiley & Sons, Ltd.

and $\boldsymbol{p}_i$ is the momentum of the ith electron. The Hamiltonian still depends on the fluctuating position of the ion. In the next approximation, the ions are assumed to lie in their equilibrium position and the effect of ionic vibration is taken as a perturbation. Thus the problem is reduced to solving the equation

$$\frac{\mathbf{p}^2}{2m_0} + \sum_i V(\mathbf{r} - \mathbf{R_{10}})\psi(\mathbf{r}) = E\psi(\mathbf{r}) \tag{A.5}$$

The ionic potential V is periodic and the eigenfunctions are Bloch functions expressed as

$$\psi_{n\mathbf{k}}(\mathbf{r}) = U_{nk}(r) \exp(j\mathbf{k}.\mathbf{r}), \tag{A.6}$$

where the cell periodic part $U(\mathbf{r})$ obeys the relation

$$U_{nk}(\mathbf{r} + \mathbf{R}) = U_{nk}(\mathbf{r}). \tag{A.7}$$

In the above equations, $\mathbf{R}$ is a vector of the Bravais lattice, n denotes the band index, and $\boldsymbol{k}$ is a wave vector of the electron in the first Brillouin zone. From Eqs. (A.6) and (A.7)

$$\psi_{nk}(\mathbf{r} + \mathbf{R}) = \psi_{nk}(\mathbf{r}) \exp(j\mathbf{k}.\mathbf{R})$$

The Bloch functions are eigenfunctions of the one-electron Schrödinger equation and therefore they are orthogonal to one another. Thus

$$\int \psi_{n'k'}\psi_{nk} d^3r = \delta_{n',n}\delta_{k'k} \tag{A.8}$$

The wavefunctions are also normalized over the volume V of the crystal and therefore

$$\psi_{nk} = V^{-1/2} U_{nk}(r) \exp(j\mathbf{k}.\mathbf{r}) \tag{A.9}$$

A.2 The k.p Perturbation Theory Neglecting Spin

Complete knowledge of the band structure of a semiconductor requires that the full E–$\boldsymbol{k}$ dispersion relation be known completely. The Schrödinger equation should therefore be solved completely. This is a rather difficult task since the form of the periodic potential $V(\mathbf{r})$ must be specified. Fortunately, in most descriptions of the electron and hole properties in semiconductors, the mostly populated electron and hole states lie within a fraction of an eV from the band edges. Thus, if the wavefunctions and energies of the carriers are known at the band extrema, then perturbation methods may be applied to find out the wavefunctions and energies at other points in the Brillouin zone, leading to knowledge of the E–$\boldsymbol{k}$ relationship. The method, known as ***k.p*** *perturbation theory*, is most widely used in the study of transport and optical processes in common semiconductors [3–9].

The ***k.p*** perturbation theory is based on the fact that the cell periodic part U_k of the electrons, for any value of k but different bands, forms a complete set. Let us consider the wavefunctions for the electrons having a value $\boldsymbol{k}$ near the minima in the nth band. For simplicity we assume that the minima are located at $\boldsymbol{k} = 0$. The theory is applicable also when the minima are located at $\boldsymbol{k} = \boldsymbol{k}_0$. The wavefunction is given by

$$\psi = U_{nk}(\mathbf{r}) \exp(j\mathbf{k}.\mathbf{r}) = \left[\sum_m c_m U_{m0}(\mathbf{r})\right] \exp(j\mathbf{k}.\mathbf{r}), \tag{A.10}$$

since U_{m0} forms a complete orthonormal set. Using this form of ψ in the Schrödinger equation, one obtains

$$\left[-\frac{\hbar^2}{2m_0}\nabla^2 + \frac{\hbar^2}{m_0}\mathbf{k}.\mathbf{p} + \frac{\hbar^2 k^2}{2m_0} + V(\mathbf{r})\right] U_{nk}(\mathbf{r}) = E_n(\mathbf{k})U_{nk}(\mathbf{r}). \tag{A.11}$$

However, U_{m0} is the wavefunction for $\boldsymbol{k} = 0$ in the nth band satisfying the equation

$$\left[-\frac{\hbar^2}{2m_0}\nabla^2 + V(\mathbf{r})\right] U_{m0}(\mathbf{r}) = E_m(0)U_{m0}(\mathbf{r}). \tag{A.12}$$

We now put Eq. (A.10) in (A.11) and use (A.12) to obtain

$$\sum c_m \left[E_m(0) + \frac{\hbar^2}{2m_0}k^2 + \frac{\hbar^2}{m_0}\mathbf{k}.\mathbf{p}\right] U_{m0}(\mathbf{r}) = \sum_{\mathbf{m}} c_m E_n(\mathbf{k})U_{m0}(\mathbf{r}). \tag{A.13}$$

Multiplying both sides of Eq. (A.13) by $U^*_{l0}(\mathbf{r})$ and integrating over a volume of a unit cell (V_c), the following set of linear homogeneous equations is obtained:

$$c_l\left[E_n(\mathbf{k}) - E_l(0) - \frac{\hbar^2}{2m_0}k^2\right] - \sum c_m \frac{\hbar}{m_0}(\mathbf{k}.\mathbf{p}_{lm}) = 0, \tag{A.14}$$

where

$$\mathbf{p}_{lm} = \int_{V_c} U^*_{l0}(\mathbf{r})\, \mathbf{p}\, U_{m0}(\mathbf{r}) d^3r. \tag{A.15}$$

By giving l successive integer values, one obtains the full set of equations.

In the general case, the set of equations has a nontrivial solution if the determinant of the coefficients c_l is zero. This condition gives the energy eigenvalues $E_n(\boldsymbol{k})$ in terms of the quantities $E_m(0)$ and p_{lm}. The relative values of the expansion coefficients c_m are then obtained by using the values of $E_n(\boldsymbol{k})$. The absolute values of c_m are obtained by imposing normalization conditions on ψ. The accuracy of the calculation is increased if many such coefficients are included. For practical reasons we need to limit ourselves to a few bands. The bands of greatest interest in common semiconductors are conduction (C), heavy hole (HH), light hole (LH), and split-off (SO) bands. Each of these four bands has two spin components, so there are altogether eight bands. Depending on the problem at hand and the degree of accuracy required, we may use some or all of these eight bands. We present below the results by using different approximations regarding the number of bands.

A.2.1 Single-Electron Band

Let us assume that U_{nk} is determined mostly by U_{n0} and the contributions from other bands are smaller. In other words, we assume that $c_m(m \neq n) \ll c_n$. Then

$$\psi = \left[c_n U_{n0}(\mathbf{r}) + \sum c_m U_{m0}(\mathbf{r})\right] \exp\,(j\mathbf{k}.\mathbf{r}). \tag{A.16}$$

Since ψ is normalized, $\sum_m |c_m|^2 = 1$; but we have assumed $c_m \ll c_n$. It follows, therefore, that $c_n \approx 1$.

To solve Eq. (A.16), stationary perturbation theory is applied. First, neglect c_m $(m \neq n)$ in comparison to c_n in the nth equation. The result is

$$E_n(\mathbf{k}) = E_n(0) + \frac{\hbar^2 k^2}{2m_0} + \frac{\hbar}{m_0}\mathbf{k}.\mathbf{p_{nn}}. \tag{A.17}$$

Next put $c_n = 1$ and obtain from Eq. (A.14)

$$c_m \approx \frac{\hbar}{m_0}\frac{\mathbf{k}.\mathbf{p}_{mn}}{E_n(0) - E_m(0)} c_n, \tag{A.18}$$

neglecting the $\boldsymbol{k.p}$ term in the denominator. If this expression for c_m is now used in Eq. (A.14), a second-order approximation results and one obtains

$$E_n(\mathbf{k}) = E_n(0) + \frac{\hbar^2 k^2}{2m_0} + \frac{\hbar}{m_0}\mathbf{k}.\mathbf{p}_{nn} + \sum_{m \neq n}\left(\frac{\hbar}{m_0}\right)^2 \frac{|\mathbf{k}.\mathbf{p}_{nn}|^2}{E_n(0) - E_m(0)}. \tag{A.19}$$

Since the extrema occur at $\boldsymbol{k} = 0$, $\boldsymbol{p}_{\mathrm{nn}} = 0$. Therefore, by choosing a proper coordinate system, one may write

$$E_n(\mathbf{k}) = E_n(0) + \frac{\hbar^2 k^2}{2m_i},$$

where, from Eq. (A.19),

$$\frac{1}{m_i} = \frac{1}{m_0} + \frac{2}{m_0{}^2}\sum_m \frac{|\mathbf{i}.\mathbf{p}_{\mathrm{nm}}|^2}{E_n(0) - E_m(0)}, \tag{A.20}$$

where $\mathbf{i}$ is a unit vector along the ith coordinate axis. The above equation predicts a parabolic E–$\boldsymbol{k}$ relation.

The analysis presented in this subsection may be improved by combining with the band under consideration other bands close to it and treating the effects of the additional bands as small perturbations.

A.2.2 Four Bands

We now consider four bands: the C, HH, LH, and SO bands. The valence bands are triply degenerate, and the C-band minima and H-band maxima occur at $\boldsymbol{k} = 0$. For the present, the spin–orbit interaction is neglected. We denote the cell periodic parts of the conduction band by U_c and those of the three valence bands by U_{v1}, U_{v2}, U_{v3}. Also the symbols E_c and E_v are used to denote, respectively, the energy for conduction band minima and valence band maxima. We may write the wavefunction for any $\boldsymbol{k}$ as

$$\psi = (a_k U_c + b_k U_{v1} + c_k U_{v3} + d_k U_{v2}) \exp(j\mathbf{k}.\mathbf{r}), \tag{A.21}$$

in accordance with Eq. (A.10). Using the symbol $E' = E - \hbar^2 k^2/2m_0$, the linear homogeneous equations are

$$\begin{aligned}
&a_k(E' - E_c) - (\hbar/m_0)\mathbf{k}.(b_k\mathbf{p}_{cv1} + d_k\mathbf{p}_{cv2} + c_k\mathbf{p}_{cv3}) = 0\\
&-a_k(\hbar/m_0)\mathbf{k}.\mathbf{p_{cv1}} + b_k(E' - E_v) - (\hbar/m_0)\mathbf{k}.(d_k\mathbf{p}_{v1v2} + c_k\mathbf{p}_{v1v3}) = 0\\
&-a_k(\hbar/m_0)\mathbf{k}.\mathbf{p_{cv2}} - b_k(\hbar/m_0)\mathbf{k}.\mathbf{p}_{v1v2} - d_k(E' - E_v) - (\hbar/m_0)\mathbf{k}. = 0\\
&-a_k(\hbar/m_0)\mathbf{k}.\mathbf{p_{cv3}} - b_k(\hbar/m_0)\mathbf{k}.\mathbf{p}_{v1v3} - d_k(\hbar/m_0)\mathbf{k}. + c_k(E' - E_v) = 0
\end{aligned} \tag{A.22}$$

The quantities p_{cv1}, p_{v1v2}, and so on are defined in Eq. (A.15). The matrix elements may be evaluated once the Us are known. Since U_c is an atomic s-like function and the U_v' s are p-like functions, the $\boldsymbol{k}.\boldsymbol{p}_{cv1}$ term may be expressed as

$$h_1 = \frac{\hbar^2}{m_0 j}\int U_t^*\left(k_x\frac{\partial}{\partial x}+k_y\frac{\partial}{\partial y}+k_z\frac{\partial}{\partial z}\right)U_m d^3r \quad t,m = s,x,y,z. \tag{A.23}$$

Since $(\partial/\partial x)U_s$ is an odd function of x, the matrix element $\int U_j^*\,(\partial/\partial x)U_s\,dx$ is nonzero only when $j=x$. The same is true for $\int U_s^*\,(\partial/\partial x)U_j dx$. The function $(\partial/\partial x)\,U_y$ is odd in both x and y. Thus

$$\iint U_j^*\frac{\partial}{\partial x}U_y dxdy = 0, \quad j = s,x,y,z.$$

The only nonvanishing matrix elements are the ones defined by the following expressions:

$$P = -\frac{\hbar^2}{m_0}\int U_j\frac{\partial}{\partial j}U_s d^3r = -\frac{\hbar^2}{m_0}\int U_s\frac{\partial}{\partial j}U_j d^3r, \quad j = x,y,z. \tag{A.24}$$

In the following, we shall assume that $\boldsymbol{k}$ is parallel to the z direction. Then we may rewrite the four homogeneous equations (A.22) as

$$\begin{aligned} &a_k(E'-E_c)-c_kPk = 0,\\ &b_k(E'-E_v) = 0,\\ &-a_kPkc_k(E'-E_v) = 0,\\ &d_k(E'-E_v) = 0. \end{aligned} \tag{A.25}$$

The energy eigenvalues are thus given by

$$E' = E_c, E_V \text{ and } (E'-E_c)(E'-E_v)-P^2k^2 = 0. \tag{A.26}$$

These equations give the dispersion relations when the conduction bands and valence bands are strongly coupled. Denoting the energy gap by $E_g = E_c - E_v$, we may write Eq. (A.26) as

$$(E'-E_c)(E'-E_c+E_g)-P^2k^2 = 0.$$

When E' tends to E_c, we may neglect E'–E_c in comparison to E_g and write

$$(E'-E_c)E_g = P^2k^2. \tag{A.27}$$

If the band edge effective mass is denoted by m_{e0}, we obtain

$$P^2 = \left(\frac{E-E_c-\hbar^2k^2}{2m_0}\right)\frac{E_g}{k^2} = \hbar^2\left(\frac{1}{m_{e0}}-\frac{1}{m_0}\right)\frac{E_g}{2}. \tag{A.28}$$

A.3 Spin–Orbit Interaction

The electron is a fermion with spin $^1/_2$ in units of $\hbar$. In classical mechanics, a point particle rotating about an axis has an angular momentum $\mathbf{L} = \mathbf{r} \times \mathbf{p}$. In the quantum picture, the angular momentum of a point particle is quantized and the intrinsic value of the momentum is called the *spin*. There is strong interaction between the spin and orbital motion of the electrons. This spin–orbit coupling may be calculated for isolated atoms; however, it is difficult to do so in crystals.

A.3.1 Spin–Orbit Interaction Term

To calculate the interaction a general form of spin–orbit interaction is assumed with a fitting parameter that is adjusted to fit experimentally observed effects. The total Hamiltonian in the presence of spin–orbit interaction is written as $H = H_0 + H_{so}$, where H_0 is the Hamiltonian without interaction and H_{so} is the spin–orbit interaction written as

$$H_{so} = \lambda \mathbf{L} \cdot \mathbf{S}. \tag{A.29}$$

Here $\boldsymbol{L}$ represents the operator for orbital angular momentum, $\boldsymbol{S}$ is the operator for spin angular momentum, and λ is treated as a constant. The total angular momentum $\boldsymbol{J}$ may be expressed as

$$\mathbf{J}^2 = (\mathbf{L} + \mathbf{S})^2 = L^2 + S^2 + 2\mathbf{L} \cdot \mathbf{S}. \tag{A.30a}$$

Thus

$$\langle \mathbf{L} \cdot \mathbf{S} \rangle = (1/2)\left\langle J^2 - L^2 - S^2 \right\rangle = \frac{\hbar^2}{2}[j(j+1) - l(l+1) - s(s+1)], \tag{A.30b}$$

where j, l, and s are the quantum numbers for the operators $\boldsymbol{J}$, $\boldsymbol{L}$, and $\boldsymbol{S}$, respectively. However, to calculate the spin–orbit interaction energy, one needs the pure angular momentum states to which Eq. (A.30) is applicable. One should note that states like $|X\alpha\rangle$ are mixed states, with the symbol α denoting the spin-up state. To illustrate this statement, we express $|X\rangle$ in terms of pure angular momentum states, that is,

$$\begin{aligned} |X\rangle &= \frac{1}{\sqrt{2}}(-\phi_{1,1} + \phi_{1,-1}) \\ |Y\rangle &= \frac{j}{\sqrt{2}}(\phi_{1,1} + \phi_{1,-1}) \\ |Z\rangle &= \phi_{1,0} \end{aligned} \tag{A.31}$$

The $\phi_{i,j}$'s are pure angular momentum states, and the expressions for the lower eigenstates are

$$\begin{aligned} \phi_{1,\pm1} &= Y_{1,\pm1}(\theta, \phi) = \mp\sqrt{\frac{3}{8\pi}} \sin\theta \exp(\pm j\phi) \\ \phi_{1,0} &= Y_{1,0}(\theta, \phi) = \sqrt{\frac{3}{4\pi}} \cos\theta \end{aligned}$$

The $\phi_{i,j}$'s are eigenfunctions of L^2 and L_z. The respective quantum numbers are $l=I$ and $l_z=j$. For example, $L^2\phi_{1,-1} = \hbar^2(1)(1+1)\phi_{1,-1} = 2\hbar^2\phi1, -1$ and $L_z\phi_{1,-1} = -1\hbar\phi_{1,-1}$.

Equation (A.31) is modified if spin is included; for example, the spin-up state $p_x = |X\alpha\rangle$ is expressed as

$$|X\alpha\rangle = \frac{1}{\sqrt{2}}(-\phi_{1,1} + \phi_{1,-1})\alpha.$$

This formulation is still in terms of mixed states. To decompose the mixed states into states of pure angular momentum, the spin and orbital angular momentum must be added to obtain the total angular momentum states. The standard Clebsch–Gordan (CG) technique is employed for this addition. The following six equations are obtained as a result:

$$\begin{aligned}
\phi_{3/2,3/2} &= \phi_{1,1}\alpha = (-1/\sqrt{2})|(X+jY)\alpha\rangle \\
\phi_{3/2,1/2} &= \frac{1}{\sqrt{3}}\phi_{1,1}\beta + \frac{2}{\sqrt{6}}\phi_{1,0}\alpha = \frac{-1}{\sqrt{6}}[|(X+jY)\beta\rangle - |2Z\alpha\rangle] \\
\phi_{3/2,-1/2} &= \frac{1}{6}\phi_{1,0}\beta + \frac{1}{\sqrt{3}}\phi_{1,-1}\alpha = \frac{1}{\sqrt{6}}[|(X-jY)\alpha\rangle + |2Z\beta\rangle] \\
\phi_{3/2,-3/2} &= \phi_{1,-1}\beta = (1/\sqrt{2})|(X-jY)\beta\rangle \\
\phi_{1/2,1/2} &= \frac{-1}{\sqrt{3}}\phi_{1,0}\alpha + \frac{2}{\sqrt{6}}\phi_{1,1}\beta = \frac{-1}{\sqrt{3}}[|(X+jY)\beta\rangle + |Z\alpha\rangle] \\
\phi_{1/2,-1/2} &= \frac{-2}{\sqrt{6}}\phi_{1,-1}\alpha + \frac{1}{\sqrt{3}}\phi_{1,0}\beta = \frac{-1}{\sqrt{3}}[|(X-jY)\alpha\rangle - |Z\beta\rangle]
\end{aligned} \tag{A.32}$$

These six equations are inverted to find states like $\phi_{1,0}$, and from the resultant equations one gets states like $|X\alpha\rangle$, and so on.

$$\begin{aligned}
|X\alpha\rangle &= \frac{1}{\sqrt{2}}\left[-\phi_{3/2,3/2} + \frac{1}{\sqrt{3}}\phi_{3/2,-1/2} - \sqrt{\frac{2}{3}}\phi_{1/2,-1/2}\right] \\
|X\beta\rangle &= \frac{1}{\sqrt{2}}\left[-\frac{1}{\sqrt{3}}\phi_{3/2,1/2} - \frac{2}{\sqrt{3}}\phi_{1/2,1/2} - \phi_{3/2,-3/2}\right] \\
|Y\alpha\rangle &= \frac{j}{\sqrt{2}}\left[\phi_{3/2,3/2} + \frac{1}{\sqrt{3}}\phi_{3/2,-1/2} - \sqrt{\frac{2}{3}}\phi_{1/2,-1/2}\right] \\
|Y\beta\rangle &= \frac{j}{\sqrt{2}}\left[\frac{1}{\sqrt{3}}\phi_{3/2,1/2} + \frac{2}{\sqrt{3}}\phi_{1/2,1/2} + \phi_{3/2,-3/2}\right] \\
|Z\alpha\rangle &= \sqrt{\frac{2}{3}}\phi_{3/2,1/2} - \frac{1}{\sqrt{3}}\phi_{1/2,1/2} \\
|Z\beta\rangle &= \sqrt{\frac{2}{3}}\phi_{3/2,-1/2} + \frac{1}{\sqrt{3}}\phi_{1/2,-1/2}.
\end{aligned} \tag{A.33}$$

The phases used in the above expressions for ϕ_{j,m_j} in terms of $|X\alpha\rangle, ...|Z\beta\rangle$ are obtained in the standard derivation of Clebsch–Gordan coefficients. The overall phase of a state is

arbitrary and has no effect on the physical predictions. The convention used by Luttinger and Kohn [10] is in widespread use and will be used here. The states are expressed in terms of CG states as

$$\begin{aligned}
\phi_{3/2,3/2}(\mathrm{LK}) &= -\phi_{3/2,3/2}(\mathrm{CG}) = \frac{1}{\sqrt{2}}|(X+jY)\alpha\rangle \\
\phi_{3/2,1/2}(\mathrm{LK}) &= -j\phi_{3/2,1/2}(\mathrm{CG}) \\
\phi_{3/2,-1/2}(\mathrm{LK}) &= \phi_{3/2,-1/2}(\mathrm{CG}) \\
\phi_{3/2,-3/2}(\mathrm{LK}) &= j\phi_{3/2,-3/2}(\mathrm{CG}) \\
\phi_{1/2,1/2}(\mathrm{LK}) &= -\phi_{1/2,1/2}(\mathrm{CG}) \\
\phi_{1/2,-1/2}(\mathrm{LK}) &= j\phi_{1/2,-1/2}(\mathrm{CG})
\end{aligned} \tag{A.34}$$

The spin–orbit Hamiltonian may be calculated now with the above states. The interaction is

$$H_{so} = \frac{\lambda\hbar^2}{2}[\,j(j+1) - l(l+1) - s(s+1)] \tag{A.35}$$

For p-type electron orbitals $l=1$ and $s=1/2$, j is given by the first subscript of ϕ in Eq. (A.33). Many terms become zero as the pure states are orthogonal. We conclude that only the following terms are nonzero:

$$\begin{aligned}
&\langle X\alpha|H_{so}|Y\alpha\rangle = \langle Y\alpha|H_{so}|Z\beta\rangle = \langle Y\beta|H_{so}|Z\alpha\rangle = -j\frac{\Delta}{3} \\
&\langle X\alpha|H_{so}|Z\beta\rangle = \frac{\Delta}{3};\ \langle X\beta|H_{so}|Z\alpha\rangle = -\frac{\Delta}{3};\ \langle X\beta|H_{so}|Y\beta\rangle = j\frac{\Delta}{3},
\end{aligned} \tag{A.36}$$

where Δ is a parameter known as *spin–orbit splitting* given by $\Delta = \Delta_{so} = 3\lambda\hbar^2/2$.

A.3.2 Conduction Band Energy

The calculation of energy levels in the conduction band by including the spin–orbit interaction is easier, and therefore we consider it first. As mentioned previously, we are interested in four basic vectors, the $|S\rangle$ state for the conduction band and the $|X\rangle.|Y\rangle, |Z\rangle$ states for the valence bands. There are four coefficients, $a_k \ldots .d_k$ as in Eq. (A.21), needed to describe a state. With the inclusion of spin, the number of basis vectors to be considered becomes eight. The secular equation containing the coefficients and basis vectors involves an 8×8 matrix. It turns out that if the basis vectors are arranged in the following manner,

$$|S\alpha\rangle.|(X+jY)\beta\rangle, |Z\alpha\rangle, |(X-jY)\beta\rangle, |S\beta\rangle, |(X-jY)\alpha\rangle, |Z\rangle, \text{ and } |-(X+jY)\alpha\rangle,$$

the matrix may be written in the form

$$\begin{bmatrix} H & 0 \\ 0 & H \end{bmatrix},$$

where H is a 4×4 matrix. Using the matrix elements between different states and the earlier elements obtained without the spin–orbit interactions, we may express the different

matrix elements in terms of P in the following form:

$$\begin{array}{ccccc} & |S\alpha\rangle & |(X+jY)\beta\rangle & |Z\alpha\rangle & |(X-jY)\beta\rangle \\ |S\alpha\rangle & E_s-E' & 0 & -jkP & 0 \\ |(X+jY)\beta\rangle & 0 & E_p-E'-\Delta/3 & \sqrt{2}\Delta/3 & 0 \\ |Z\alpha\rangle & jkP & \sqrt{2}\Delta/3 & E_p-E' & 0 \\ |(Xxx-jY)\beta\rangle & 0 & 0 & 0 & E_p-E'+\Delta/3 \end{array} \tag{A.37}$$

To simplify the calculation, we choose the $\boldsymbol{k}$-vector in (A.37) along the z-direction. Furthermore, to account for the shift of band energies due to spin–orbit interaction, notations E_p and E_s are used. The difference in signs in $E_p-E'\pm\Delta/3$ is due to the fact that $L_zS_z|(X+jY)\beta\rangle = -|(X+jY)\beta\rangle$ while $L_zS_z|(X-jY)\beta\rangle = |(X-jY)\beta\rangle$. Expanding (A.37) one obtains

$$E' = E_p + \Delta/3 \tag{A.38a}$$

$$\left(E'-E_p+\frac{2\Delta}{3}\right)\left(E'-E_p-\frac{\Delta}{3}\right)(E'-E_s)-k^2P^2\left(E'-E_p+\frac{\Delta}{3}\right)=0 \tag{A.38b}$$

For small values of k^2, the cubic equation can easily be solved by treating the term k^2P^2 as a small perturbation. This yields

$$E_1' = E_s + \frac{k^2P^2(E_s-E_p+\Delta/3)}{(E_s-E_p+2\Delta/3)(E_s-E_p-\Delta/3)}. \tag{A.38c}$$

Let $E_v = E_p+\Delta/3 = 0, E_c = E_s = E_{g0}$ and $E_c-E_v = E_{g0}$, the direct gap. We then rewrite Eq. (A.38a) as

$$E'(E'-E_{g0})(E'+\Delta)-k^2P^2(E'+2\Delta/3)=0. \tag{A.39}$$

Taking $E_c = E_{g0}$ in the first approximation,

$$E_c(\mathbf{k}) = E_{g0} + \frac{\hbar^2k^2}{2m_0} + \frac{k^2P^2}{3}\left[\frac{2}{E_{g0}} + \frac{1}{E_{g0}+\Delta}\right]. \tag{A.40a}$$

We also obtain, by putting $E_v = E_p+\Delta/3 = 0$,

$$E_{v1}(\mathbf{k}) = \frac{\hbar^2k^2}{2m_0}. \tag{A.40b}$$

Taking $E=0$,

$$E_{v2} = \frac{\hbar^2k^2}{2m_0} - \frac{2k^2P^2}{3E_{g0}}. \tag{A.40c}$$

Finally, taking $E' = -\Delta$ as a first approximation,

$$E_{v3} = -\Delta + \frac{\hbar^2k^2}{2m_0} - \frac{k^2P^2}{3(E_{g0}+\Delta)}. \tag{A.40d}$$

In Eq. (A.40a), E_c is the energy of the conduction band electrons, while E_{v1}, E_{v2}, and E_{v3}, are, respectively, the energies of the three valence bands. From Eq. (A.40a), we may define the band edge effective mass for conduction band electrons by writing

$$E_c = E_{g0} + \hbar^2 k^2 / 2m_{e0}.$$

It then follows that

$$\frac{1}{m_{e0}} = \frac{1}{m_0} + \frac{2P^2}{3\hbar^2}\left(\frac{2}{E_{g0}} + \frac{1}{E_{g0}+\Delta}\right). \tag{A.41}$$

The momentum matrix element, P, which is central to all calculation of transition probabilities from the valence band to the conduction band, may be expressed in terms of m_{e0} as

$$P^2 = \frac{\hbar^2}{2m_{e0}} \frac{E_{g0}(E_{g0}+\Delta)}{E_{go}+2\Delta/3} \frac{m_0 - m_{e0}}{m_0}. \tag{A.42}$$

A.3.3 Valence Band Energies

The earlier treatment of the dispersion relation cannot explain the properties of electrons in the valence band. Referring to Eq. (A.40a), one notices that the energy of the electrons increases with k, which however is opposite to what is observed experimentally. To treat the valence band properties correctly, degenerate perturbation theory is needed [11].

Let the state of the electron in the lth band at $k = 0$ be degenerate, having f-fold degeneracy. It follows from the theory of perturbation of degenerate states that the second-order corrections $\Delta E^{(2)}$ due to the $(\hbar/m_0)(\mathbf{k}.\mathbf{p})$ perturbation are the roots of the secular equation

$$\left[\left(\frac{\hbar^2}{m_0}\right)^2 \sum_{n,s}{}' \frac{\langle l,r'|\mathbf{k}.\mathbf{p}|n,s\rangle\langle n,s|\mathbf{k}.\mathbf{p}|l,r\rangle}{E_l(0)-E_n(0)} - \Delta E^{(2)}\delta_{rr'}\right] = 0, \tag{A.43}$$

where the primed summation sign indicates that the summation is over all $n \neq l$ and over $s.|l,r\rangle$ and $|l,r'\rangle$ are the unperturbed f-fold degenerate wavefunctions (r, $r' = 1,2,\ldots f$) satisfying Eq. (A.43) for the energy eigenvalue $E_l(0)$. The$|n,s\rangle$'s are the wavefunctions for energy level $E_n(0)$. The order of the determinant of the secular equation is equal to the degree of degeneracy of the level $E_l\,(0)$. In the present situation, the degenerate eigenstates at $k = 0$ are the three $l = 1$, $m_l = 0$, and ± 1 states. To the second order in perturbation, the energy in the nth band is

$$E_n(k) = E_n(0) + \frac{\hbar^2 k^2}{2m_0} + \Delta E_n^{(2)}(k), \tag{A.44}$$

where $\Delta E_n^{(2)}$ is given by Eq. (A.43) above. Hence we have the set of three equations

$$\sum_{i=1}^{3}\left[\sum_m{}' \frac{\langle j|H_l|m\rangle\langle m|H_l|l\rangle}{E_l(0)-E_m(0)} + \left\{E_l + \frac{\hbar^2k^2}{2m_0} - E_n(k)\right\}\delta_{j,l}\right]\langle n|k\rangle = 0. \tag{A.45}$$

Nontrivial solutions of this set of N coupled homogeneous equations occur only if

$$\det[|H| - E_n(k)\mathbf{I}] = 0,$$

where $\mathbf{I}$ is the identity matrix and H is a 3×3 matrix whose elements are

$$H_{jl} = \left[E_l + \frac{\hbar^2k^2}{2m_0}\right]\delta_{jl} + \sum_m{}' \frac{\langle j|H_l|m\rangle\langle m|H_l|l\rangle}{E_l(0)-E_m(0)}.$$

The calculation of the matrix element is first made by ignoring the spin–orbit interaction for the present. We take the basis sets as $|X\rangle.|Y\rangle$ and $|Z\rangle$. Then

$$H_{11} = \langle X|H|X\rangle = E_1(0) + \frac{\hbar^2k^2}{2m_0} + \sum_m{}' \frac{|\langle X|H_1|m\rangle|^2}{E_l(0)-E_m(0)}. \tag{A.46}$$

Since $|X\rangle$ is proportional to $xf(\mathbf{r})$, we may verify that

$$\frac{m_0^2}{\hbar^2}|\langle X|H_l|m\rangle|^2 = |\langle X|p_x|m\rangle|^2k_x^2 + \left|\langle X|p_y|m\rangle\right|^2k_y^2 + |\langle X|p_z|m\rangle|^2k_z^2.$$

Thus we may write

$$H_{11} = E_1 + \sum_{j=x,y,z}\left[\frac{\hbar^2}{2m_0} + \frac{\hbar^2}{m_0^2}\sum{}' \frac{\left|\langle X|p_j|m\rangle\right|^2}{E_l - E_m}\right]k_j^2.$$

Due to symmetry at $k=0$,

$$\left|\langle X|p_y|m\rangle\right|^2 = |\langle X|p_z|m\rangle|^2.$$

Therefore,

$$H_{11} = E_1 + Ak_x^2 + B(k_y^2 + k_z^2), \tag{A.47a}$$

where

$$A = \frac{\hbar^2}{2m_0} + \frac{\hbar^2}{m_0{}^2}\sum_j{}' \frac{|\langle X|p_x|j\rangle|^2}{E_l - E_j}, \tag{A.47b}$$

$$B = \frac{\hbar^2}{2m_0} + \frac{\hbar^2}{m_0^2}\sum_j{}' \frac{\left|\langle X|p_y|j\rangle\right|^2}{E_l - E_j}. \tag{A.47c}$$

The remaining matrix elements can be evaluated in a similar way to give the following Hamiltonian matrix,

$$H = \begin{bmatrix} E_1 + Ak_x^2 + B(k_y^2 + k_z^2) & Ck_xk_y & Ck_xk_z \\ Ck_xk_y & E_1 + Ak_y^2 + B(k_x^2 + k_z^2) & Ck_yk_z \\ Ck_xk_z & Ck_yk_z & E_1 + Ak_z^2 + B(k_x^2 + k_y^2) \end{bmatrix} \tag{A.48}$$

where

$$C = \frac{\hbar^2}{m_0^2} \sum_j{}' \frac{\langle X|p_x|j\rangle\langle j|p_y|Y\rangle + \langle X|p_y|j\rangle\langle j|p_x|Y\rangle}{E_l - E_j}. \tag{A.49}$$

Let us now consider the effect of spin. As noted earlier in Eq. (A.30) for $j = 1/2$, $\langle \mathbf{L}.\mathbf{S}\rangle = -\hbar^2$, and for $j = 3/2$, $\langle \mathbf{L}.\mathbf{S}\rangle = +\hbar^2/2$. Thus the states are split by an amount Δ proportional to $(3/2)\hbar^2$, the doubly degenerate state with $j = 3/2$ moving up by $\Delta/3$ and the single $j = 1/2$ state moving down by $2\Delta/3$. Since in the designation $|ls; jm_j\rangle$, $l = 1$ and $s = 1/2$, we shall use only the symbol $|jm_j\rangle$ to denote the states. As noted in this appendix, the Hamiltonian for the spin–orbit coupling is diagonalized if the states are chosen according to Eq. (A.32).

We shall treat the $j = 3/2$ states and $j = 1/2$ states separately, since the splitting energy is large. The Hamiltonian matrix $|H|$ now becomes a 4×4 matrix for $j = 3/2$ states and a 2×2 matrix for $j = 1/2$ states. We may evaluate the matrix elements for the 4×4 matrix using Eq. (A.47). Thus

$$\begin{aligned} H_{11} &= \langle 3/2, 3/2|H|3/2, 3/2\rangle = (1/2)\langle (X + jY)\alpha|H|(X + jY)\alpha\rangle \\ &= (1/2)[\langle X\alpha|H|X\alpha\rangle + \langle Y\alpha|H|Y\alpha\rangle + j\langle X\alpha|H|Y\alpha\rangle - j\langle Y\alpha|H|X\alpha\rangle] \\ &= E_1 + \frac{A}{2}\left(k_x{}^2 + k_y{}^2\right) + \frac{B}{2}\left(k_x{}^2 + k_y{}^2 + 2k_z{}^2\right) \\ &= H_{44} \end{aligned} \tag{A.50}$$

where the symbols 1, 2, 3, and 4 are used in the order in which the states are written in Eq. (A.48). Similarly,

$$H_{12} = \frac{1}{2\sqrt{3}}\langle (X + jY)\alpha|H|[(X + jY)\beta - 2Z\alpha]\rangle = \frac{1}{\sqrt{3}}(H_{xz} - jH_{yz}).$$

Instead of computing matrix elements $\langle n|p|m\rangle$ from first principles, one replaces them with experimentally determined parameters called *Luttinger parameters*, defined as

$$\gamma_1 = -2m_0(A + 2B)/3\hbar^2, \quad \gamma_2 = -m_0(A - B)/3\hbar^2, \quad \gamma_3 = -m_0C/3\hbar^2. \tag{A.51}$$

In terms of Luttinger parameters,

$$H_{11} = E_1 - \frac{\hbar^2 k_z^2}{2m_0}(\gamma_1 - 2\gamma_2) - \frac{\hbar^2(k_x^2 + k_y^2)}{2m_0}(\gamma_1 + \gamma_2) \tag{A.52}$$

Since in measurements the parameters conform to holes, and since the hole energy is positive, we write $H_{11} = -H_{hh}$ and the zero energy reference is $E_1 = 0$. Repeating the above

calculation for the other matrix elements, we obtain for the Luttinger Hamiltonian,

$$H = \begin{bmatrix} H_{hh} & -c & -b & 0 \\ -c* & H_{lh} & 0 & b \\ -b* & 0 & H_{lh} & -c \\ 0 & b* & -c* & H_{hh} \end{bmatrix} \tag{A.53}$$

where

$$\begin{aligned} H_{lh} &= \frac{\hbar^2 k_z^2}{2m_0}(\gamma_1 + 2\gamma_2) - \frac{\hbar^2 (k_x^2 + k_y^2)}{2m_0}(\gamma_1 - \gamma_2) \\ c &= \frac{\sqrt{3}\hbar^2}{2m_0}\left[\gamma_2 (k_x^2 - k_y^2) - 2j\gamma_3 k_x k_y\right] \\ b &= \frac{\sqrt{3}\hbar^2}{m_0}\gamma_3 k_z (k_x - jk_y). \end{aligned} \tag{A.54}$$

In the vicinity of $k = 0$, one may use the axial approximation, where γ_2 and γ_3 are replaced by an effective Luttinger parameter,

$$\bar{\gamma} = (1/2)(\gamma_2 + \gamma_3) \tag{A.55}$$

The function c is then expressed as

$$c \cong \frac{\sqrt{3}\hbar^2\bar{\gamma}}{2m_0}\left(k_x - jk_y\right)^2 \tag{A.56}$$

The dispersion relation for valence band holes may be written as

$$\begin{aligned} E_v &= -\frac{\hbar^2}{2m_0}\left[Ak^2 \pm \left\{B^2k^4 + C^2(k_x^2k_y^2 + k_x^2k_z^2 + k_y^2k_z^2\right\}^{1/2}\right] \\ &= \frac{\hbar^2}{2m_0}\left[-\gamma_1 k^2 \pm \left\{4\gamma_2^2 k^4 + 12(\gamma_3^2 - \gamma_2^2)(k_x^2k_y^2 + k_x^2k_z^2 + k_y^2k_z^2\right\}^{1/2}\right] \end{aligned} \tag{A.57}$$

Introducing the spherical polar coordinate system with the polar axis along the z-direction, we obtain

$$E_v = \frac{\hbar^2 k^2}{2m_0}\left[A \pm (B^2 + C^2/5)^{1/2}\right].$$

This enables us to define effective masses for heavy and light holes as

$$m_{hh} = \frac{m_0}{A - (B^2 + C^2/5)^{1/2}} \tag{A.58a}$$

$$m_{lh} = \frac{m_0}{A + (B^2 + C^2/5)^{1/2}} \tag{A.58b}$$

Example A.1

Using the values of γ's in Table A.1, the band edge effective masses for Ge are $m_{hh} = 0.33\,m_0$ and $m_{lh} = 0.04\,m_0$.; the values for Si are $m_{hh} = 0.56\,m_0$ and $m_{lh} = 0.16\,m_0$. For GaAs, the values are $m_{hh} = 0.059\,m_0$ and $m_{lh} = 0.08\,m_0$.

Table A.1 Band structure parameters for Ge, Si, and GaAs

	m_e/m_0	γ_1	γ_2	γ_3
Ge	1.58/0.082	13.25	4.20	5.56
Si	0.916/0.191	4.26	0.34	1.45
GaAs	0.067	6.8	2.1	2.9

A.3.4 Momentum Matrix Elements

It follows from Eq. (A.41) that the conduction band effective mass is expressed in terms of P and is related to p_{cv}, the momentum matrix element. The momentum matrix element also appears in the calculation of the optical absorption coefficient or the recombination rate in semiconductors. The conduction band edge state for a direct-gap semiconductor has been found to have s-type symmetry and is denoted by $|S\alpha\rangle$ and $|-S\beta\rangle$. The valence band states are written in terms of angular momentum spin representation in Eq. (A.32).

From symmetry, we find that only the matrix elements of the form

$$\langle X|p_x|S\rangle = \langle Y|p_y|S\rangle = \langle Z|p_z|S\rangle$$

are nonzero. The nonvanishing matrix elements are

$$\begin{aligned} &\langle \pm 3/2|p_x|\pm S\rangle = (1/\sqrt{2})\langle X|p_x|S\rangle, \quad \langle \pm 1/2|p_x|\mp S\rangle = (1/\sqrt{6})\langle X|p_x|S\rangle \\ &\langle \pm 3/2|p_x|\pm S\rangle = (2/\sqrt{6})\langle X|p_x|S\rangle \end{aligned}.$$

One may define a quantity:

$$E_p = \frac{2}{m_0}|\langle X|p_x|S\rangle|^2 = \frac{2}{m_0}p_{cv}^2$$

The values of E_p for different semiconductors are remarkably close to ~25 eV.

A.4 Quantum Wells

The subband structures for electrons and holes have been calculated in Chapter 3 by using simple theory. However, for refined calculation, complete knowledge of the E–$\boldsymbol{k}$ dispersion relation is needed in order to explain the experimental results. Here we shall give the outline of the theory for valence band states in a quantum well.

A.4.1 Subband Structures for Holes

The degenerate nature of the valence bands prompts us to employ the multiband effective mass approximation. The Hamiltonian is written as

$$\sum_{\nu'} [H_{\nu\nu'}(\mathbf{k}) + V(z)\delta_{\nu\nu'}]\phi_m^{\nu'} = E_m\phi_m^{\nu}. \tag{A.59}$$

The complete wavefunction for the valence band hole in the mth subband in the vth valence band $\varphi_{mk}(\mathbf{r})$ is written in terms of the envelope function ϕ_m^v as

$$\varphi_{mk}(r) = \sum_{v=1}^{4} \phi_m^v(\mathbf{k}, z) \exp(j\mathbf{k}.\mathbf{r})\, U^v(\mathbf{r}). \tag{A.60}$$

We have assumed as before that the conduction band is decoupled from the valence bands and only the two top valence bands are considered. We therefore consider four eigenstates: $|1\rangle = |3/2, 3/2\rangle, |2\rangle = |3/2, -1/2\rangle, |3\rangle = |3/2, 1/2\rangle, |4\rangle = |3/2, -3/2\rangle$.

The Luttinger Hamiltonians for the hole states are given by Eq. (A.53) and the coefficients are given in Eq. (A.54).

For a rectangular QW, $H_{vv'}(\mathbf{k})$ given in the matrix form described here should be replaced by $H_{vv'}(\mathbf{k}, -j\partial/\partial z)$, where $\mathbf{k}$ is now the in-plane wave vector. The simple solution for $\mathbf{k} = 0$ has been worked out in Chapter 3 (see Eq. (3.57)).

The in-plane effective masses for holes in different subbands have been calculated by different authors using various degrees of approximations. We assume that the band gap as well as the separation between the heavy-hole and split-off bands are large, so that the Hamiltonian matrix is treated as a 4×4 matrix as before. The eigenvalues are obtained by solving the secular determinant of the 4×4 matrix, and the expression is given by (A.57).

The upper and lower signs correspond, respectively, to heavy- and light-hole bands. The character of the bands becomes increasingly mixed for higher values of $\mathbf{k}$. The effective mass in the xy plane depends, in general, on the direction, and the magnitude of the anisotropy is determined by the difference in γ_2 and γ_3. In many cases the difference is small and it is justified to take a spherical average. It can be shown that the average $\left\langle k_x^2 k_y^2 + k_y^2 k_z^2 + k_z^2 k_x^2 \right\rangle \cong k^4/5$. Putting this in Eq. (A.57), we obtain

$$E = \frac{\hbar^2}{2m_0}(-\gamma_1 \pm 2\bar{\gamma})k^2 \tag{A.61}$$

where

$$\bar{\gamma}^2 = (2\gamma_2^2 + 3\gamma_3^2)/5. \tag{A.62}$$

Each of the four eigenfunctions is of the form:

$$\varphi = [A|3/2, 3/2\rangle + B|3/2, -1/2\rangle + C|3/2, 1/2\rangle + D|3/2, -3/2\rangle] \exp(j\,\mathbf{k}.\mathbf{r}) \tag{A.63}$$

The four eigenfunctions may be expressed as column matrices.

In the practical situation when the barrier height is finite, a parameter $\bar{\gamma} = (\gamma_2 + \gamma_3)/2$ is introduced. The 4×4 Hamiltonian matrix is transformed to a new matrix $\bar{H}$ by using a unitary matrix U such that [12]

$$\bar{H} = UHU^+ = \begin{bmatrix} H^U & 0 \\ 0 & H^L \end{bmatrix}, \tag{A.64}$$

where

$$H^U = \begin{bmatrix} H_{hh} & R \\ R^* & H_{hh} \end{bmatrix} \quad \text{and} \quad H^L = \begin{bmatrix} H_{lh} & R \\ R^* & H_{lh} \end{bmatrix} \tag{A.65}$$

$$R = |c| - j|b|$$

The upper and lower blocks are then decoupled. We write now the upper- and lower-block envelope functions as

$$\varphi_{mk}^{U}(\mathbf{r}) = \sum_{v=1,2} g_m^{(v)}(\mathbf{k}, z) \exp(j\mathbf{k}.\mathbf{r})|v\rangle \tag{A.66a}$$

and

$$\varphi_{mk}^{L}(\mathbf{r}) = \sum_{v=2,4} g_m^{(v)}(\mathbf{k}, z) \exp(j\mathbf{k}.\mathbf{r})|v\rangle, \tag{A.66b}$$

where $\{|v\rangle\}$ denotes the transformed basis set, and the envelope functions satisfy

$$\sum_{v'=1,2} \left[H_{vv'}^{L}\left(\mathbf{k}.-j\frac{\partial}{\partial z}\right) + V(z)\delta_{vv'}\right] g_m^{(v'+2)}(\mathbf{k}, z) = E_m^L(\mathbf{k}) g_m^{(v+2)}(\mathbf{k}, z). \tag{A.67}$$

A.4.2 Subband Structures for Strained Ge on GeSiSn

The theoretical calculation by Chang and Chuang [13] is somewhat along the lines shown, but includes the strain effect. The upper- and lower-block Hamiltonians are expressed as

$$H_U = \begin{bmatrix} V_v(z) - P - Q & \bar{R} \\ \dot{R}^+ & V_v(z) - P + Q + \Delta Q \end{bmatrix} \tag{A.68a}$$

$$H_L = \begin{bmatrix} V_v(z) - P + Q + \Delta Q & \bar{R} \\ \bar{R}^+ & V_v(z) - P - Q \end{bmatrix} \tag{A.68b}$$

$$P = -\frac{\hbar^2}{2m_0}\frac{\partial}{\partial z}\gamma_1\frac{\partial}{\partial z} + \frac{\gamma_1\hbar^2 k_t^2}{2m_0} + P_\varepsilon$$

$$P_\varepsilon = -a_v(\varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz})$$

$$Q = \frac{\hbar^2}{2m_0}\frac{\partial}{\partial z}\gamma_2\frac{\partial}{\partial z} + \frac{\gamma_2\hbar^2 k_t^2}{2m_0} + Q_\varepsilon$$

$$Q_\varepsilon = -\frac{b_v}{2}(\varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz})$$

$$\Delta Q = \frac{1}{2}\left[\sqrt{(\Delta + Q_\varepsilon)^2 + 8Q_\varepsilon^2} - (\Delta + Q_\varepsilon)\right]$$

$$\bar{R} = \frac{\sqrt{3}\hbar^2}{2m_0}\left[\left(\frac{\gamma_2 + \gamma_3}{2}\right)k_t^2 - k_t\left(\frac{\partial}{\partial z}\gamma_3 + \gamma_3\frac{\partial}{\partial z}\right)\right]$$

$$\bar{R}^+ = \frac{\sqrt{3}\hbar^2}{2m_0}\left[\left(\frac{\gamma_2 + \gamma_3}{2}\right)k_t^2 + k_t\left(\frac{\partial}{\partial z}\gamma_3 + \gamma_3\frac{\partial}{\partial z}\right)\right].$$

The symbols have their usual meanings.

In a QW grown along the [100] plane, all four L valleys are equivalent. The Hamiltonian for the [111] L valley may be written as

$$H_L^{[111]}\left(k_t, k_z = -j\frac{\partial}{\partial z}\right) - \frac{\hbar^2}{2}\frac{\partial}{\partial z}\left(\frac{1}{3m_l}+\frac{2}{3m_t}\right)\frac{\partial}{\partial z} - j\frac{\sqrt{2}\hbar^2 k_1}{6} \times \left[\frac{\partial}{\partial z}\left(\frac{1}{m_l}-\frac{1}{m_t}\right)+\left(\frac{1}{m_l}-\frac{1}{m_t}\right)\frac{\partial}{\partial z}+\left(\frac{2}{3m_l}+\frac{1}{3m_t}\right)\frac{\hbar^2 k_1^2}{2} + \frac{\hbar^2 k_2^2}{2m_t} + V^{[111]}(z) + V_\varepsilon^{[111]}(z)\right] \tag{A.69}$$

$$V_\varepsilon^{[111]}(z) = a_L(\varepsilon_{xx}+\varepsilon_{yy}+\varepsilon_{zz})$$
$$k_1 = \frac{1}{\sqrt{2}}\left(k_x + k_y - \frac{2\pi}{a}\right)$$
$$k_2 = \frac{1}{\sqrt{2}}\left(-k_x + k_y\right)$$

A.4.3 Expression for Gain

In order to calculate the gain spectra, one needs to evaluate the squared momentum matrix elements for transverse electric (TE) and transverse magnetic (TM) configurations. We quote here the results for the TE momentum matrix element as given in Chang and Chuang.

$$M_{nm}^{U,TE}(k_t) = \frac{3}{2}\left|\int_{-\infty}^{\infty} dz\phi^*{}_{n,k_t=0}(z) M_b g_m^{(1)}(k_t,z)\right|^2 + \frac{1}{2}\left|\int_{-\infty}^{\infty} dz\phi^*{}_{n,k_t=0}(z) M_b g_m^{(2)}(k_t,z)\right|^2$$
$$M_{nm}^{L,TE}(k_t) = \frac{3}{2}\left|\int_{-\infty}^{\infty} dz\phi^*{}_{n,k_t=0}(z) M_b g_m^{(4)}(k_t,z)\right|^2 + \frac{1}{2}\left|\int_{-\infty}^{\infty} dz\phi^*{}_{n,k_t=0}(z) M_b g_m^{(3)}(k_t,z)\right|^2$$
$$M_b^2 = \frac{|\langle S|p_x|X\rangle|^2}{3} = \frac{m_0}{6}E_p. \tag{A.70}$$

where M_b^2 is the bulk momentum matrix element squared and E_p is the corresponding energy parameter. $g_m^{(1)}(k_t,z)$ and $g_m^{(2)}(k_t,z)$ are the eigencomponents of the upper Hamiltonian H_U; $g_m^{(3)}(k_t,z)$ and $g_m^{(4)}(k_t,z)$ are the eigencomponents of the lower Hamiltonian H_L, and ϕ is the eigenfunction for the Γvalley. Similar equations may be obtained for TM waves [13].

References

1. Madelung, O. (1978) *Introduction to Solid State Theory*, Springer, Berlin.
2. Ridley, B.K. (1982) *Quantum Processes in Semiconductors*, Oxford Univ. Press, Oxford.

3. Basu, P.K. (2003) *Theory of Optical Processes in Semiconductors: Bulk and Microstructures*, Clarendon Press, Oxford.
4. Singh, J. (2003) *Electronic and Optoelectronic Properties of Semiconductor Structures*, Cambridge Univ. Press, Cambridge, New York.
5. Nag, B.R. (1980) *Electron Transport in Compound Semiconductors*, Springer, Berlin.
6. Wang, S. (1989) *Fundamentals of Semiconductor theory and Device Physics*, Prentice Hall, Englewood Cliffs, NJ.
7. Anselm, A.I. (1981) *Introduction to Semiconductor Theory*(translation: M.M. Samokkvalov), Mir Publishers, Moscow.
8. Kane, E.O. (1957) Band structure of indium antimonide. *J. Phys. Chem. Solids*, **1**, 249–261.
9. Chuang, S.L. (1995) *Physics of Optoelectronic Devices*, John Wiley & Sons, New York.
10. Luttinger, J.M. and Kohn, W. (1955) Motion of electrons and holes in perturbed periodic fields. *Phys. Rev.*, **97**, 869–883.
11. Landau, L.D. and Lifshitz, E.M. (1977) *Quantum Mechanics: Non-relativistic Theory*, Pergamon, Oxford.
12. Broido, D.A. and Sham, L.J. (1985) Effective masses of holes at GaAs–AlGaAs heterojunctions. *Phys. Rev. B*, **31**, 888–892.
13. Chang, S-W. and Chuang, S.L. (2007) Theory of optical gain of Ge–$Si_xGe_ySn_{1-x-y}$ quantum well lasers. *IEEE J. Quantum Electron.*, **43**, 249–256.