DEPARTMENT OF

ELECTRONICS AND COMMUNICATION ENGINEERING

EC402 NANOELECTRONICS

COURSE MATERIAL

JAWAHARLAL COLLEGE OF ENGINEERING & TECHNOLOGY

JAWAHAR GARDENS,LAKKIDI,,MANGALAM ,PALAKKAD-679301

MODULE I
INTRODUCTION TO NANOTECHNOLOGY

1.1 DEFINITION

The design, characterization, production, and application of structures, devices, and systems by controlled manipulation of size and shape at the nanometer scale (atomic, molecular, and macromolecular scale) that produces structures, devices, and systems with at least one novel/superior characteristic or property.

1.2 THE SIGNIFICANCE OF THE NANOSCALE

A nanometer (nm) is one thousand millionth of a meter. For comparison, a red blood cell is approximately 7,000 nm wide and a water molecule is almost 0.3nm across.

People are interested in the nanoscale – which we define to be from 100nm down to the size of atoms (approximately 0.2nm) – because it is at this scale that the properties of materials can be very different from those at a larger scale. We define nanoscience as the study of phenomena and manipulation of materials at atomic, molecular and macromolecular scales, where properties differ significantly from those at a larger scale; and nanotechnologies as the design, characterisation, production and application of structures, devices and systems by controlling shape and size at the nanometer scale.

The properties of materials can be different at the nanoscale for two main reasons:

First, nanomaterials have a relatively larger surface area when compared to the same mass of material produced in a larger form. This can make materials more chemically reactive (in some cases materials that are inert in their larger form are reactive when produced in their nanoscale form), and affect their strength or electrical properties.

Second, quantum effects can begin to dominate the behavior of matter at the nanoscale - particularly at the lower end - affecting the optical, electrical and magnetic behavior of materials. Materials can be produced that are nanoscale in one dimension (for example, very thin surface coatings), in two dimensions (for example, nanowires and nanotubes) or in all three dimensions (for example, nanoparticles).

1.3 IMPACTS OF NANO TECHNOLOGY

The **impact of nanotechnology** extends from its medical, ethical, mental, legal and environmental applications, to fields such as engineering, biology, chemistry, computing, materials science, and communications.

Below are 10 ways nanotechnology impacts our lives on a daily basis.

1. Faster, smaller, and more powerful computers that consume far less power, with longer-lasting batteries. Circuits made from carbon nanotubes could be vital in maintaining the growth of computer power, allowing Moore's Law to continue.

2. Faster, more functional, and more accurate medical diagnostic equipment. Lab-on-a-chip technology enables point-of-care testing in real time, which speeds up delivery of medical care. Nanomaterial surfaces on implants improve wear and resist infection.

3. Nanoparticles in pharmaceutical products improve their absorption within the body and make them easier to deliver, often through combination medical devices. Nanoparticles can also be used to deliver chemotherapy drugs to specific cells, such as cancer cells.

4. Improved vehicle fuel efficiency and corrosion resistance by building vehicle parts from nanocomposite materials that are lighter, stronger, and more chemically resistant than metal. Nanofilters remove nearly all airborne particles from the air before it reaches the combustion chamber, further improving gas mileage.

5. Nanoparticles or nanofibers in fabrics can enhance stain resistance, water resistance, and flame resistance, without a significant increase in weight, thickness, or stiffness of the fabric. For example, “nano-whiskers” on pants make them resistant to water and stains.

6. Water filters that are only 15-20 nanometers wide can remove nano-sized particles, including virtually all viruses and bacteria. These cost-efficient, portable water treatment systems are ideal for improving the quality of drinking water in emerging countries.

7. Carbon nanotubes have a variety of commercial uses, including making sports equipment stronger and lighter weight. For example, a tennis racket made with carbon nanotubes bends less during impact, and increases the force and accuracy of the delivery. Nanoparticle-treated tennis balls can keep bouncing twice as long as standard tennis balls.

8. Most sunscreens today are made from nanoparticles that effectively absorb light, including the more dangerous ultraviolet range. They also spread more easily over the skin. These same nanoparticles are also used in food packaging to reduce UV exposure and prolong shelf life.

9. Many drink bottles are made from plastics containing nanoclays, which increase resistance to permeation by oxygen, carbon dioxide, and moisture. This helps retain carbonation and pressure and increases shelf life by several months.

10. A huge variety of chemical sensors can be programmed to detect a particular chemical at amazingly low levels, for example, a single molecule out of billions. This capability is ideal for surveillance and security systems at labs, industrial sites, and airports. On the medical front, nanosensors can also be used to accurately identify particular cells or substances in the body.

1.4 LIMITATIONS OF CONVENTIONAL MICROELECTRONICS

MICROELECTRONICS

Microelectronics is that area of electronics technology associated with electronics systems built from extremely small electronic parts or elements. Most of today's computers, weapons systems, navigation systems, communications systems, and radar systems make extensive use of microelectronics technology.

The research and development industry has presently been relying on the silicon based technology which was working at micro scale. The miniaturization of this technology is governed by the famous Moore's law. But with the advent in time, silicon based technology has reached its maximum limits of the reduction in size.

- The main obstacles are in the way:
- The limits of the lithography techniques,
- The rising cost of fabrication and the size of the transistor
- Contamination occurs during high temperature treatment of silicon
- The use of the silicon in device manufacturing is going through the fundamental limitations including the tunneling current and the sub threshold current which lead to the high power consumption and circuit failure.

Ultimately the goal of the scaling down process is to build an individual transistor which is smaller, faster, cheaper and consumes less power. Unfortunately, the scaling down of the lithographically patterned transistors could not continue forever, but the nanoelectronics may be able to continue this scaling down process.

1.5 TRENDS IN NANOELECTRONICS AND OPTOELECTRONICS

The evolution of microelectronic devices is influenced by factors such as growing demand in memory capacity of integrated circuits, high transmission data speed, optical communications, etc. This requires electronic devices with faster speed operation and smaller size as is the case of silicon integrated circuits, in particular dynamic random access memori (DRAMs). Figure shows the reduction of feature size of metal-oxide-semiconduct (MOS) transistors, as well as the number of bits per chip for the period 1970-2000 [1 For example, a 256 M-bit DRAM contains about 10^9 transistors with a feature size L close to 100 nm. For structures with these dimensions,

transport can still be treated classically but we are already at the transition regime to quantum transport.

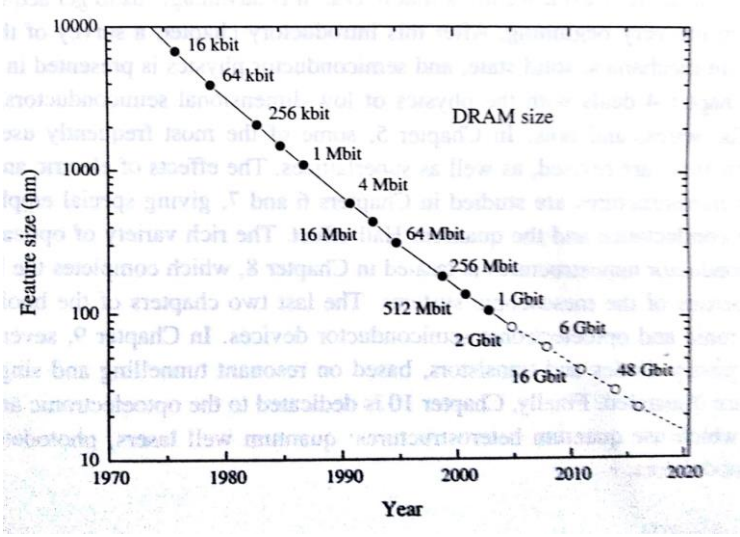


Fig 1.1 : Evolution of the minimum feature of Si DRAM

Today it believed that present silicon technology will evolve towards feature sizes still one order of magnitude lower, i.e. L~10 nm; but below this size, transistors based on new concepts like single electron transistors, resonant tunnelling devices, etc. will have be developed. The operation of this new kind of devices has to be described by the concepts of mesoscopic and quantum physics. It is interesting to remark that quantum effects show up in III-V devices for larger feature sizes, as a consequence of the smaller value the effective mass, and therefore larger value of the de Broglie wavelength In the near future, and due to the growing demands of calculus from industries like communications, information, military, space, etc. microelectronics will be replaced nanoelectronics since the feature size of electronic devices will be reduced to about 10nm. Although there has been an impressive advance at the device and technology level during the last decade, the progress in the development of new system architectures dealing with billions of transistors is scarce. In this sense, new architectures for parallel data processing are under current development.

The evolution towards nanoelectronics has been possible because of advances in the deposition of very thin films to form heterostructures in which electrons could be confined to a 2D mesoscopic system. Standard evaporation and sputtering techniques did not produce heterostructures of enough quality. However, during the 19803 both Molecular Beam Epitaxy (MBE) and Metal Organic Chemical Vapour Deposition (MOCVD) became available .In MBE, a film of the desired material is epitaxially grown over a substrate under ultrahigh vacuum conditions (pressure less than 10^{-10} mbar). The materials are evaporated in tubular effusion cells of the Knudsen type at fairly high temperatures. The atoms or molecules emerge from the effusion cells forming a molecular beam that travels in straight lines towards the substrate where the molecules condense.

Very thin films of elemental semiconductors (Si, Ge), III-V (AlGa, AlGaAs, InP), and II-VI compounds (CdTe, PbS) can be deposited by MBE. This technique allows a layer-by-layer growth of films and superlattices as well as the doping of the material from sources such as B, Al, As, etc. Although the deposited films are of very high quality, the technique is somewhat slow and expensive.

MOCVD is the preferred industrial technique to produce semiconductor heterostructures. For instance, GaAs can be grown over a substrate at about 500 °C by the reaction of arsine (AsH_3) gas and trimethyl gallium $(CH_3)_3Ga$ at sub-atmospheric pressures. Diluted dopant gases are also simultaneously introduced; for instance Si n-type doping of GaAs can be obtained by means of SiH4 diluted in hydrogen which is used as carrier gas. This technique allows for simultaneous deposition on several wafers and is used for the commercial production of semiconductor lasers.

In general, mesoscopic systems require the formation of nanostructures in the range close to 100nm, that is, a decrease in specifications of about one order of magnitude in comparison to the state-of-the-art some 20 years ago. Therefore at present we are close to the limits of conventional optical lithography, and other high resolution nanolithographic techniques (electron beam, ion beam, x-ray, etc.) have to be industrially implemented. As for resists, the most commonly used in nanolithography is the positive tone resist known as PMMA (poly-methylmethacrylate). Although the molecular weight of PMMA is close to 10^6, its roughness, once spin coated onto the substrate, is only about 2 nm.

A frequent matter of present discussion is the ultimate limits in device size. taking into account the evolution shown in Figure 1.1. It seems reasonable that the rate of scaling down predicted by Moore's law will have to slow down. It is expected that the limits of further miniaturization, from an industrial and economic point of view, will be reached in about one decade. The technological limits are related to several factors of which we will mention only two. The first has to do with the amount of heat generated by the

power consumed, and which cannot be eliminated because of thermal conductivity limits of the materials and the increasing number of overlayers. The other factor is related to the so-called "parameter spread" in fabrication. For instance many of the electrical parameters of the MOS transistors are set by doping; however, as the size of the region to be doped is decreased to about 0.1 μm^3, the number of doping atoms becomes so low (about 10) that the parameter spread cannot be controlled appropriately.

In addition to the above mentioned technological limits, there are others of fundamental nature, which are called *physical limits*. Although at present integrated devices are still far from these limits, we think it is important to revise them. They are the following:

- ***Thermal limit*:** The energy necessary to write a bit should be several times kT which is the average energy of thermal fluctuations. In CMOS the lowest values contemplated to write a bit should not be smaller than about 2eV, i.e. 100kT at room temperature or ~3 x 10^{-19} J.

- ***Relativistic limit:*** Signals cannot propagate faster than the speed of light.Therefore assuming that the nucleus of a microprocessor has a size of a few cm, it takes 10^{-10} s for the signal to propagate, which corresponds to a frequency of about 10 GHz.
- ***Uncertainty principle***: According to the Heisenberg's uncertainty principle, the energy and time needed to write or read a bit should be related by $\Delta E \times \Delta t \geq h$. To be safe, we ask the product $\Delta E \times \Delta t = 100h$. Since for future circuits ΔE could be as low as 10^{-19} J, we can appreciate that we can approach the quantum limit as the frequency increases.

At present nanoelectronics is moving simultaneously along several directions. One of them is solid state nanoelectronics, usually consists of heterostructures of well-known materials (Si, Si02, III-V compounds), and several types of transistors: heterojunction, single-electron, resonant-tunnelling, ballistic, etc. However, the amount of computational capacity for some tasks like speech and visual recognition is so large that other radically different alternatives are being sought. Some of these alternatives, like superconductivity electronics and spintronics, use fabrication techniques not too different from those employed in present integrated circuit technologies.

Superconducting electronics, proposed in the 1970s, and developed to the prototype stage during the 1980s, is based on the switching properties of the Josephson junction, which consists of two superconducting layers separated by a very thin oxide insulating film that can be tunnelled by superconducting pairs of electrons. The advantages of superconducting electronics are based on the fact that Josephson junctions can operate at high switching speeds (switching times between 1 and l0 ps), the amount of dissipated power is very low and the resistance of interconnect superconducting lines is practically null.

Another technology being pursued is spintronics which exploits the spin orientation of electrons. Electron-spin transistors are built by enclosing a semiconductor layer (base) sandwiched between two ferromagnetic layers (emitter and collector). Electrons acquiring the magnetization state of the emitter can only travel through the collector across the base if their spins are aligned with the magnetization of the collector. These developments can be considered as running parallel to the efforts in magnetoelectronics to develop MRAM memories, based both in the giant magnetoresistance effect and in the magnetic tunnel junctions developed in 1995. Electron-spin transistors have a big potential if successfully integrated with CMOS circuits.

Other radically new alternatives for future nanoelectronics have been proposed. Molecular electronics is based on the different states or configurations that molecules can take, like "trans" or "cis", as well as parallel or antiparallel alignment of the spin of unpaired electrons. The change between states must be fast, consume little energy, should be addressed by some external signal, and should be readable by a probe. If this technology is able to be put to work, we will have the ultimate step in miniaturization, since molecules are much smaller than present feature sizes in integrated circuits. Besides, molecules have the advantage of being capable of self-organization in 3D supramolecular entities, although, after the development of scanning atomic force microscopy, molecules can be, in principle, individually manipulated. Examples of molecules that can be used in molecular electronics are: azobenzene, hydrazobenzene, etc. One advantage of organic molecules, when compared to inorganic ones, is that it is easier to isolate them in single molecular systems. Although in the future it is expected that molecular wires or nanotubes can be developed for contacting molecules, today only metallic or semiconducting electrodes are

contemplated. Even with this limitation, the interfacing of molecules to the external world for addressability in the case of large systems seems, at present, an enormous problem.

Lastly, we mention biology-inspired electronics, also called bioelectronics. In trying to copy nature, we are not concerned with the size of the building blocks, since for instance, a neuron is very large for nanotechnology standards. What nanoelectronics would like to imitate of biological neurons is the capabilities in parallel processing as well as their 3D architectures and the topology of the interconnects. This results from the large number of computations, needed for instance in pattern recognition of the visual systems of humans and animals, which has to be performed simultaneously at different sites. In addition to parallel processing, neural networks try to simulate the integration of computing and memory functions, which in CMOS microprocessors are performed separately.

To conclude this section, the present situation in optoelectronics will be shortly considered. Optoelectronic devices, based mainly in III-V direct gap semiconductors, have received a great upsurge since the development of optical fibre communications. In addition, there is at present a tendency to replace, whenever possible, electronic devices by photonic ones. The III-V semiconductors more frequently used are based on AlGaAs-GaAs and GaInAsP-InP heterostructures which cover the 0.8-1.6μm wavelength range. GaN blue lasers for short wavelength applications were developed about ten years ago. In the last two decades, quantum well semiconductor lasers with very low threshold currents and photodetectors are replacing the conventional ones, especially in long distance optical communications. One very interesting type of quantum well lasers, which operate at still lower threshold currents, is based on strained-layer quantum heterostructures.

At present, laser diodes are manufactured in chips by standard integrated circuit technology, coupled to transistors and optical interconnects, constituting the so-called optoelectronic integrated circuits (OEIC). In all cases, the trend in optoelectronic devices is to achieve a high level of integration which implies smaller sizes, but still in the micron range. Here again there is a lot of research in efficient integration architectures.

Perhaps, the greatest advances in optoelectronics based on quantum semiconductor heterostructures can be found in the field of electro-optical signal modulation. In effect, modulators based on the confined quantum Stark effect are several orders of magnitude more effective than their bulk counterparts. This is due to the fact that excitons in quantum wells have much higher ionization energy than in the bulk, and therefore can sustain much higher electric fields.

1.6 CHARACTERISTIC LENGTHS IN MESOSCOPIC SYSTEMS

Mesoscopic physics deals with structures which have a size between the macroscopic world and the microscopic or atomic one. These structures are also called mesoscopic systems, or nanostructures in a more colloquial way since their size usually ranges from a few nanometres to about 100 nm. The electrons in such mesoscopic systems show their wavelike properties and therefore their behaviour is markedly dependent on the geometry of the samples. In this case, the states of the electrons are wavelike and somewhat similar to electromagnetic radiation in waveguides. For the description of the behaviour of electrons in solids it is very convenient to define a series of characteristic lengths. If the dimensions of the solid in which the electron is embedded is of the order of, or smaller than these characteristic lengths, the material might show new properties, which in general are more interesting than the corresponding ones in

macroscopic materials. In fact, the physics needed to explain these new properties is based on quantum mechanics. On the contrary, a mesoscopic system approaches its macroscopic limit if its size is several times its characteristic length. Let us next describe some of the most commonly used characteristics lengths in mesoscopic systems.

i. ***De Broglie Wavelength***: It is well known from quantum mechanics that for an electron of momentum p, there corresponds a wave of wavelength given by the de Broglie wavelength:

$$\lambda_B = \frac{h}{p} = \frac{h}{m * v}$$

In Eq. we have substituted p by m∗ v in a semiclassical description, where m* is the electron effective mass. From solid state physics, we know that inside a semiconductor, electrons behave dynamically as if their mass was m*, instead of the mass m_0 of the electron in vacuum. This observation is very important because for many interesting semiconductors, like GaAs or InSb, m* is much smaller than m_o, For instance, for GaAs and InSb, m* is equal to 0.067mo and 0.014mo, respectively

ii. Mean free path

As the electron moves inside a solid, it is usually scattered by interactions with crystal imperfections like impurities, defects, lattice vibrations (phonons), etc. In most cases, these scattering events or "collisions" are inelastic, i.e. the values of energy and momentum of the system after the interaction, differ from the corresponding ones before they interact. The distance covered by the electron between two inelastic collisions is usually called the mean free path ***l_e*** of the electron in the solid. If *v* is

the speed of the electron, then

$$l_e = v\tau_e$$

where τ_e is known as the relaxation time.

iii. Diffusion length

In a mesoscopic system of typical size L, the electrons can move either in the ballistic regime or in the diffusive regime. If the previously defined mean free path to is much larger than L, the particle moves throughout the structure without scattering; this is the so-called ballistic transport regime in which the surfaces usually are the main scattering entities. In hot electron transistors ,electron transport is ballistic and the electrons can reach energies much higher than the ones corresponding to the lattice thermal energy. 0n the other hand, if (l_e « L, transport can be explained as a diffusion process. In this case, the system is characterized by a diffusion coefficient D. In terms of D, the diffusion length L. is defined by

$$L_e = (D\tau_e)^{1/2}$$

where τ_e is the relaxation time. In semiconductor theory, the concept of diffusion length is used very often; for instance, if electrons diffuse within a p-type

semiconductor, their concentration diminishes exponentially with distance with a decay length equal to Le.

In the diffusive regime, transport in the mesoscopic systems is usually explained by means of the Boltzmann equation, as in the bulk. On the contrary, in the ballistic regime, the Boltzmann transport model is not valid, and electrons move through the structure essentially without scattering.

iv. Screening length

In extrinsic semiconductors, the dopants or impurities are usually ionized and constitute a main factor contributing to scattering. However, in general we cannot consider that the electrical potential produced by these impurities varies with distance as 1/r . Because of the screening of free carriers by charges of the opposite polarity, the effect of the impurity over the distance is partially reduced. It is found that the variation of the potential is modulated by the term exp(-r/λs,) where λs; is called the screening length and is given by

$\lambda s = \left(\frac{\in kT}{e^2 n}\right)^{1/2}$

where e is the electronic charge, ∈ the dielectric constant of the semiconductor, and n the mean background carrier concentration. One should be careful about nomenclature because some authors call λs; the Debye length or the Fermi-Thomas length. In a typical semiconductor, As is in the range 10-100nm, and is an indication of the attenuation of charge disturbances in a semiconductor. From Eq. it is determined that λs; should be much smaller in metals than for semiconductors.

v. Localization length

The localization length can be understood in terms of transport in disordered materials, in which we know from solid state physics that, in addition to Bloch extended states, there can also be localized states. In disordered materials, the electrons move by "hopping" transport between localized states or from a localized state to a bound state.

In order to describe the hopping transport and other mesoscopic properties of the localized states, it is assumed that the electron wave function is described by

$$\boldsymbol{\psi = e^{\left(\frac{-r}{\lambda_{loc}}\right)}}$$

where λ_{loc} is known as the localization length. Evidently, the electrical conductivity of a material will be proportional, among other factors, to the overlap between the wave functions. If the sample dimensions are of the order of λ_{loc}. we can say that our system is mesoscopic.

1.7 QUANTUM MECHANICAL COHERENCE

In a mesoscopic structure of dimensions similar to the electron dc Broglie wavelength λ_B. the behaviour of the electron should be described quantum mechanically, i.e. by using Schrodinger equation. 1f the electron interacts inelastically with a defect, or any impurity, the electrons change their energy and momentum, as well as the phase of their wave function. ***The phase coherence length L_Φ, is defined as the distance travelled by the electron without the carrier wave changing its phase***. Evidently, interference effects in the electron waves should only be observed if the particles move over distances of the order of, or smaller than, $\boldsymbol{L_\Phi}$. Since we are usually interested in mesoscopic systems in the quasi-ballistic regime, in which electrons are practically unscattered, $\boldsymbol{L_\Phi}$, should be a length similar to the inelastic scattering mean free path *le* previously defined. Coherent states can evidently show interference effects. On the contrary, once the coherent states loose their coherence, by inelastic scattering, the corresponding waves cannot be superposed and cannot interfere (in some way, the matter-wave becomes a particle). In mesoscopic physics, the loss of coherence is usually called dephasing. Evidently, the coherence processes are characteristic of mesoscopic systems.

From the above definition of $\boldsymbol{L_\Phi}$, electrons can show interference effects over distances smaller than $\boldsymbol{L_\Phi}$.If electrons with phase Φ1 interfere with electrons of phase $\boldsymbol{\Phi 2}$, we know from the wave theory that the amplitude of the resultant wave varies as cos(*Φ1-Φ2*) and the amplitudes can add up to each other, or they can be subtracted depending on the phase difference.

1.8 SCHRODINGER WAVE EQUATION

As it was established by Schrodinger in I926. The dual wave-like and particle-like properties of matter are described by the so-called wave function $\boldsymbol{\Phi}(\vec{\boldsymbol{r}}, \boldsymbol{t})$. Which is continuous and has continuous derivatives. The wave function is also complex. I.e. It has a real part end an imaginary one. The wave function satisties the second-order. Linear, differential Schrodinger equation,

$$\left[-\frac{h^2}{2m}\nabla^2 + \boldsymbol{V}(\vec{\boldsymbol{r}}, \boldsymbol{t})\right]\Phi = ih\frac{d\Phi}{dt} \quad \text{.....(1)}$$

Where ∇^2 is the operator $\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}$ and V is the potential energy. Which is generally a function of position and possibly of time. Although the function Φ does not have a physical meaning, the product of Φ by its complex conjugate is a real quantity, such that the probability dP of finding a particle in a small volume d V is given by

$$dP = |\Phi^2|dV \text{.....(2)}$$

If the potential energy V is not time dependent, we can search for a solution to Eq (1)

$$\Phi(\vec{r}, t) = \psi(\vec{r})e^{-j\omega t} \text{.....(3)}$$

Substituting Eq. (3) in (1) and writing E = hω, we find the time-independent Schroidinger equation

$$\left[-\frac{h^2}{2m}\nabla^2 + \boldsymbol{V}(\vec{\boldsymbol{r}})\right]\psi(\vec{r}) = E\,\psi(\vec{r}) \text{.....(4)}$$

for the time-independent wave function $\psi(\vec{r})$. Schroidinger's equation can be solved exactly only in a few cases. Probably the simplest one is that of the free particle, as for instance a free electron of energy E and mass m. In this case $\boldsymbol{V}(\vec{\boldsymbol{r}})$ = 0 and the solution of Eq. (1) is easily found to be

$$\Phi = Ae^{i(kx-\omega t)} + Be^{i(-kx-\omega t)} \text{.....(5)}$$

Where

$$k = \left(\frac{2mE}{h^2}\right)^{1/2} \text{.....(6)}$$

Therefore, the free electron is described by a wave. which according to the de Broglie relation has momentum and energy given. respectively, by

$$p = hk, \qquad E = \frac{p^2}{2m} \text{.....(7)}$$

in general we will assume that the electron travels in one direction, for instance, along the x-axis from left to right. and therefore the coefficient B in Eq. (5) is zero. The wave equation for the free electron can simply be written as:

$$\Phi = Ae^{i(kx-\omega t)} \text{.....(8)}$$

Another example in which Schrodinger's equation can be solved exactly is that of the hydrogen atom for which the potential is Coulombic, i.e. V varies with the distance r between the proton and electron in the form l/r. Solving Schrodinger's equation, one gets the well-known expression for the electron energies:

$$En = -\frac{m_r e^4}{2(4\pi\varepsilon_0)^2 h^2 n^2} = -\frac{13.6}{n^2}\ (eV)\ , \qquad n = 1,2,3.. \text{.....(9)}$$

where m, is the reduced proton-electron mass. In solid state physics, the mathematical model of the hydrogen atom is often used, as for example, in the study of the effects of impurities and excitons in semiconductors. Although the equation giving the values of the energy is very similar to Eq. (9), the values of the energy are much smaller, since the dielectric constant of the medium has to substitute the value of the permittivity of vacuum ε_0. For instance, in the case of silicon, the value of the dielectric constant is about $12\varepsilon_0$.

The model of the harmonic oscillator is very often used in solid state physics.. In general terms, we can say that the harmonic oscillator model can be used to describe any system which

performs vibrations of small amplitude about an equilibrium point. The allowed energies of the harmonic oscillator can be obtained from Schrodingers's equation and are given by:

$$En = \left(n - \frac{1}{2}\right) h\omega, \quad n = 1,2,3 \ldots \qquad \ldots\ldots\ldots\ldots(10)$$

Another potential that allows the exact solution of Schrodinger's equation is the so-called infinite square well potential. which consists of a flat potential of width a surrounded by infinite potentials at its extremes. It is a fairly good approximation to modulation-doped quantum wells, which is the basic structure of many quantum well transistors and lasers.

1.9. QUANTUM WELLS, WIRES, AND DOTS

We have defined a series of characteristic lengths λ which correspond to physical properties of electrons which are size dependent. We have also seen that when the dimensions of the solid get reduced to a size comparable with, or smaller than λ, then the particles behave wave-like and quantum mechanics should be used.

Let us suppose that we have an electron confined within a box of dimensions L x, Ly, Lz. If the characteristic length is λ, we can have the following situations:

i. λ<< in Lx,Ly, LZ
In this case the electron behaves as in a regular 3D bulk semiconductor.

ii. λ > Lx, and Lx << Ly, Lz In this situation we have a 2D semiconductor perpendicular to the x-axis. This mesoscopic system is also called a quantum well .

iii. λ > Lx, Ly and Lx, Ly < Lz; corresponds to 1D semiconductor or quantum wire. located along the z-axis

iv. (iv) λ >> Lx,Ly, L_Z ,In this cease it is said that we have a 0D semiconductor or quantum dot

In general, we say in mesoscopic physics that a solid, very often a semiconductor, is of reduced dimensionality if at least one of its dimensions Li, is smaller than the characteristic length. For instance, if Lx and Ly are smaller than λ. we have a semiconductor of dimensionality equal to one. We could also have the case that λ is comparable, or a little larger, than one of the dimensions of the solid but much smaller than the other two. Then we have a quasi 2D system, which in practice is a very thin film, but not thin enough to show quantum size effects.

1.10. DENSITY OF STATES AND DIMENSIONALITY

Although the density of states (DOS) of physical systems will be derived formally in Chapter 4, in this section we see from a mathematical point of view the consequmces of the dimensionality of the system in the DOS. As we know from solid state physics, most physical properties significantly depend on the DOS function ρ. The DOS function, at a given value E of energy, is

detined such that ρ (E)ΔE is equal to the number of states (i.e. solutions of Schrodinger equation) in the interval energy ΔE around E. We also know that if the dimensions

L_i(i = x, y, z) are macroscopic and if proper boundary conditions are chosen, the energy levels can be treated as a quasi-continuous. 0n the other hand, in the case where any of the dimensions Li gets small enough, the DOS function becomes discontinuous. Let us next obtain the DOS function for several low-dimensional solids. First, let us remind that for bulk solids ρ(E) varies with energy in the form $\sqrt{E}$

If each quantum state or Bloch state in a solid is designated by a quantum number k (Bloch state), the general expression for the DOS function should be

$\rho(E) = \sum_k \delta(E - E_k)$ ……1.8.1

where the quantized energies are given by Ek and δ (E) is the Dirac's delta function. If we take into account the electron spin degeneracy, a factor 2 should also appear in the above expression. Let us recall for simplicity the case of a cubic shaped 3D macroscopic crystalline solid, of edge

L = Na, where a is the lattice constant and N the number of sites along the one-dimensional directions which is supposed to be large. In this case, the eigenstates can be considered as quasi-continuous and the summatory in k of Eq. can be replaced by an integral, i.e.

$\sum_k \rightarrow \frac{L}{(2\pi)^3} \int dk$ ……1.8.2

for the case of a cube of size L and volume $V = L^3$ .We also know from simple solid state theory that if we assume that the energy E_k only depends on the magnitude of k in a parabolic energy; dependence between the momentum hk, i.e.

$E_k = \frac{h^2 k^2}{2m^*}$ ……1.8.3.

$\rho_{3D}(E) = \frac{V}{2\pi^2}\left(\frac{2m^*}{h^2}\right)^{3/2}\sqrt{E}$ ……1.8.4

We can follow exactly the same procedure for 2D and 1D semiconductors of area A and length L, respectively, reaching the following expressions

$\rho_{2D}(E) = \frac{A}{\pi}\left(\frac{m^*}{h^2}\right)$ ……1.8.5

$$\rho_{1D}(E) = \frac{L}{2\pi}\left(\frac{2m^*}{h^2}\right)^{1/2}\frac{1}{\sqrt{E}} \qquad \text{.....1.8.6}$$

Some important considerations can already be made: the DOS function in 3D semiconductors is proportional to $\sqrt{E}$, in 2D is constant, and in 1D varies inversely pmpomonal t0 $\sqrt{E}$. This implies in the last case that at the bottom of bands, the DOS plays a very important role. because there is a singularity for E = 0.

Eqns 1.8.5. & 1.8.6. were derived for perfectly 2D and 1D solids. but in the real world a 2D solid. for instance, is really a 3D one where the perpendicular dimension is very short. We will see that since electrons can move almost freely in the (x, y) plane, Eq. 1.8.5. should be written for a quasi-2D solid as

$$\rho_{2D}(E) = \frac{A}{\pi}\left(\frac{m^*}{h^2}\right)\sum_{n_z}\theta(E - E_{n_z})$$

where n_z; refers to the quantization in the contined z-axis and θ is the step function.

Similarly, for a quasi-lD solid or quantum wire along the z-direction

$$\rho_{1D}(E) = \frac{L}{2\pi}\left(\frac{2m^*}{h^2}\right)^{1/2}\frac{1}{\sqrt{E - E_{nx,ny}}}$$

where n_x and n_y, are the quantum numbers for the confined x and y directions. This function looks like a series of peaks, one for each value of $E_{nx,ny}$.

Evidently, in the case of a quasi-OD solid or quantum dot, there is no continuous DOS function, since there is quantization in the three spatial directions. Therefore the DOS function consists of a series of peaks given by

$$\rho_{0D}(E) = \sum_i \delta(E - E_i)$$

Where i=(nx,ny,nz) , the peaks in quantum dots are not perfect δ -functions since there is a broadening effect as a consequence of scattering mechanisms.

1.11 BASIC PROPERTIES OF 2D SEMICONDUCTOR NANO STRUCTORS

One of the most practical two-dimensional semiconductor structures consists of a sandwich of gallium arsenide (GaAs), with a thickness in the nanometre range, surrounded on each side by a semiconductor such as aluminium gallium arsenide ($Al_xGa_{1-x}As$) of higher bandgap. The bandgap of $Al_xGa_{1-x}As$ ($x \sim 0.3$) is close to 2.0 eV while that of GaAs; is 1.4 eV. As a consequence, the potential energy profile has the shape of a square well , with a barrier height of 0.4 eV for electrons and 0.2 eV for holes. In reality, the profile of the potential barrier is somewhat more complicated, since the potential varies with atomic distances, which also affects the wave functions. However, in most cases, it is a good approximation to consider the average over a few atomic distances (envelope function approximation). As seen in Figure (a), carrier motion for both electrons and holes is not allowed in the direction perpendicular to the well, usually taken as the z-direction because of the potential walls. However in the other two spatial directions (x, y), parallel to the semiconductor interfaces, the motion is not restricted. is the electrons behave as free electrons.

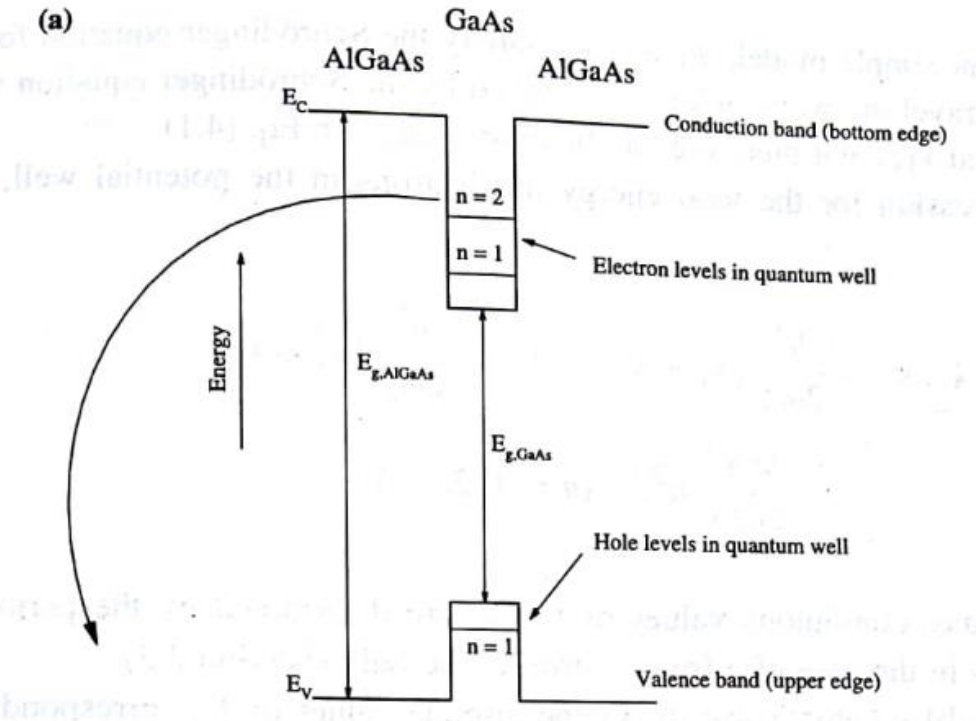


Fig 1.9; (a) AlGaAs-GaAs-AlGaAs Square potential well

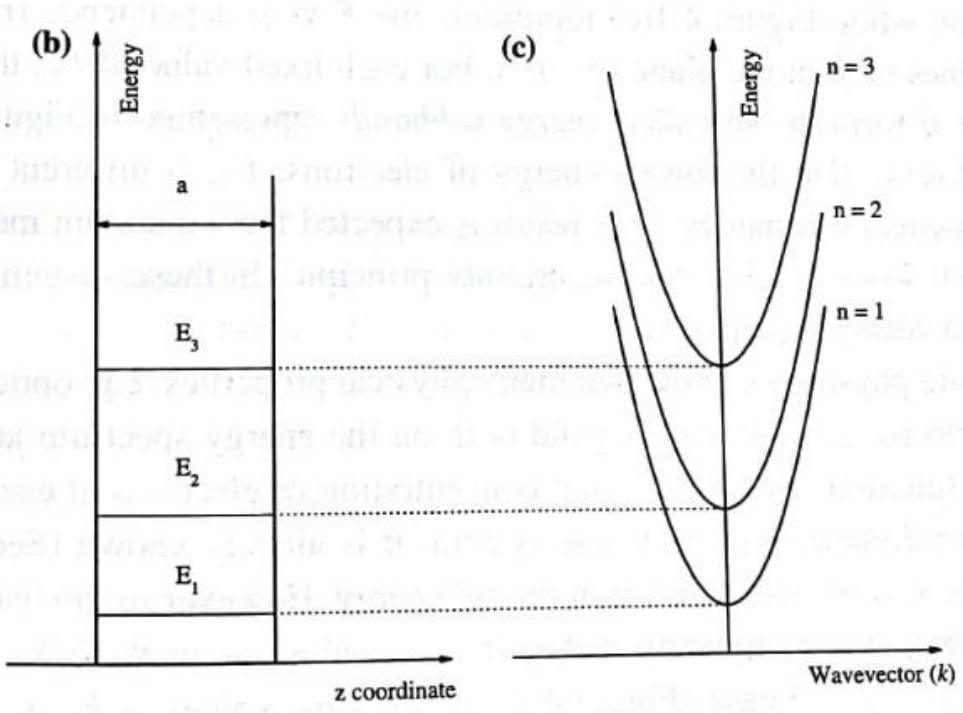


Fig1.9 (b): Energy Levels Fig1.9 : (c) Energy subbands

The behaviour of electrons when their motion is restricted along one direction in the wells of infinite height corresponds to a well-known problem in quantum mechanics, d! so-called particle in a box of infinite wells. It is well known from quantum mechanics that, in the case of infinite potentials barriers, the wave functions and energy levels of the bound electrons are given by

$$\psi_n(z) = \left(\frac{2}{a}\right)^{1/2} sin\left(\frac{\pi n z}{a}\right) \quad \text{.....(1)}$$

$$En = \frac{h^2\pi^2}{2m_e{}^* a^2} n^2 \quad , \quad \text{n=1,2,3,.....} \quad \text{.....(2)}$$

where $m_e{}^*$ is the effective mass of the electrons in the well material for the motion along the

z-direction and a is the width of the well. From Eq. (2) we can derive several important consequences:

(1) In general, quantum size effects will be more easily observable in quantum structures of very small size a, and for materials for which the electron effective mass is as small as possible. In this sense, GaAs nanostructures are very convenient since $m_e{}^*$ ~0.067mo, where mo is the free electron mass. This is equivalent to saying that in materials for which the electron mobility or the free electron path are large, quantum effects are easier to observe.

(2) Quantum size effects, which require energy transitions of electrons between levels, are better observed at low temperatures. since the mean thermal energy of carriers is of the order of kT.

As it has been described, the motion of electrons in the quantum well is confined only in one direction, z:, but in the (x, y) planes the electrons behave as in a three-dimensional solid. Therefore the electron wave function is separable as the product of **ψ**x, **ψ**y, **ψ**z, ie.

ψ= **ψ**x **ψ**y **ψ**z ……..(3)

where, in our simple model, **ψ**x and **ψ**y, satisfy the Schrondinger equation for a free electron, i.e. a travelling wave, while **ψ**z is given by the Schrodinger equation with a Square Well potential V(z) and therefore can be expressed as in Eq. (1).

The expression for the total energy of electrons in the potential well, can then be written as

$$E(kx, ky, n) = \frac{h^2}{2m_e^*}\left({k_x}^2+{k_y}^2\right) + En = \frac{h^2}{2m_e^*}\left({k_x}^2+{k_y}^2\right) + \frac{h^2\pi^2}{2m_e^* a^2}n^2 \quad \text{……..(4)}$$

where the quasi-continuous values of kx, ky, are determined by the periodic boundapy conditions as in the case of a free electron in the bulk.

Figure (b) schematically shows the discrete values of En corresponding to motion in the z-direction while Figure (c) represents the E vs p dependence (remember the p = hk) for values of $\vec{p}$ in the plane (px, py). For each fixed value of En... the values of E as a function of $\vec{p}$ form the so-called energy subbands represented in Figure(c).It is interesting to observe that the lowest energy of electrons, E1, is different from zero, in contrast with classical mechanics. This result is expected from quantum mechanics since a value of E = 0 would violate the uncertainty principle. In these systems the valued E = E1 is called zero-point energy.

From solid state physics we know that many physical properties, e.g. optical absorption. transport of electronic current, etc. depend both on the energy spectrum and the denim of states (DOS) function, which gives the concentration of electrons at each value oi the energy. In a three-dimensional electronic system, it is already known that the DOS depends as a parabolic function on the energy. However in the two-dimensional case, this dependence is completely different. Proceeding as in the 3D case, it can N appreciated that for the 2D case the possible values of kx,

ky, are separated by 2π/L, where L is the dimension of the sample, which has been assumed to be square without loss of generality. The number of states in the k-space within a circular ring limited by the circumferences of radii k and k + dk is therefore:

$$n_{2D}(k)dk = \frac{2\pi k dk}{\left(2\pi/_L\right)^2} \quad \text{.....(5)}$$

And the number of states in k –space per unit area is

$$n_{2D}(k) = \frac{k}{2\pi} \quad \text{.....(6)}$$

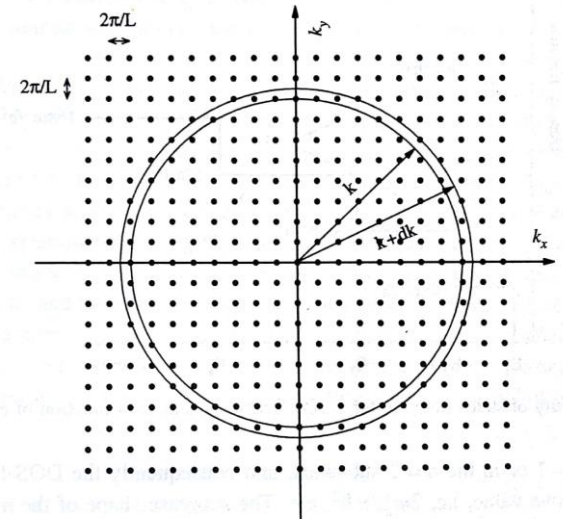


Fig: 1.10 Representation in k space for the states (kx,ky) of a 2D electron system

If we wish to calculate the DOS in energy, we define $n_{2D}(E)$ such that $n_{2D}(E)\ \delta E$ is the number of states in the range δ E. The densities of states in energy and wave vector are related by:

$n_{2D}(E)\ \delta E = n_{2D}(k)\ \delta k$ .....(7)

where E and k are related by $E = \frac{h^2k^2}{2m_e{}^*}$. Differentiating this expression and taking into account Eq. (6), we have, after substitution. in Eq. (7), and adding a factor 2 which accounts for the spin:

$$n_{2D}(E) = 2\frac{k}{2\pi}\frac{\delta k}{\delta E} = 2\frac{k}{2\pi}\frac{m_e{}^*}{h^2k} = \frac{m_e{}^*}{h^2\pi} \quad \text{.....(8)}$$

Note that in the 2D case, the DOS function is a constant, independent of energy. Let us show next that the DOS function, for the two-dimensional case, exhibits a staircase-shaped energy dependence (Figure 1.11) in which all the steps are of the same height, but located at energies En given by Eq. (4.2). In effect, from Figure 1.9 (c) it can be appreciated that the interval of energy between 0 and E1 is not allowed. For E such that E1 < E < E2 the electrons will be located in the subband corresponding to n = 1 and the value will be $\frac{m_e{}^*}{h^2\pi}$. For the energy interval between E1; and E3, the electrons can be located either in the n =1 or in the n = 2 subbands, and consequently the DOS function would be twice the above value, i.e. $\frac{2m_e{}^*}{h^2\pi}$, etc. The staircase shape of the $n_{2D}(E)$ function can be directly observed by optical absorption measurements. We have also

represented the parabolic 3D case in Figure 4.3, from which it can be appreciated that the differences between the 2D and 3D cases are more discernable for low values of n.

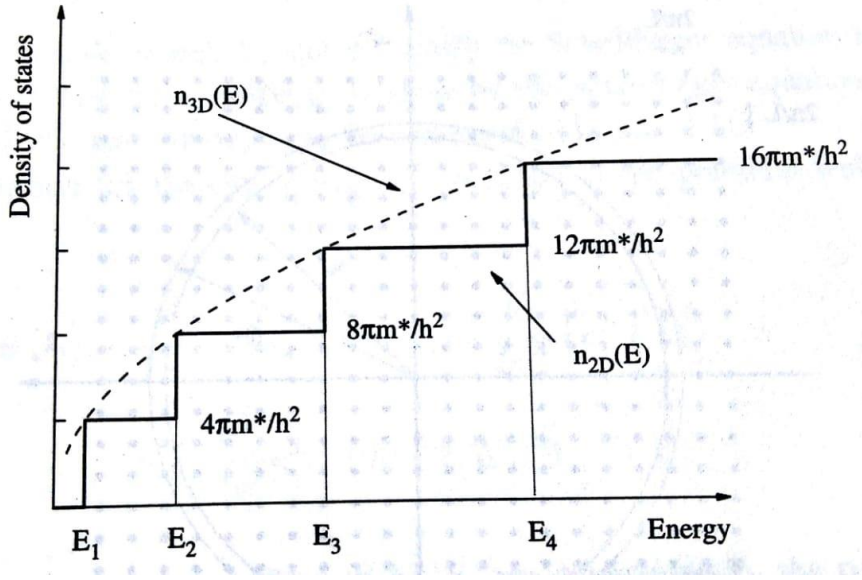


Fig: 1.11 Density states function for 2D electron system as a function of energy

According to the subband in which the electrons are located, their kinetic energy is differently partitioned as a consequence of the relationship expressed in Eq. (4). For instance, for a given energy in the interval E2 to E3, for the same value of energy, the electrons located in subband

n = 2 have higher energy in the z-direction, E2, than those in the n =1 subband for which the corresponding value is El. Therefore the energy of motion corresponding to the plane (px , py) should be smaller for the electrons in subband n = 2. Evidently the separation of energy in different "components" (remember that the energy is a scalar function) is a direct consequence of the simple forms adopted for **ψ** and energy in eqns 3 & .4

1.12 SQUARE QUATUM WELL OF FINITE DEPTH

The quantum wells for electrons and holes in GaAs nanostructures surrounded by higher gap AlGaAs,are not of infinite height.In fact, the value of the height of the potential for electrons should coincide with the discontinuity ΔEc that appears at the interface in the conduction bands of AlGaAs and GaAs, which for the above system is of the order of some tenths of an eV. However, it is fairly easy to deduce that for electron energies in the quantum well not too close to

the barrier ΔEC (we take the energy as zero at the bottom of the well), the values obtained for the case of infinite wells do not differ too much of those obtained in the case of finite depth wells.

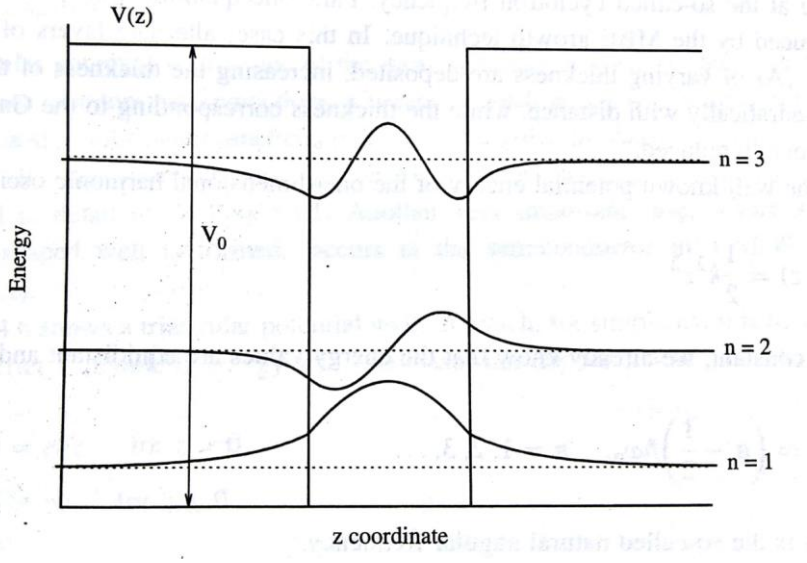


Fig 1.12: Finite potential square well

If we call Vo in Figure 1.12 the height of the finiite square well, it is evident that for states with energy E < Vo we have bound states, i.e. the electrons are trapped inside the well of width a, while for E > Vo, we have continuous propagation states, in which the electrons are free to move from z = -∞ to z = +∞. Since this problem presents inversion symmetry around the centre of the well, this point is chosen as origin for the z-direction. In relation to the bound states, the wave functions inside the well should have the same shape as in the case of the infinite well, i.e. the solutions are symmetric or antisymmetric, and therefore should be sine or cosine functions, respectively. We also know from quantum mechanics that the solutions outside the well, which are obtained from the Schrodinger equation with a potential energy equal to Vo, are exponential decay functions. Therefore the solutions for the wave functions should be linear combinations of the functions:

$$\psi_n(z) = \begin{cases} D\exp(kz), & z > -\frac{a}{2} \\ C\cos(kz), C\sin(kz), & -\frac{a}{2} < z < \frac{a}{2} \\ D\exp(-kz), & z > a/2 \end{cases} \quad \text{.....(1)}$$

Where

$$k = \left(\frac{2m_e{}^*E}{h^2}\right)^{1/2} \text{.....(2)}$$

$$k = \left(\frac{2m_e{}^*(Vo-E)}{h^2}\right)^{1/2} \text{ outside the well .....(3)}$$

Note that in Eqs (2 & 3) we have assumed the same value for the effective electron mass in the barrier and in the well. Usually, this is a good approximation because the barrier material is similar to the well material (e.g. GaAs and AlGaAs with a small Al mole fraction) and because the penetration of electron wave functions into the barriers is small for the lowest states. In order to proceed with the solution of the problem we should next realize that since (z) is a continuous function, therefore, the functions of Eq. (1) must be equal for z = ± a / 2, and the same should hold for their derivatives in the homogeneous effective mass approximation. From this fact a transcendental equation is derived which can be easily solved numerically . From this solution, several important facts are derived. For instance, it can be shown that in a one-dimensional well there is always at least one bound state, independent of how small the value of Vo might be. For the case of weakly bound states, the exponential decay constant k in Eq. (1) is small and therefore the wave function represented in Figure 1.12 penetrates deeply into the barrier region. Evidently, the opposite should be true for strongly bound states, in which the penetration in the energy forbidden region is very small.

1.13 PARABOLIC AND TRIANGULAR QUANTUM WELLS

Parabolic well

The case of the parabolic well is well known in solid state physics since the vibrations of the atoms in a crystal lattice, whose quantification gives rise to phonons, are described in a first approximation by harmonic oscillators. In addition, a magnetic field applied to a two-dimensional electron system gives rise to a parabolic potential, and the electrons oscillate at the so-called cyclotron frequency. Parabolic quantum well profiles can also be produced by the MBE growth technique. In this case, alternate layers of GaAs and $Al_xGa_{1-x}As$ of varying thickness are deposited, increasing the thickness of the AlGaAs layer quadratically with distance, while the thickness corresponding to the GaAs layer is proportionally reduced.

For the well-known potential energy of the one-dimensional harmonic oscillator

$V(z) = \frac{1}{2}\, k^2 z^2$ .....(1)

With k constant

$En = \left(n - \frac{1}{2}\right) h\omega_o \qquad n = 1,2,3 \ldots$ .....(2)

Where ω_o is the natural angular frequency.

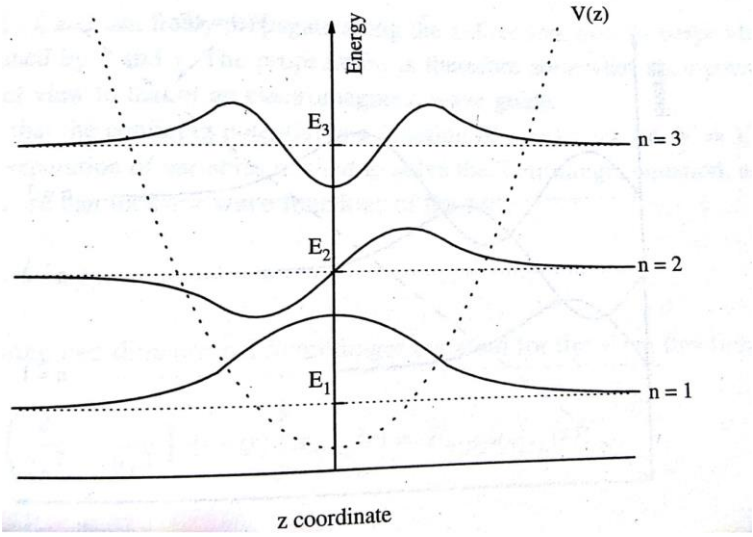


Fig 1.13 : Harmonic oscillator potential well V(z)

Figure 1.13 shows the case of a parabolic well potential and the wave functions of three electron bound states. The wave functions are mathematically expressed in terms of the Hermite polynomials. Note, as in the case of the square well, the symmetric or antisymmetric character of the wave functions and their exponential decay in the forbidden energy zone.

Triangular wells

The triangular potential well is one of the most common geometries, since the potential profile across quantum heterojunctions , such as the well-known modulation doped AlGaAs-GaAs heterojunction, is almost triangular in shape for electrons within GaAs. of all heterostructures, this is probably the most investigated one. Another very important case, where an almost triangular-shaped well is formed, occurs at the semiconductor in a MOS structure.

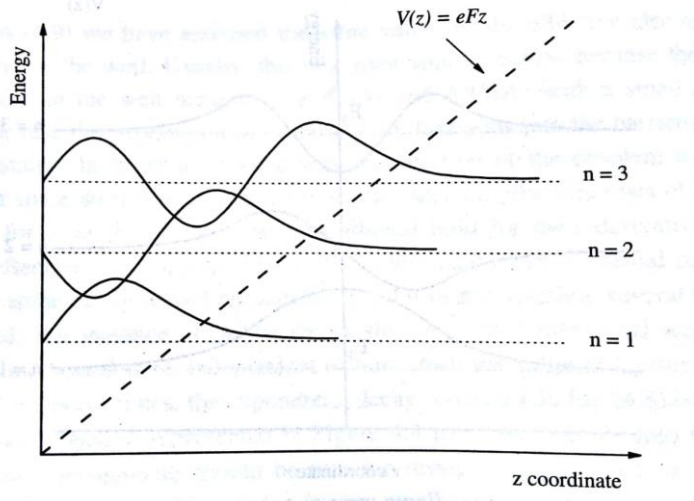


Fig 1.14 : triangular potential well V(z)

Figure 1.14 shows a triangular potential well, in which, for simplicity, it is assumed that the left barrier is infinite in energy and it increases linearly for z > 0:

$$V(z) = e\,Fz, \qquad for\ z > 0$$

$$V(z) = \infty, \qquad for\ z \leq 0$$

where e is a constant equal to the electron charge and F is a uniform electric held. As in the other cases, the electron energies and states are found by solving the Schrodinger equation subject to the boundary condition (z = 0) = 0. In this case, the eigenvalues are given in terms of the Airy functions. However, for small values of n, it can be demonstrated by applying the WKB quantum-mechanical approximation, that

$$En \approx \left[\frac{3}{2}\pi\left(n-\frac{1}{4}\right)\right]^{2/3}\left[\frac{e^2h^2F^2}{2m}\right]^{1/3}, \qquad \text{n=1,2,3,....}$$

Figure 1.14 shows the spacing between the energy levels, which get a little closer as n increases, in contrast to the square well where the levels become further apart as n increases (in the parabolic case they were equally spaced). In the same figure, the wave functions are also represented. Observe that as n increases, the wave functions add one more half' cycle. However at difference with the previous parabolic case, the wave functions are neither symmetric or antisymmetric due to the asymmetry of the potential well.

1.15 QUANTUM WIRES

Having considered the two-dimensional electron gas , it is easy to understand that in the one-dimensional electron gas, the electrons should be confined in two directions, (x, y), and can freely propagate along the z-direction, usually perpendicular to the plane defined by x and y. The propagation is therefore somewhat analogous from a formal point of view to that of an electromagnetic wave guide.

Supposing that the confining potential is a function of r = (x , y), ie. V = V (r), and following the separation of variables method to solve the Schrodinger equation, we can look for wave functions of the form:

$$\psi(r) = e^{jk_z z}u(r) \text{.....(1)}$$

With the following 2D Schrodinger equation for wave function u(r)

$$\left[-\frac{h^2}{2m}\left(\frac{d}{dx^2}+\frac{d}{dy^2}\right)+\boldsymbol{V(r)}\right]u_{n1n2}(r) = E_{n1n2}(r) \text{ .....(2)}$$

Where n1,n2 = 1,2,3.....

The expression for the total energy of electron in the quantum wire should be of the form

$$E_{n1n2}(kz) = E_{n1n2} + \frac{h^2 {k_z}^2}{2m_e^*} \quad \text{.....(3)}$$

where the last term represents the kinetic energy of the electron propagating along the z-direction.

As an example to get specific expressions for the energy, we consider now the simplest case of a two-dimensional rectangular potential of infinite depth and size ax, ay,. That is:

$$V(x,y) = 0 \qquad 0 < x < ax, 0 < y < ay$$

$$V(x,y) = \infty \qquad x \le 0\,, \quad x \ge ax, y \le 0\,, y \ge ay$$

Now energy

$$E_{n1n2} = \frac{h^2\pi^2}{2m_e^*}\left\{\frac{{n_1}^2}{{a_x}^2} + \frac{{n_2}^2}{{a_y}^2}\right\}, \quad \text{n1, n2 = 1,2,.....} \quad \text{.....(4)}$$

Another case, relatively easy to solve in cylindrical coordinates, is that of a wire with circular cross section, in which case the solutions are given in terms of the Bessel functions. Therefore in the case of quantum wires, the energy levels corresponding to the transverse direction are specified by two quantum numbers, and each value $E_{n1,n2}$ is now the bottom Of a parabolic one-dimensional subband in kz space. Observe also that as the electron moves in a narrower wire, the energy corresponding to the $E_{n1,n2}$ levels increases; .

Let us now calculate w density of states for the one-dimensional electron gas. The concentration of states in energy is related to that in wave number by the expression:

$$n_{1D}(E)\partial E = n_{1D}(E)\frac{dE}{dk}\partial k = 2n_{1D}(k)\partial k \quad \text{.....(5)}$$

The factor 2 appears because the wave number could be either positive or negative corresponding to the two directions along the wire. The density of states in k-space per unit length is $1/2\pi$, as seen from considering the one-dimensional version of Figure 1.10 Substituting in Eq. (5) and taking into account that $E = \frac{h^2k^2}{2m_e^*}$, we obtain:

$$n_{1D}(E) = \frac{1}{\pi h}\sqrt{\frac{2m_e^*}{E}} \quad \text{.....(6)}$$

Which diverge for E=0

In terms of group velocity vg,

$$n_{1D}(E) = \frac{2}{\pi h v_g} \quad \text{.....(7)}$$

One interesting result of this equation is that the current in a one-dimensional system is constant and proportional to the velocity and the density of states. The expression (7) for the DOS function in a quantum wire, will have important consequences as, for example, the quantized conductance.

The expression of the total DOS per unit length for a quantum wire can be expressed as

$$n_{1D}(E) = \sum_{n1,n2} \frac{1}{\pi h} \sqrt{\frac{2m_e^*}{E - E_{n1,n2}}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(8)$$

Figure 1.15 represents the DOS for a one dimensional System, which is compared to the parabolic 3D case. Now the DOS diverges at the bottom of the subbands given by the energy values $E_{n1,n2}$ This result will have important consequences in the physical properties of quantum wires.

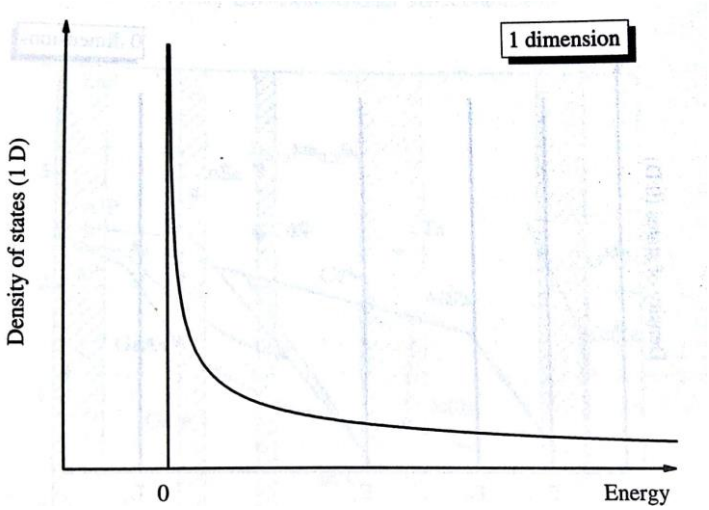


Fig 1.15 : DOS function for 1D electron system, as a function of energy

1.16 QUANTUM DOTS

Quantum dots are often nano crystals with all three spatial dimensions in the nanometer range. Sometimes, as is the case of the II-VI materials, such as CdSe or CdS, the nano crystals can be grown from liquid phase solutions at well-specified temperatures. Conversely they can also be prepared by lithographic etching techniques from macroscopic materials

Although the word "dot" implies an infinitely small size, in practice dots might have a large number of atoms: 10^4-10^6, and still have their three dimensions in the nanometre legion, so that the electron de Broglie wavelength is comparable to the size of the dot. In this case, the wave nature of the electron becomes important. Quantum dots are often referred to as ***artificial***

atoms, because, the spectrum of the energy levels resembles that of an atom. In addition, at least in theory the energy spectrum can be engineered depending on the size and shape of the dot. In analogy to atoms, we can also define an ionization energy, which accounts for the energy necessary to add or remove an electron from the dot. This energy is also called the ***charging energy*** of the dot, in an image similar to the concept of capacitance of a body, in which the addition or subtraction of electric charge is specified by the Coulomb interaction. Therefore the atom-like properties of the quantum dots are often studied via the electrical characteristics. From this point of view, it is very important to remark that even the introduction or removal of one single electron in quantum dots, in contrast to the case of 2D or 1D systems, produces dominant changes in the electrical characteristics, mainly manifested in large conductance oscillations and in the Coulomb blockade effect .

Let us now study the energy spectrum of quantum dots. The simplest case would be that of a confining potential that is zero inside a box of dimensions ax, ay,, and az, and infinite outside the box. Evidently the solution to this well-known problem are standing Waves for the electron wave function and the energy levels are given by

$$E_{n1,n2,n3} = \frac{h^2\pi^2}{2m_e{}^*}\left(\frac{{n_1}^2}{{a_x}^2}+\frac{{n_2}^2}{{a_y}^2}+\frac{{n_3}^2}{{a_z}^2}\right) \quad ; \quad n1, n2, n3 = 1,2,3, \ldots.$$

In contrast to the 2D and 1D cases, now the energy is completely quantized, and, as in the case of atoms, there is no free electron propagation. However, the levels are frequently degenerate, for instance, if two or three of the dimensions of the box are equal.

The case of a spherical dot in which the potential is zero inside the sphere and infinite outside can also be exactly solved, and the solutions are expressed in terms of the Spherical functions. This problem resembles that of a spherically symmetric atom and the energy depends on two quantum numbers, the principal quantum number n, arising from the one-dimensional radial equation, and the angular momentum quantum number *l.*

Since in the case of quantum dots the electrons are totally confined, the energy spectrum is totally discrete and the DOS function is formed by a set of peaks in theory with no width and with infinite height (Figure 1.16).

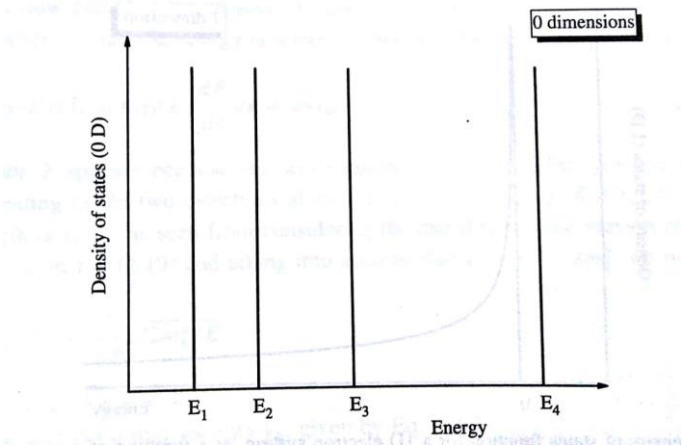


Fig 1.16: DOS function for 0D electron system.

Evidently, in practice, the peaks should have a finite width, as a consequence, for instance, of the interaction of electrons with lattice phonons and impurities.

1.17 CARBON NANOTUBES

One of the greatest impacts on nanoscience and nanotechnology has been made by the carbon nanotubes. A schematic illustration of carbon nanotubes is presented in Fig, and an atomically resolved image generated using scanning tunnelling microscope.

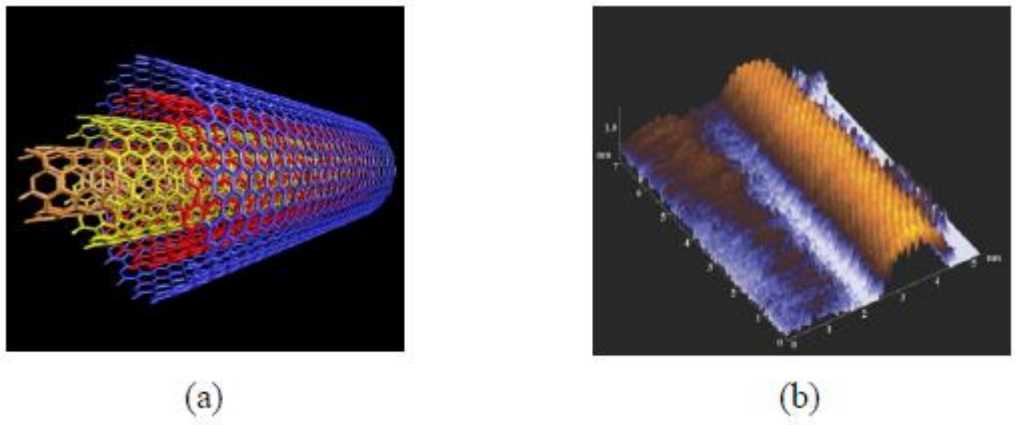


Fig 1.2 . Carbon nanotube.

- The nanotubes can be single- and multi-walled (which are abbreviated as SWNT and MWNT, respectively). The reason for this is that the carbon nanotubes possess remarkable mechanical, electrical and thermal properties that equal, or even surpass, those of other benchmark materials such as steel, copper, and diamond, respectively. The Young's modulus of the nanotubes can be as high as 1000 GPa (which is approximately

five times higher than steel) while their tensile strength can be up to 63 GPa (around 50 times higher than steel). These

properties, coupled with their low density, give nanotubes huge potential in a range of structural applications. Applications on the nanometer and micrometer scale, such as SWNT-based transistors and chemical sensors are progressing rapidly.

The multi-walled carbon nanotubes consist of several nested coaxial singlewalled tubules. Typical outer and inner diameters of multi-walled carbon nanotube are 2–20 nm and 1–3 nm, respectively. The typical length is 1–100 μm. The intertubular distance is 0.34 nm.

1.17.1 GRAPHENE

Graphene is the basic structural element of carbon nanotubes. It is one-atom thick planar sheet of (*sp*2 bonded) carbon atoms that are densely packed in a honeycomb crystal lattice, as shown in Fig

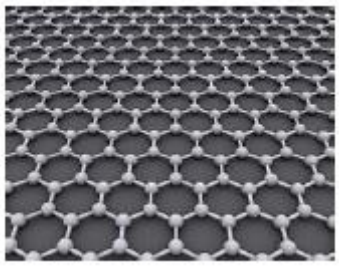


Fig.1.3 Graphene

- The definition of graphene is "a flat monolayer of carbon atoms tightly packed into a two-dimensional honeycomb lattice, which is a basic building block for graphitic materials of all other dimensionalities. It can be wrapped up into fullerenes, rolled into nanotubes or stacked into graphite".
- Graphene is stronger and stiffer than diamond. It, however, can be stretched like rubber. Its surface area is the largest known for its weight.
- The C–C bond length in graphene is ~0.142 nm. The graphene sheets stack to form graphite with an interplanar spacing of 0.335 nm, which means that a stack of $3x10^6$ sheets will be ~1 mm thick.

SWNTs and MWNTs

· The synthesis of C_{60} and other fullerenes stimulated intense interest in the tubular variant of the fullerenes. Iijima (1991) prepared multi-walled nanotubes as shown in Fig

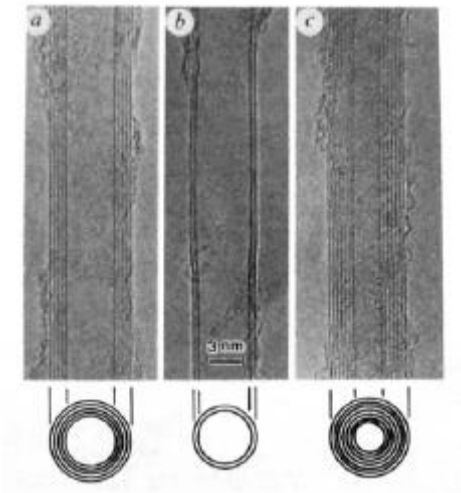


Fig 1.4 .: Electron micrographs of nanotubes of graphitic carbon. The parallel dark lines correspond to the (002) lattice images of graphite. A cross-section of each tubule is illustrated: (a) tube consisting of five graphitic sheets, diameter 6.7 nm, (b) two-sheet tube, diameter 5.5 nm, and (c) seven-sheet tube, diameter 6.5 nm, which has the smallest hollow diameter of 2.2 nm

These nanotubes were grown on the negative end of the carbon electrode used in the dc arc-discharge evaporation of carbon in an argon-filled vessel at 13.3 kPa pressure.

- The tubes grew on certain regions of the electrode. They were made of a single curved sheet of graphite connected at its edge. The electrical energy and the energy associated with the dangling bonds in a graphite sheet sealed the rolled-up sheets in the form of nanotubes.
- Electron microscopy revealed that each needle comprised coaxial tubes of graphitic sheets, ranging in number from 2 up to about 50. On each tube, the hexagons of carbon atom were arranged in a helical fashion about the needle axis. The helical pitch varied from needle to needle and from tube to tube within a single needle. The diameter of the tubes varied between 4 nm and 30 nm, and their length was up to 1 μm.
- The single-walled nanotubes were synthesized in 1993 by Iijima and Ichihashi, and Bethune *et al*. Iijima and Ichihashi used two vertical electrodes in the carbonarc chamber. The anode was a graphitic carbon rod and the cathode was a carbon rod. The latter held a small piece of iron. The chamber was filled with a mixture of methane and argon at low pressure. The carbon discharge arc was generated by running a dc current of 200 A at 20 V between the electrodes. The iron melted and formed a droplet. The iron vapor cooled and condensed into small particles of iron carbide on the electrode. Iron acted as catalyst in the vapor phase in the formation of the single-walled nanotubes. The nanotubes were

found in the sootlike deposits. The diameter of the tubes varied between 0.7 nm and 1.6 nm.

- Bethune *et al.* (1993) synthesized single-walled carbon nanotubes by covaporizing carbon and cobalt in an arc generator. The tubes had about 1.2 nm diameter. The tubes formed a web-like deposit woven through the fullerenecontaining soot, giving it a rubbery texture.
- Thess *et al*. (1996) presented an efficient route for the synthesis of arrays of single-walled nanotubes. Their process involved condensation of a laservaporized carbon□nickel□cobalt mixture at 1473 K. The yield was more than 70%. X-ray diffraction and electron microscopy showed that these single-walled nanotubes were almost uniform in diameter and they self-organized into 'ropes' which consisted of 100 to 500 nanotubes in a two-dimensional triangular lattice.

The ropes were metallic with a single-rope resistivity of less than 10□6 □ m at
300 K.

- The high-pressure-carbon-monoxide (HiPCO) method developed by Nikolaev *et al.* (1999) can produce large quantities of carbon nanotubes. The catalyst for the growth of single-walled nanotube form *in situ* by thermal decomposition of iron pentacarbonyl in a heated flow of carbon monoxide at pressures in the range of 101.325 to 1013.25 kPa, and temperatures between 1073 and 1473 K. The yield of SWNT and the diameter of the nanotubes produced by this process can vary over a wide range, which is determined by the condition and geometry of the flow-cell.
- Nikolaev *et al*. (1999) produced SWNTs of 0.7 nm diameter. The products of thermal decomposition of Fe(CO)5 reacted to produce iron clusters in the gas phase. These clusters acted as nuclei upon which the SWNTs nucleated and grew. The solid carbon was formed through disproportionation of CO (known as the *Boudouard reaction*).

$$CO + CO \rightarrow C(s) + CO_2$$

- This reaction occurred catalytically on the surface of the iron particles. The iron particles promoted the formation of the tube's characteristic graphitic carbon lattice. The rate at which the reactant gases were heated determined the amount and quality of the SWNTs produced. The temperature and pressure had important effects on the yield of the nanotubes. TEM image of the single-walled nanotube produced by this method is shown in Fig.

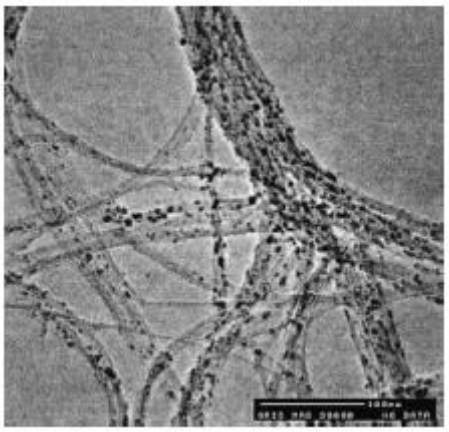


Fig 1.5 : . High-magnification TEM image of single-walled nanotube produced by the HiPCO process

- Ericson *et al*. (2004) produced well-aligned macroscopic fibers of single-walled carbon nanotubes from the concentrated dispersions of SWNTs in 102% sulphuric acid employing a wet-spinning technique. Because of the high-temperature stability of the SWNTs, melt-spinning is not a suitable option and wet-spinning is the only viable approach. The main challenge to the production of neat SWNT fibers is dispersing the SWNTs at high enough concentrations suitable for efficient alignment and effective coagulation.
- However, due to their chemical inertness and strong van der Waals attractions, SWNTs aggregate into ropes with limited solubility in aqueous, organic or acidic media. If a surfactant is used to disperse the SWNTs, there are complications of removing the surfactant from the fiber during coagulation or after processing. In sulfuric acid of concentration greater than 100%, SWNTs form charge-transfer complexes of individual positively charged nanotubes surrounded by a finite number of sulfuric acid anions.
- At very low concentrations, such charged tube-anion complexes behave as Brownian rods. At concentrations greater than 0.03 wt. %, a small amount of dissolved individual tubes coexists with a SWNT spaghetti-phase consisting of seemingly endless swollen ropes of well-aligned positively charged SWNTs intercalated by sulfuric acid anions.
- The SWNTs in the spaghetti are mobile and at a high enough concentration (> 4 wt. %), they coagulate and form ordered domains behaving similarly to nematic liquid crystalline rod-like polymers. The SWNT-acid system is very sensitive to water, and if a very small amount of moisture enters the system, phase separation occurs and discrete needle-like crystal-solvates precipitate. This ordered SWNT dispersion can be extruded and

coagulated in a controlled fashion using conventional fiber-spinning techniques to produce continuous lengths of macroscopic neat SWNT fibers.

- The scanning electron micrographs depicting the aligned macroscopic fibers consisting solely of the single-walled nanotubes are shown in Fig. 1.7 . The morphology of the fibers was strongly dependent on the coagulation conditions.

1.17.2 STRUCTURE OF CARBON NANOTUBES

In the theoretical analysis of carbon nanotubes, the major focus has been on single-walled tubes, cylindrical in shape with caps at each end, such that the two caps can be joined together to form a fullerene. The cylindrical portions of the tubes consist of a single graphene sheet (a single layer of carbon atoms from a three dimensional graphite crystal) that is shaped to form the cylinder.

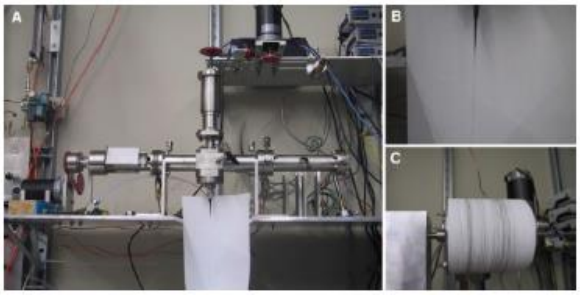


Fig 1.6: The spinning process of single-walled carbon nanotubes (SWNTs) in 102% sulfuric acid: (A) the apparatus for mixing and extruding neat fibers, (B) a jet of SWNT dispersion being extruded from a capillary tube, and (C) a 30 m spool of water-coagulated fiber

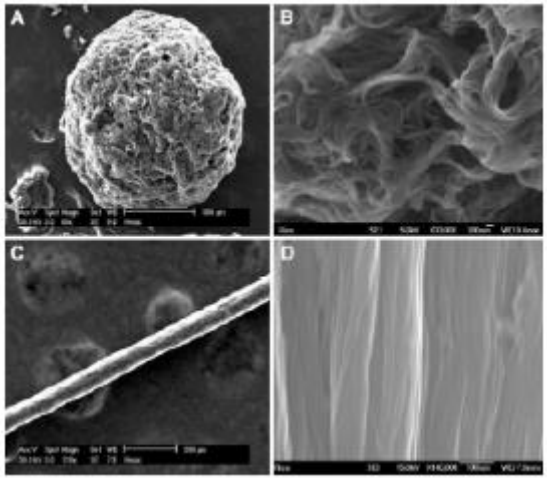


Fig. 1.7: SEM images showing the evolution of purified SWNTs into continuous fiber:
(A) SWNTs after the purification process, (B) view inside the purified SWNTs showing the tangled mass of SWNT ropes that are 20–30 nm in diameter, (C) an annealed neat SWNT fiber spun from 8 wt.% dispersion in 102% sulphuric acid and coagulated in water, and (D) higher magnification of the neat fiber surface showing that all the ropes have merged into aligned super-ropes that are 200 nm or larger in diameter

There are two possible high-symmetry structures for nanotubes, known as 'zigzag' and 'armchair'. In practice, it is believed that most nanotubes do not have these highly symmetric forms but have structures in which the hexagons are arranged helically around the tube axis. These structures are generally known as 'chiral', because they can exist in two mirror-related forms.

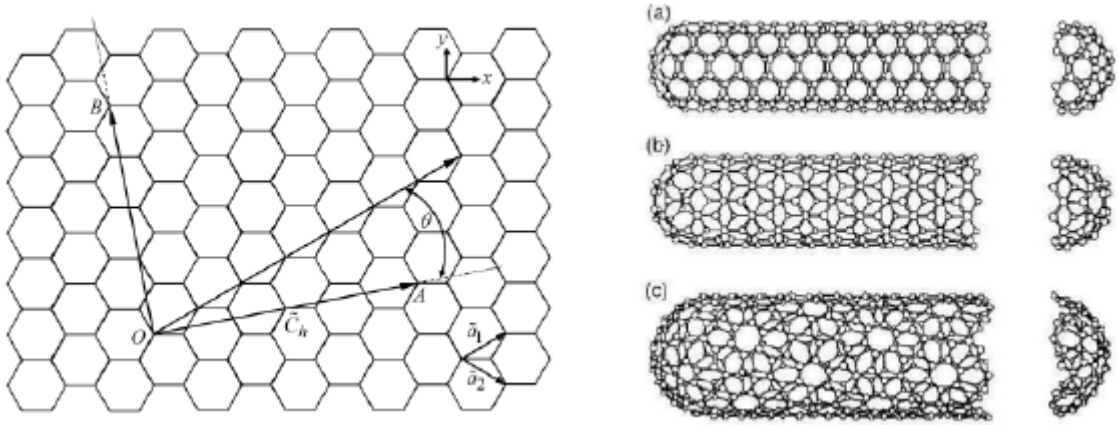


Fig 1.8: Schematic model for single-walled carbon nanotubes with the tubeaxis, *OB*, normal to: (a) θ = π/6 rad direction (an 'armchair' nanotube), (b) θ = 0 direction (a 'zigzag' nanotube), and (c) 0 < θ < π/6 rad (a 'chiral' nanotube)

It is convenient to describe a carbon nanotube in terms of the chiral vector, $\widetilde{C_h}$, and the chiral angle, θ , as shown in this Fig. 1.8. The points, *O*, and A, are crystallographically equivalent on the graphene sheet. These points are connected by the chiral vector, $\widetilde{C_h}$, which is defined in terms of the basis vectors, $\widetilde{a_1}$ and $\widetilde{a_2}$, of the honeycomb lattice.

$$\widetilde{C_h} = n\widetilde{a_1} + m\widetilde{a_2}$$

where *n* and *m* are integers. The chiral angle, θ , is shown in the figure. *OB* represents the normal to $\widetilde{C_h}$ at point *O*. Depending on the value of the chiral angle, a single-walled carbon nanotube can have three basic geometries: armchair, zigzag and chiral, as illustrated in the Fig 1.8

- The diameter of a carbon nanotube is given by

$$d_t = \frac{|\widetilde{C_h}|}{\pi}$$

The integers *n* and *m* uniquely determine *dt* and θ

- The circumference of the nanotube is given by

$$L = |\widetilde{C_h}| = a(n^2 + m^2 + nm)^{\frac{1}{2}} \qquad 0 \le |m| \le n$$

where *a* is the length of unit vector

Problem 1. Calculate the diameter of a single-walled nanotube if $n = m = 5$ and the length of the unit vector is 0.246 nm.

$$L = \left|\widetilde{C_h}\right| = a(n^2 + m^2 + nm)^{\frac{1}{2}}$$

$$L = .246 \times \sqrt{5^2 + 5^2 + 5 \times 5} = 2.13\text{nm}$$

$$d_t = \frac{|\widetilde{C_h}|}{\pi} = \frac{2.13}{\pi} = 0.678nm$$

- The carbon nanotubes can either have metallic conductivity or can be semiconducting depending on the tube diameter and the chiral angle. Metallic conduction in carbon nanotubes can be achieved without the introduction of doping or defects.
- Ballistic electron transport and electrical conductivity without phonon and surface scattering have been observed in metallic carbon nanotubes. Semiconducting nanotubes have a band gap between their conduction and valence bands, which is proportional to $1/dt$.
- It is possible to fabricate a transistor based on a single helical carbon nanotube. This is a step towards molecular electronics

An AFM image of a single nanotube contacting three platinum electrodes is shown in Fig

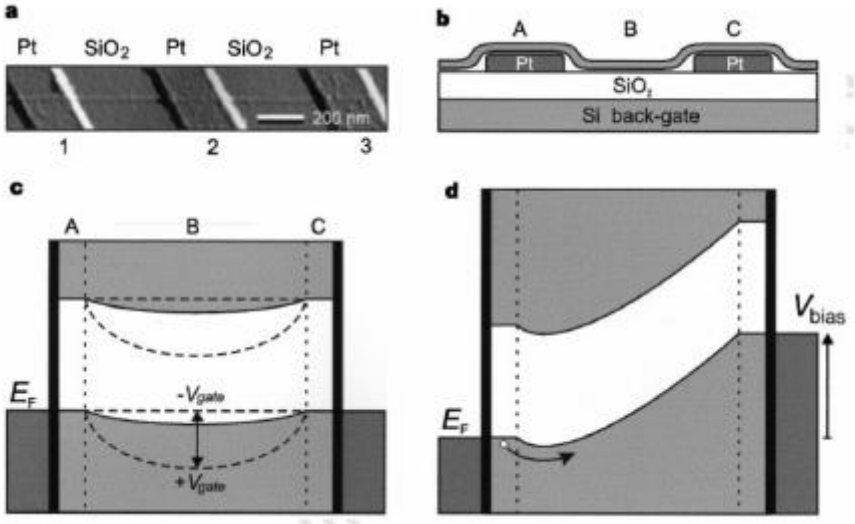


Fig. 1.9 (a) Tapping-mode AFM image of an individual carbon nanotube on top of three Pt electrodes, (b) Schematic side view of the TUBEFET (single carbon nanotube field-effect transistor) device. A single semiconducting nanotube is contacted by two electrodes. The Si substrate, which is covered by a layer of SiO2 300 nm thick, acts as a back-gate, (c) Suggested band diagram of the device. The nanotube with a gap of ~0.6 eV is connected to the leads with Fermi energy *EF* by tunnelling contacts, indicated by the black vertical bars. At A and C (see b), the valence-band edge is pinned to the Fermi energy of the leads. Owing to a difference in work function between the tube and the electrodes, the bands bend towards lower energy in between the electrodes (B). For positive *V*gate the bands bend more strongly, leading to an insulating state. For negative *V*gate the bands flatten, resulting in a metal-like conductance, and (d), Application of a bias voltage results in a suppression of the barrier

The semiconductor Si-substrate covered with a 300 nm layer of thermally grown SiO2 was used as a back-gate. The metallic variety of tubes had linear current I versus voltage (Vbias) curves, and showed no dependence on the gate voltage (Vgate). For the sample shown in the figure, the I versus Vbias curve was slightly nonlinear for Vgate = 0 . When Vgate was increased to positive values, a pronounced gap-like nonlinearity developed around Vbias = 0 .

- The curves exhibited a power-law behavior: $I \propto (V_{bias})^{\propto}$ with $\propto$ lying between 1 and 12. Upon application of a negative Vgate , the I versus Vbias curve became linear with a resistance that saturated around 1 MΩ. This resistance is mainly due to the contact resistance between the tube and the electrodes. Therefore, the device showed controllable semiconductor-to-metal transition in a one dimensional system. The nonlinearity at room temperature and the asymmetric dependence of the conductance on the gate-voltage polarity indicate that the nanotube is semiconducting. The conductance could be modulated by about six orders of magnitude if Vgate was changed by 10 V.
- Both nanowires and single-walled nanotubes can be used to make molecular field-effect transistors. Advanced electronic devices based on carbon nanotubes and various types of nanowires can have a very important role in next-generation semiconductor architectures. However, at present there is lack of a general fabrication method which has held back the development of these devices for practical applications.
- A few assembling strategies have been suggested for devices based on these nanomaterials. Lee *et al*. (2006) used inert surface molecular patterns to direct the adsorption and alignment of nanotubes and nanowires on bare surfaces to form device structures without the use of linker molecules. They have used this method to demonstrate large scale assembly of nanotube- and nanowire-based integrated devices and their applications.

1.17.3 CARBON NANOTUBE COMPOSITE MATERIALS

- Many of the outstanding properties of carbon nanotubes can be best utilized by incorporating the nanotubes into some form of matrix. The exceptional mechanical properties, in particular, have prompted huge interest in the production of composite

materials containing nanotubes for structural applications using polymers, ceramics and metals.

- The materials which form composites with carbon nanotubes are mostly polymers. The simplest method for preparing nanotube polymer composites is mixing the nanotube dispersion with solution of the polymer, and then evaporate the solvent in a controlled way. This method has been used with various polymers such as polyvinyl alcohol, polystyrene, polycarbonate and poly(methyl methacrylate).
- In order to facilitate solubilization and mixing, the nanotubes are often functionalized prior to adding to the polymer solution. For example, acid treatment is used to disperse catalytically produced MWNTs in water, and then nanotube–polymer composite is made by simply mixing one of these dispersions with an aqueous solution of the polymer and casting the mixture as film.
- The solution mixing approach is limited to polymers that freely dissolve in common solvents. An alternative is to use thermoplastic polymers and apply melt processing techniques. One problem associated with this method is that achieving homogeneous dispersions of nanotubes in melts is generally more difficult than with solutions, and high concentrations of tubes are hard to achieve due to the high viscosities of the mixtures.
- An alternative method for preparing nanotube polymer composites is to use the monomer rather than the polymer as a starting material, and then carry out *in situ* polymerization.
- Ceramics have high stiffness and thermal stabilities but relatively low breaking strengths. Incorporating carbon nanotubes into a ceramic matrix produces a composite with both toughness and high-temperature stability. However, achieving a homogeneous dispersion of tubes in an oxide, with strong bonding between tubes and matrix, presents rather more of a challenge than incorporating tubes into a
- Nanotubes have been used to improve the anti-static properties of fuel-handling components and body panels of automobiles. Nanotube-containing sporting equipment (e.g., tennis rackets and high-performance racing bicycles) has been manufactured recently
- It has been shown that considerable improvement in the performance of Li-ion batteries can be achieved by the addition of catalytically-produced MWNTs.

Carbon nanotube reactor

- Carbon nanotubes (CNTs) have well-defined hollow interiors and exhibit exceptionally good mechanical and thermal stability as well as electron conductivity. This opens up possibilities to introduce other matter into the cavities, which may lead to nanocomposite materials with interesting properties or behavior different from the bulk.
- Pan *et al.* (2007) have reported enhancement of the catalytic activity of Rhodium particles confined inside nanotubes for the conversion of CO and H2 to ethanol. The overall formation rate of ethanol inside the nanotubes exceeds that on the outside of the nanotubes by more than an order of magnitude, although the latter is much more accessible.
- The production of ethanol inside a carbon nanotube reactor containing catalytic particles is schematically shown in Fig

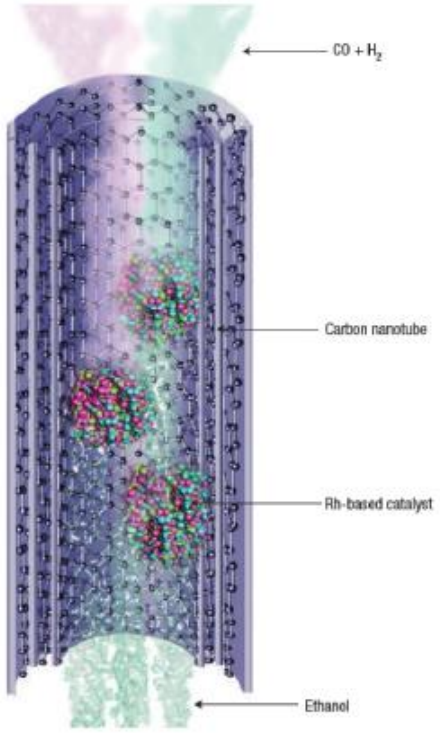


Fig 1.10 : **Schematic diagram showing ethanol production from syngas inside Rh-loaded carbon nanotubes. The black spheres denote carbon atoms, which form the graphene layers of the carbon nanotubes. The streams in light orange and green entering the nanotubes indicate the gas mixture of CO and H2, respectively. The three stacks of small spheres in rose, blue, green and red inside the tubes represent catalyst particles that may comprise more than one component. The streams in light cyan tailing behind the catalyst particles along the axis of the nanotubes represent ethanol**

MODULE II
FABRICATION OF NANO-LAYERS & NANO- PARTICLES

2.1 PRODUCTION OF NANO-LAYERS

2.1.1 Physical Vapor Deposition:

In general, physical vapor deposition (PVD) from the gas phase is subdivided into four groups.

- Evaporation,
- Sputtering,
- Ion Plating,
- Laser Ablation.

The first three methods occur at low pressures.

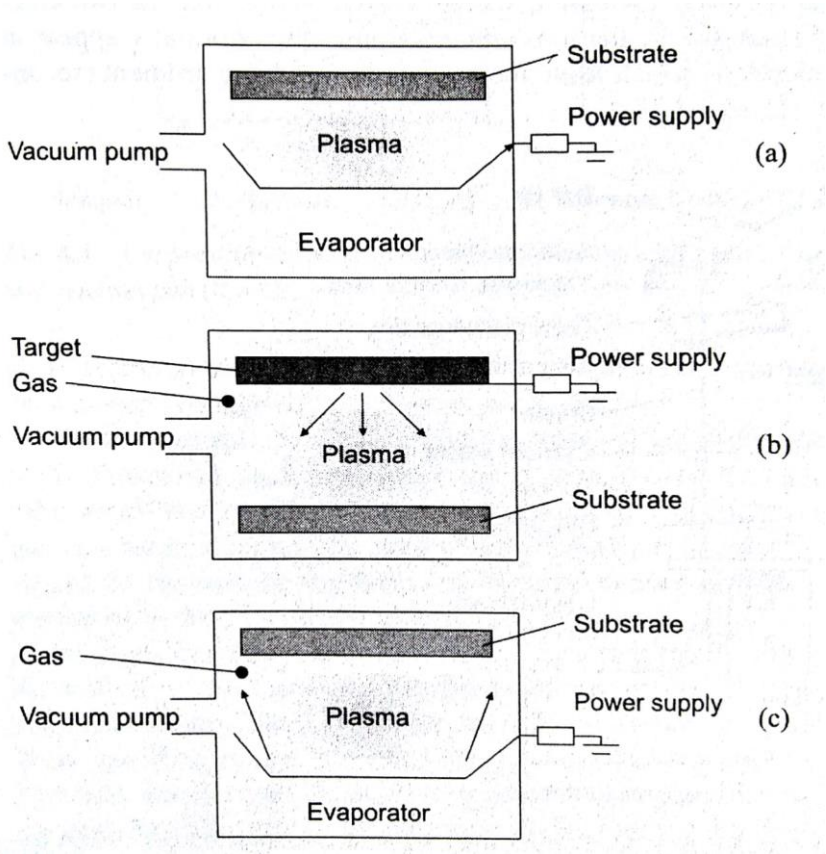


Fig:2.1 Three fundamental PVD (a).Evaporation (b). Sputtering (c.).Ion Implanting

1. Evaporation.

This procedure is carried out in a bell jar as depicted in Fig. 2.2

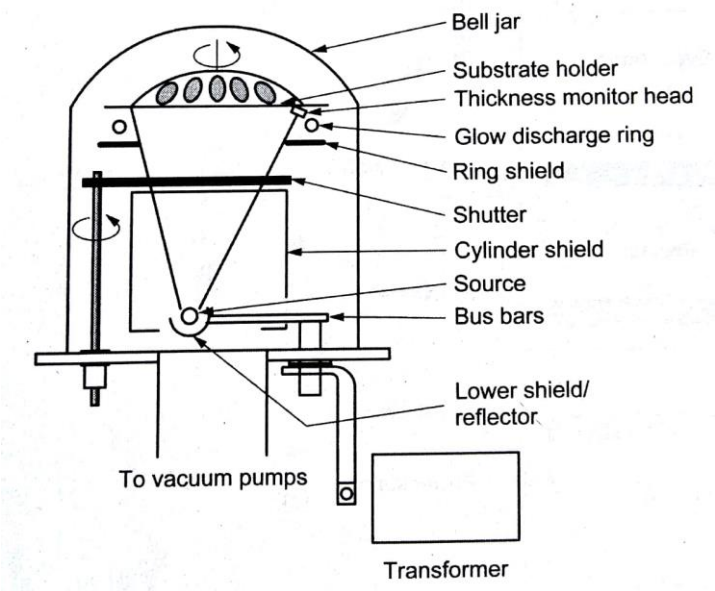
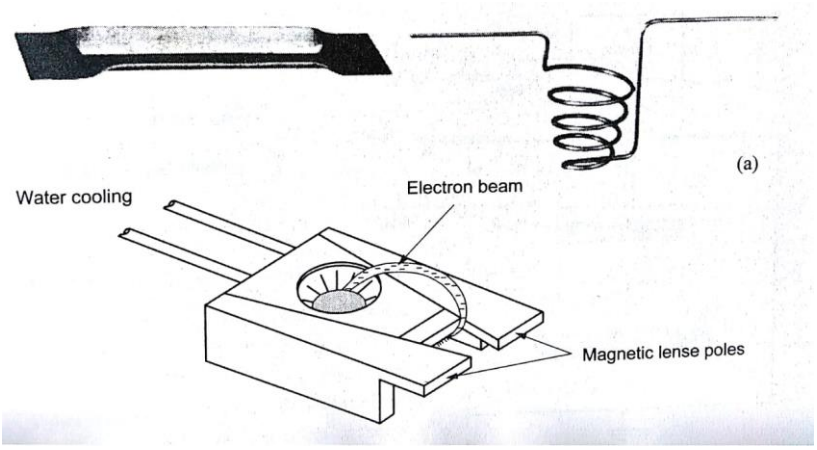
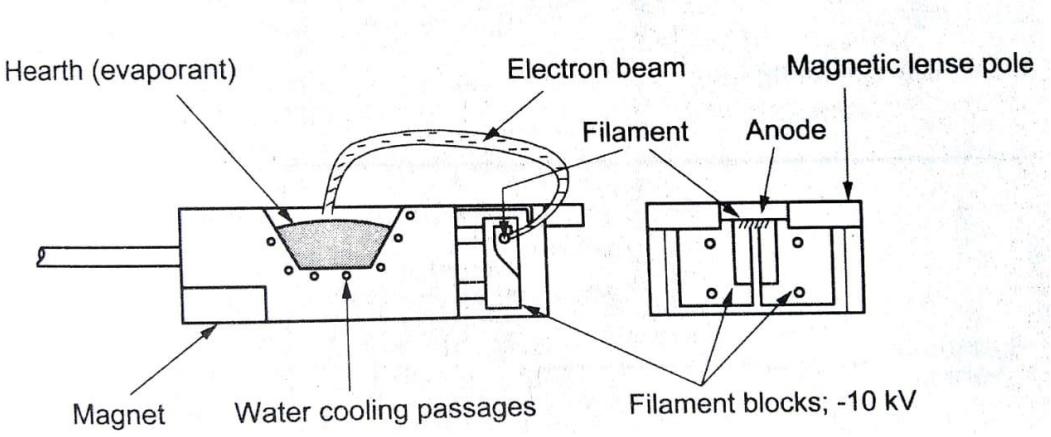


Fig 2.2 Vacuum system for the vaporization from resistance –heated sources.

A crucible is heated up by a resistance or an electron gun until a sufficient vapor pressure develops. As a result, material is deposited on the substrate. Technically, the resistance is wrapped around the crucible, or a metal wire is heated up by a current and vaporized. The electron gun (e-gun) produces an electron beam of, e.g., 10 keV. This beam is directed at the material intended for the deposition on the substrate. The gun's advantage is its unlimited supply of evaporating material and applicability of non-conductive or high-melting materials. Its shortcomings lie in the production of radiation defects, for instance in the underlying oxide coating. Both procedures are more precisely depicted in Fig. 2.3



2.3 (a). Evaporation by means of resistance-heating with a tungsten boat & winding



2.3 (b). Electron Gun

(II) SPUTTERING.

In literature, there is no clear definition of the term sputtering. Generally, an atom or a molecule, usually in its ionized form, hits a solid state (target) and knocks out surface atoms. This erosion is accompanied by a second process, namely the deposition of the knocked out atoms on a second solid state (substrate). The latter process is relevant when forming thin layers.

(a) Glow discharge. In its simplest form, sputtering is achieved by glow discharge with dc voltage. A cross section of the arrangement is schematically represented in Fig. 2.4.

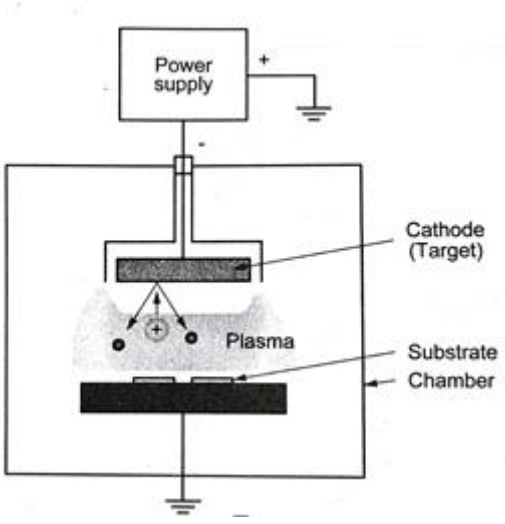


Fig: 2.4 : DC Voltage Sputtering

After mounting the samples on a holder, the chamber is rinsed repeatedly with Ar. Eventually, a constant gas pressure of some 100 mPa is built up. The target, being attached a few centimeters above the substrate, is raised to a negative dc potential from -500 to -5000 V, while both chamber and substrate are grounded. The discharge current requires a conducting target.

When the voltage is slowly increased, a small current flows over the two electrodes. This current is caused by the ions and electrons which normally appear in the gas and by the electrons which leave the target after ion bombardment (secondary electrons). At a certain voltage value, these contributions rise drastically. The final current-voltage curve is shown in Fig. 2.5.

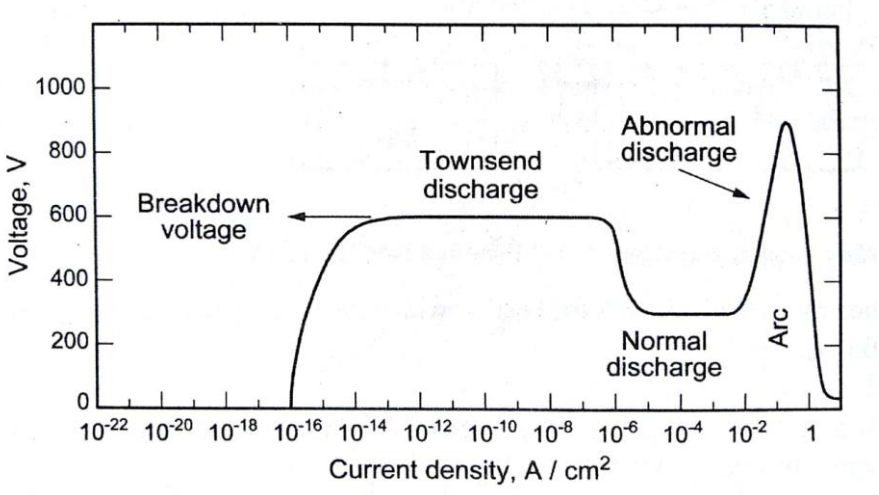


Fig 2.5: Applied Voltage *vs* Discharge current

The first plateau (at 600 V in our example) of the discharge current is referred to as Townsend discharge. Later the plasma passes through the "normal" and "abnormal" ranges. The latter is the operating state of sputtering. A self-contained gas discharge requires the production of sufficient secondary electrons by the impact of the ions 9n the target surface and .conversely, the production of sufficient ions in the plasma by the secondary electrons.

(b) High frequency discharge. When replacing the dc voltage source from Fig. 2.4 with a high frequency generator (radio frequency, RF, generator), target and substrates erode alternately depending on the respective polarity.

But even with these low frequencies, a serious shortcoming becomes apparent: due to the substantially small target surface (compared to the backplate electrode consisting of the bell, the cable shield, etc.) a proportionally large ion current flows if this

backplate electrode is negatively polarized. This would mean that the substrates are covered with the material of the bell, which is not intended.

In order to overcome this shortcoming, a capacitor is connected in series between the high frequency generator and the target, and/or the conducting target is replaced with an insulating one. During the positive voltage phase of the RF signal, the electrons from the discharge space are attracted to the target. They impact on the target and charge it; current flow to the RF generator is prevented by the capacitor. During' the negative half-wave of the RF signal, the electrons cannot leave the target due to the work function of the target material. Thus, the electron charge on the target remains constant.

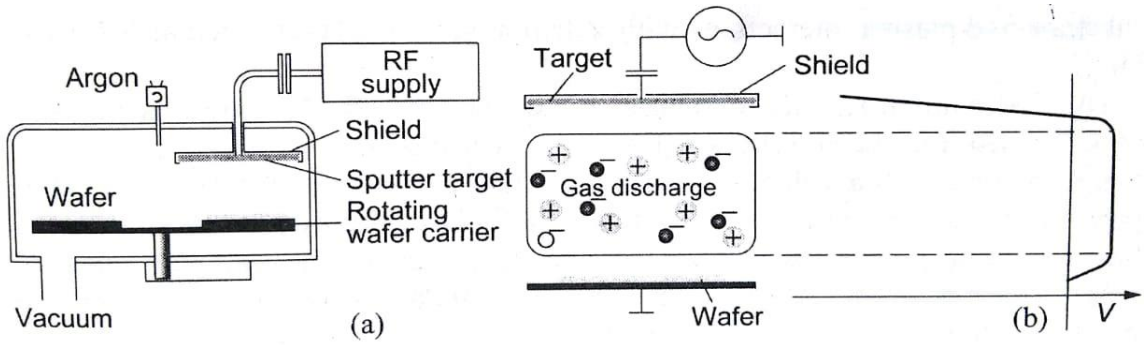


Fig: 2.6 (a) RF sputter System (b) Distribution of potential in an RF Plasma

Due to their mass, the positively charged ions are not capable of following the RF signal with frequencies above 50 kHz. Therefore, the ions are only subjected to the average electrical field which is caused by the electron charge accumulated on the target. Depending on the RF power at the target, the captured charge leads to a bias of 1 000 V or mere and causes an ion energy within the range of 1 keV.

When using a capacitively coupled target, the limitations of the glow discharge can be overcome, i.e., a conducting target is no longer required. Therefore, the number of layers which can be deposited by sputtering is greatly increased.

(III) ION PLATING.

This process is classed between resistance evaporation and glow discharge. A negative voltage is applied to the substrate, while the anode is connected with the source of the metal vaporization. The chamber is subsequently filled with Ar with a

pressure of a few Pa, and the plasma is ignited. After cleaning the wafer by sputtering, the e-gun is switched on and the material is vaporized. The growing of the layer on the substrate is improved by the plasma in some properties such as adhesion and homogeneity compared to a sole PVD.

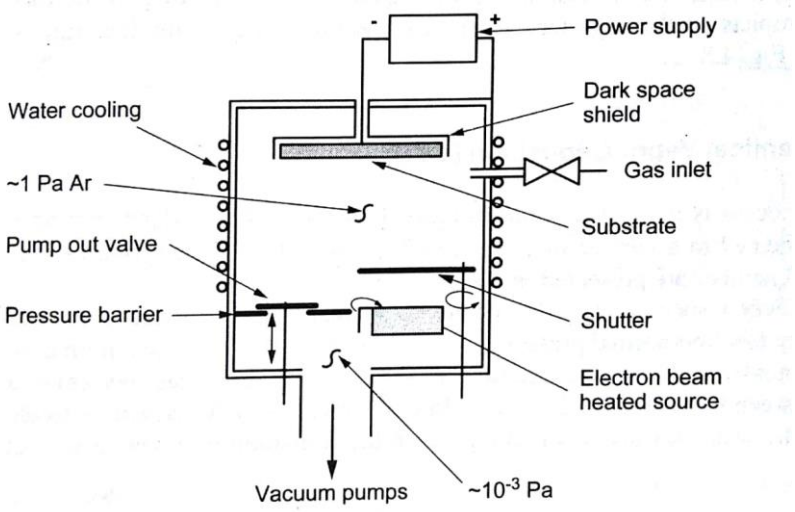


Fig 2.7: Ion Plating system

The advantages of ion plating are higher energies of the vaporized atoms and therefore better adhesion of the produced films. The disadvantage is heating of the substrate and plasma interactions with radiation-sensitive layers such as MOS oxides.

(IV) LASER ABLATION.

The following process data are typical values. A high-energy focused laser beam (100 m], l J / cm^2) is capable of eroding the surface of a target rotating with a velocity of one revolution per second. The material is vaporized on the substrate, and as a result, a film is produced on it at a rate of 0.07 nm/ laser pulse. The growth can be supported by heating the substrate (750 °C) and by chemical reactions (oxygen at 50 Pa). So far, the used lasers are excimer, Nd:YAG, ruby, and $C0_2$ lasers.

Advantages of laser ablation are the deposition of materials of high-melting points, a good control over impurities, the possibility of the vaporization in oxidizing environments, and stoichiometric vaporization. A shortcoming is the formation of

droplets on the vaporized layer. A system described in the literature is presented in Fig. 2.8.

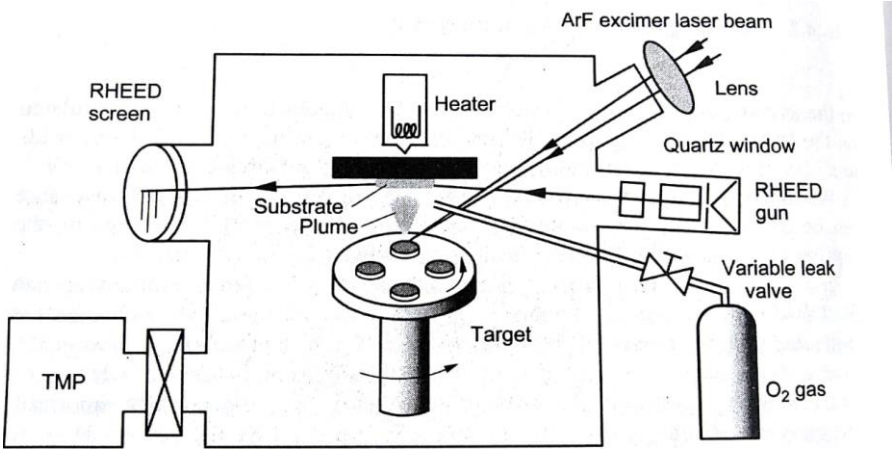


Fig 2.8 : Laser Ablation system under O_2 partial pressure

2.1.2 CHEMICAL VAPOR DEPOSITION (CVD)

The CVD process is performed in an evacuated chamber. The wafer is put on a carrier and heated to a temperature between 350 and 800 °C. Four possible versions of the chamber are presented in Fig. 2.9

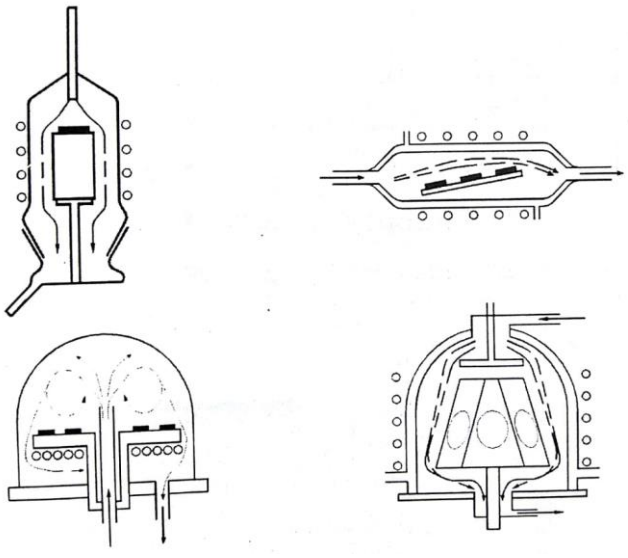


Fig 2.9 : Four versions of CVD chamber

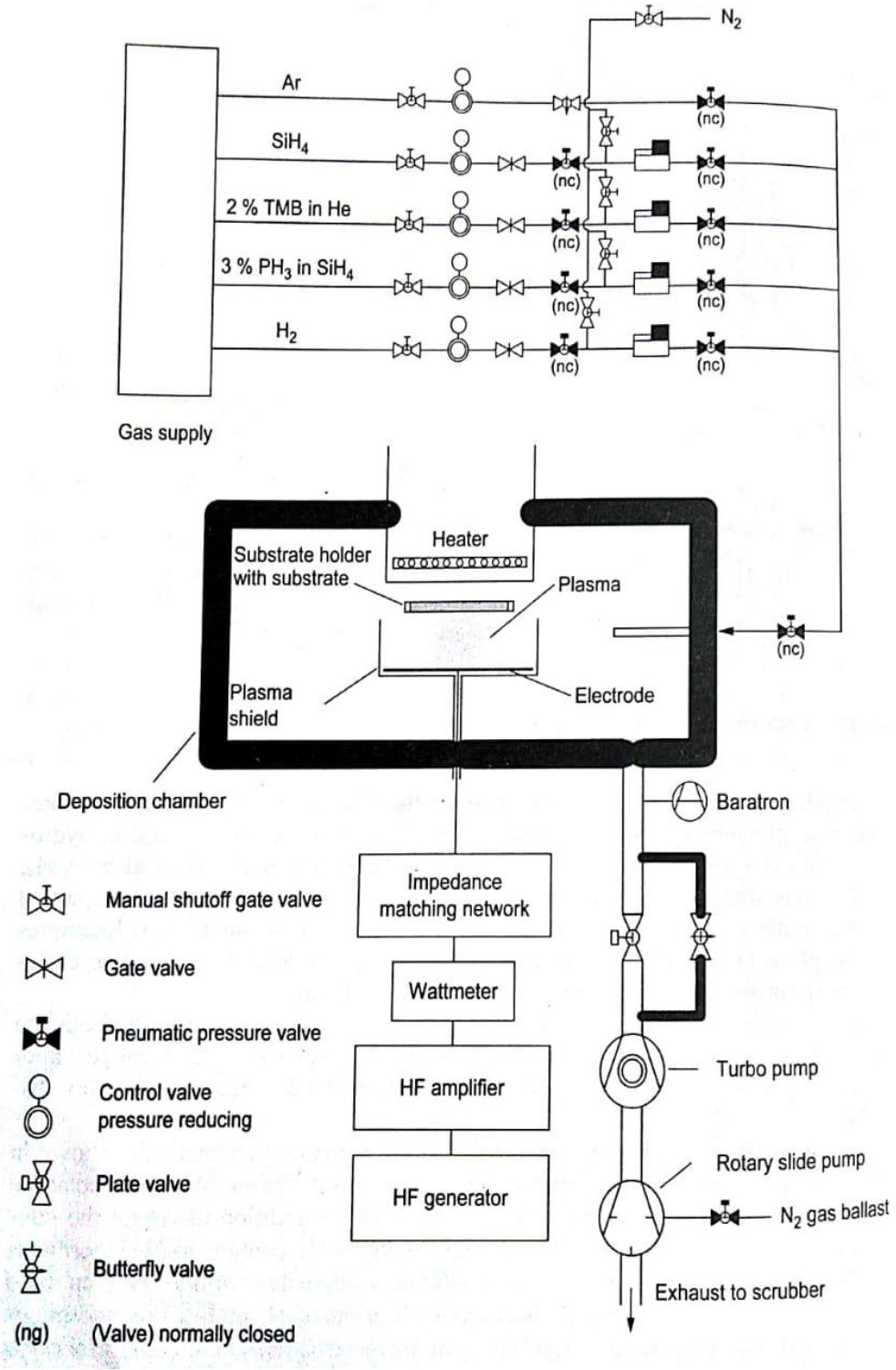


Fig 2.10: Block Diagram of PECVD

One or several species of gases are let in so that a gas pressure is formed between very low and normal pressure. The gas flow hits the wafer at a normal or a glancing incidence. Now a dissociation (in the case of a single gas species) or a reaction between two species takes place. In both cases, a newly formed molecule adheres to the wafer surface and participates in the formation of a new layer.

Let us consider silane (SiH4) as an example of the first case. On impact, it disintegrates into elementary silicon, which partly adheres to the surface, and to hydrogen, which is removed by the pumps. The second case is represented by SiH4, which reacts with N_20 to form SiO_2. The process can of course be accompanied by other types of gases which act as impurities in the deposited layer. Examples are phosphine (PH_3) or diborane (B_2H_6), which also disintegrate and deliver effective phosphorus or boron doping of the deposited silicon.

CVD is a layer structure without the continuation of the underlying lattice. The reverse case is called vapor phase epitaxy.

CVD deposition can be supported by an RF plasma, as schematically shown in Fig. 2.10, an example of an amorphous or micro-crystalline silicon deposition. The major difference to the conventional CVD is the addition of Ar for the ignition of the plasma and of H_2. The degree of the SiH_4 content in 11; determines whether amorphous or microcrystalline silicon is deposited. In the first step, both types are deposited. However, a high concentration of H_2 etches the amorphous portion, and only the microcrystalline component remains. The etching process is even more favoured if higher frequencies (e.g., 110 MHz) other than the usual 13.56 MHz are used. In Fig. 2.11, a typical PECVD system is depicted.

2.1.3 EPITAXY

We are dealing with epitaxy if a layer is deposited on a (crystalline) substrate in such a way that the layer is also monocrystalline. The layer is often referred to as film. In many cases, the film takes 99.9 % of the entire solid state, as in the example of a Czochralski crystal, which is pulled from a narrow seed nucleus. If film and substrate are from the same material, we are dealing with homoepitaxy (e.g., silicon-on-silicon), otherwise with heteroepitaxy (e.g., silicon-on-sapphire). Another distinction is made by the phase from which the film is made: vapor phase epitaxy, liquid phase epitaxy (LPE), and solid state epitaxy. A subclass of vapor phase epitaxy is molecular beam epitaxy (MBE).

The setup of a molecular beam epitaxy (MBE) is depicted in detail in Fig. 2.12.

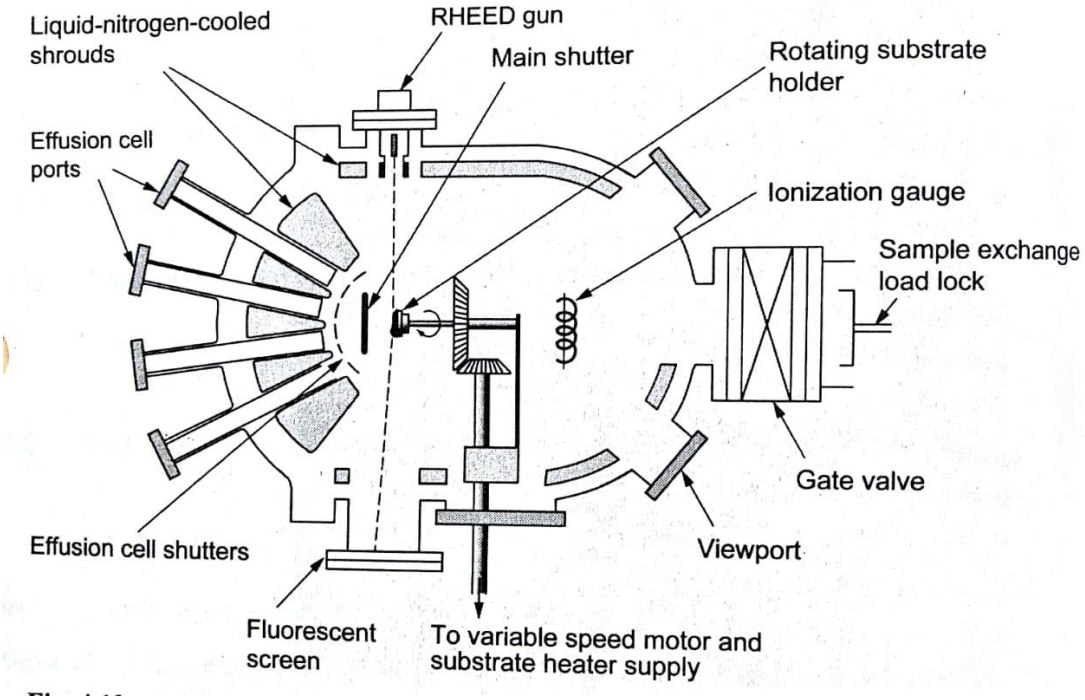


Fig 2.12 : Schematic Structure of MBE

The constituents of the deposited film are contained in mini furnaces as elements, the so-called Knudsen cells, which are discussed below. During heating some vapor pressure develops and an atom beam is emitted, which is bundled by successive apertures. The beam hits the wafer surface to which the atoms remain partially adhered. There, they can react with atoms of a second or third beam, which is also directed towards the wafer surface. A favourable reaction and finally the film deposition depend on the selection of the parameters, i.e., wafer temperature, the ratios of the beam densities, the purity of the surface, etc. As shown in the same figure, the effusion cell can be replaced by an evaporation with the electron gun. The chamber contains many devices for the in situ inspection of the growing layers, for example low energy electron diffraction (LEED), secondary ion mass spectroscopy (SIMS), and Auger and Raman spectroscopy. The quality of the vacuum is controlled by a residual gas analyzer. The effusion (Knudsen) cell is seen in detail in F ig. 2.13.

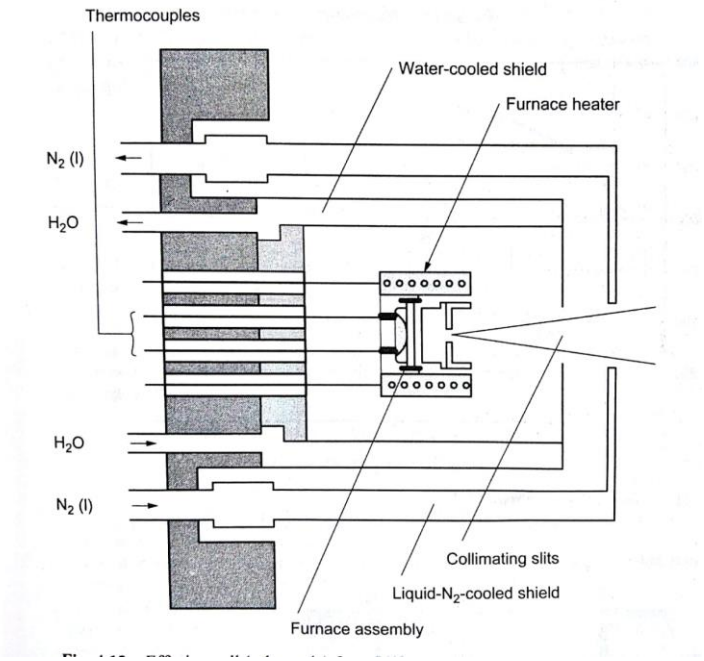


Fig 2.13 : Effusion Cell

The material to be deposited is contained in the innermost cell which is heated up. Its temperature is controlled by a set of thermocouples and resistance heaters. Without mrther measures, the high temperature leads to molecular desorption ' from all warmed up surfaces, to the emission of impurities into the substrate, and in the worst case, to the breakdown of the vacuum. Therefore, a screen cooled with liquid air is installed around the internal cell. Conversely, in order to avoid high thermal flows between furnace and screen, a water-cooled shield is inserted between them.

2.1.4 ION LMPIANTATION

Ion implantation is a doping technique with which ions are shot into a substrate (e.g., a silicon wafer) using an accelerator. The basic principle is presented in Fig. 2.16.

The desired ion species is let in as a gaseous compound through a needle valve (alternatively, solid state Sputtering sources are used). The compound is dissociated and ionized with an electron beam. The arising ions (including the unwanted ones) are pulled out of the source area and brought to an energy of 30 keV by a first, preliminary acceleration (all figures are typical values). Then, the ions pass through a magnetic field filter which is adjusted in such a way that only the desired ion type can run into the acceleration tube. The magnetic field filter is based on the fact that the Lorentz force for a moving charged particle compels a circular path. The radius of the circle depends on the magnetic field B, the velocity v, and the mass m of the particle. More exactly, it depends on *m/e* (e is the elementary charge). For the desired ion species, i.e., for a given m/e, the magnetic field is adjusted in such a way that the circular path of these particles

terminates exactly at the end of the accelerator tube. There, the ions acquire a total energy of 360 keV. This energy can be doubled or multiplied by the use of double or multiple charged ions. However, the ion yield, i.e., the available ion current, is exponentially reduced with the state of charge (ionization state).

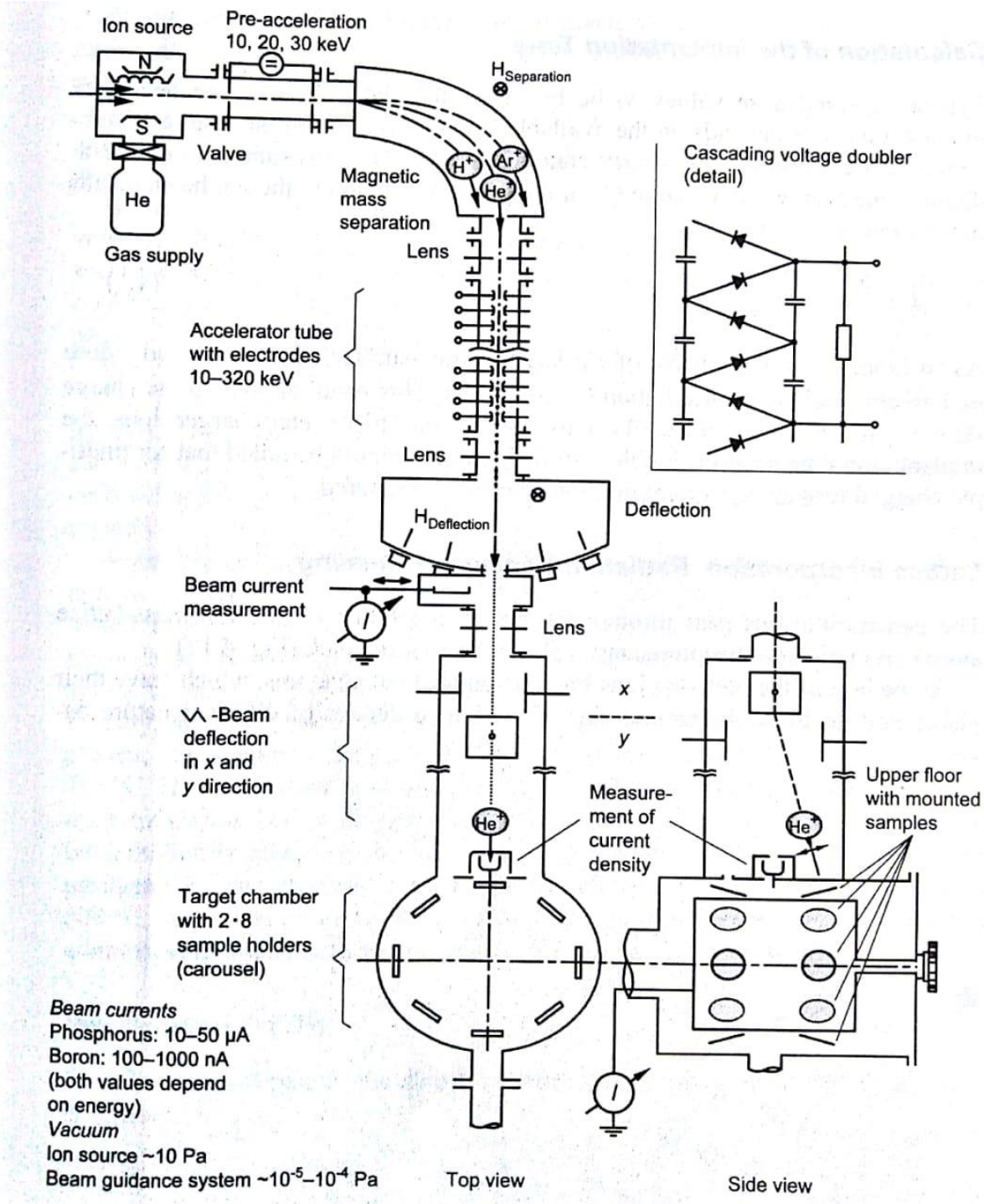


Fig 2.16: Ion implantation Equipment

The beam can be positioned by a combination of an aperture with a Faraday cage. With two capacitor disks each, the beam can be scanned upwards and downwards or left and right. In order to avoid Lissajou figures, horizontally and vertically incommensurable scanning frequencies are chosen. The beam current is measured by an ammeter, which is connected to the substrate holder isolated against ground. The {substrate holder is designed as a carousel, for example, in order to be able to implant several samples without an intermediate ventilation.

2.1.5 FORMATION OF SILICON OXIDE

Thin oxide layers are contained in almost all electronic devices. They appear as gate oxide in MOS transistors or MIS solar cells, field oxide for isolation purposes, antireflection layers in solar cells, or as passivation layers for long-term protection.

We begin with thermal oxidation. A Si wafer is cleaned so that any organic or heavy metal impurity on the surface is removed and the natural oxide is dissolved. Then, the wafer is inserted in a quartz tube heated to a temperature of about 1100 °C. A flow of an oxidizing gas, either pure oxygen (dry oxidation) or nitrogen driven through water (wet oxidation), is maintained. When oxygen penetrates into the substrate, the Si surface reacts with the oxygen and forms silicon dioxide. While this procedure sounds simple, it requires highest cleanliness, which is the critical step for the MOS production.Numerous scientific and technical investigations have focused on the properties of silicon dioxide. Fields of interest are the growth laws, deep levels, capture of charge carriers from the silicon, segregation and rearrangement of the dopant in the neighbouring silicon, masking properties against diffusion and ion implantation, etc. Their in-depth discussion is beyond the scope of this book:

Technical alternatives to thermal oxidation are CVD and PECVD of oxides. These are treated in the section on CVD. There are some technical CVD versions such as TEOS deposition shown in Fig. 2.17.

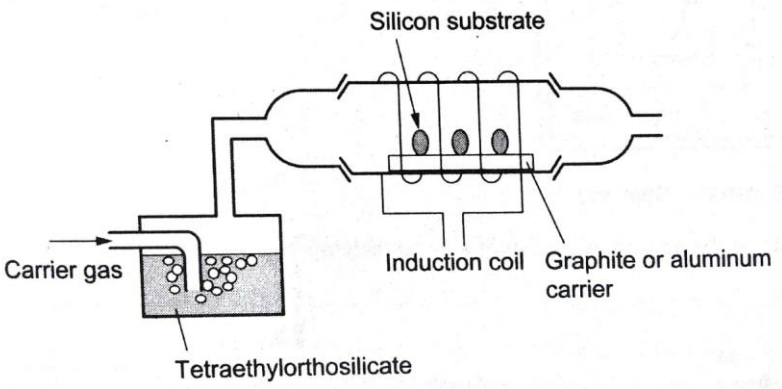


Fig 2.17: TEOS Process

A feed gas (usually nitrogen) is driven through a container filled with tetra ethylorthosilicate (TEOS). TEOS is a liquid at room temperature. Its chemical structure is presented in Fig. 2.18 The enriched nitrogen flows to the wafers where SiOz is deposited on their surfaces. The deposition is maintained at a temperature of about 650 to 850 °C by means of an external induction coil. It should be noted that the silicon in the TEOS is already oxidized, in contrast to the silane process.

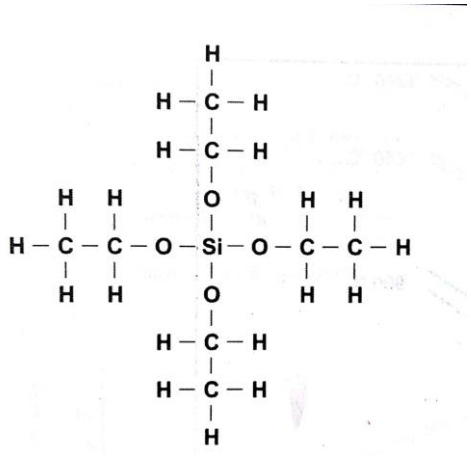


Fig 2.18: Structure of tetraethylorthosilicate (TEOS)

In this state, TEOS finds only limited application because the deposition temperature (>650 °C) prevents its use after metallization. In order to obtain lower deposition temperatures, the application of a more aggressive oxidant, i.e., ozone, is required. After adding some few molar per cent of ozone, the optimum deposition temperature is reduced to 400 °C.

The anodic oxidation is shown in Fig. 2.19. The wafer is immersed into a 0.04 M solution of KNO_3 in ethylene glycol with a small addition of water. Afier mounting it to a holder with a vacuum, it is positively charged, while a platinum disk acts as a backplate electrode.

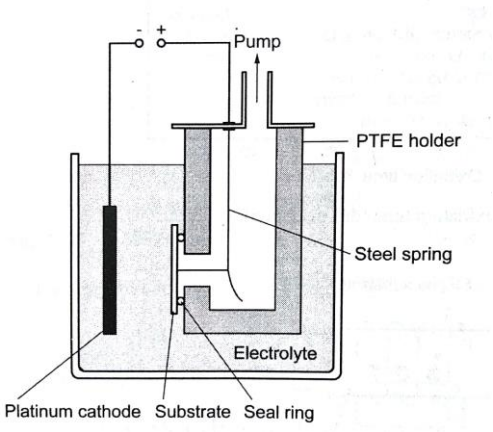


Fig 2.19: Assembly of the anodic oxidation of silicon

The current causes a reaction on the surface of silicon

$$Si + 2H_2O + 4h^+ \rightarrow SiO_2 + 4H^+$$

$$CH_2OH - CH_2OH \rightarrow 2HCHO + 2H^+ + 2e^-$$

$$H_2O \rightarrow \frac{1}{2}O_2 + 2H^+ + 2e^-$$

Technically, anodic oxidation is of little importance. The quality of the oxide is too low, and the process is time-consuming or not compatible with other applications.

Rarely, sol gels are used to manufacture oxide layers. During this process, a suspension of oxide particles in an organic solvent is distributed over the wafer. A centrifuge facilitates homogeneous distribution of the liquid on the wafer. Then, the wafer is baked and the solvent evaporates. The required temperatures range from 500 to 800 °C. The oxide can have the quality of a gate oxide, but no largescale application has been reported so far.

A further manufacturing process is SOI (silicon-on-oxide), i.e., the implantation of oxygen up to the stoichiometric dose and its reaction with silicon to form SiO_2. SOI has gained in importance because of better radiation resistance and heat distribution compared to conventional technologies.

2.2 FABRICATION OF NANOPARTICLES

Nanoparticle is defined as a ball or a ball-like molecule which consists of a few 10 to some 10,000 atoms interconnected by interatomic forces but with little or no relationship to a solid state. A first example is a nano-crystalline Si particle which is embedded into an amorphous matrix.

2.1.1 GRINDING WITH IRON BALLS

First of all, a container is filled with stainless steel balls of a few millimeters m diameter. The material to be crushed is added in the form of a powder of about 50 µm diameter grain size. After filling the container with liquid nitrogen, a rotating shaft grinds the material. The grinding periods are within the range of minutes to some 100 hours. This process is simple; its weakness, however, lies in the fact that the grinding balls contribute to impurities.

2.1.2 GAS CONDENSATION

A typical system is shown in Fig. 2.20. The operation occurs in an evacuated chamber with a pressure of 10-5 Pa. After mounting the raw material on one or more crucibles, it is evaporated thermally, by an electron gun, or by ion sputtering.

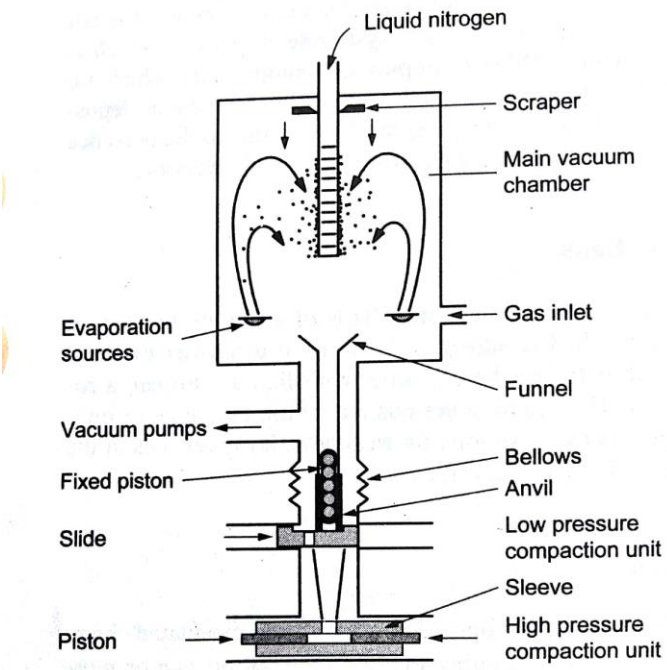


Fig 220: System for inert gas condensation of nanocrystalline power

The evaporated atoms or molecules unite and form particles of different sizes. Finally they are captured with a cold linger from which they are scraped off and collected with a funnel. The particle diameter is usually within the range of 5 to 15 nm.

2.2.3 LASER ABLATION

The raw material is provided as a solid. Its dissolution is achieved by a focused laser beam-similar to the cutting of a metal or a semiconductor. The advantage of this procedure is a 1:1 transfer of the material composition from the raw material to the particles. The system has already been shown in Fig. 2.8.

2.2.4 REDUCTION METHODS

Some metal compounds (e.g., chlorides) can be reduced to elementary metallic nanopowder by the application of $NaBEt_3H$, $LiBEt_3H$, and $NaBH_4$, for example. The reaction equation can be written as (M: metal, Et: ethyl):

$$MCl_x + xNaBEt_3H \rightarrow xNaCl + xBEt_3 + \frac{1}{2}xH_2$$

2.2.5 SOL GELS

A sol (hydrosol) is a colloidal dispersion in liquid. A gel is a jelly-like substance formed by coagulation of a sol into a gel.

The best known example of a so] gel process is probably the production of SiO_2. A catalyst (acid or base) is added to a solution of tetramethoxysilane (TMOS), water, and methanol. Hydrolysis of the Si-OMe (Me: methyl) bonds leads to the formation of Si-OH groups

$$Si(OMe)_4 + 4H_2O \rightarrow Si(OH)_4 + 4MeOH$$

Further dehydration reduces the "$Si(OH)_4$" to SiO_2; gel. If hydrolysis and condensation are completed,a silicon oxide xerogel is formed (in Greek, xerox means dry). During the reaction, the gel reaches a viscosity so low that it can be applied onto a centrifuge and distributed over the wafer. When annealing over 800 °C, homogeneous oxides, comparable to bad MOS gates, can be manufactured. lf doped SiO_2 layers are produced, they can be used as diffusion sources in a subsequent

At moderate and low solidification temperatures, the procedure delivers the so called nanocomposite. By definition, nanocomposites contain nanoparticles of less than 1000 nm in a host matrix. The following nanocomposites have already been manufactured: nano-Co/Mo, Cu, Fe, Ni, Pd, Pt and Ru in $A1_20_3$, SiO_2, TiO_2 and ZrO_2 gels, nano-C, Cu/Ni, Plei, and Pt in silica gel, nano-Ag, Ge, Os, C, Fe, Mo, Pd, Pt, Re, Ru and PtSn in silica gel-xerogel.

2.2.6 PRECIPITATION OF QUANTUM DOTS

Quantum dots are three-dimensional semiconductor materials in or on a matrix. Nano composites from semiconductor materials and the above-mentioned self assembled islands belong to this group. Sometimes it is difficult to differentiate between quantum dots and nano defects. An example is SiO_2 implanted with Ge, which is used for photoluminescence experiments. The earliest descriptions of quantum dots took place with the investigation of semiconductor precipitation in glasses. Precipitation is still used in the manufacturing of CdS, CdSe, CdTe, GaAs, and Si nanocrystallites in silica glasses. The contaminants are added to the melt, and after a further annealing step from 600 to 1400 °C, they form precipitates of controllable size, for instance, 2 nm for CdTe dots in boron silicate glass.

Another procedure is codeposition of quantum dots with thin films. There are several modifications, but the common principle is the production of nanocrystallites in a separate step

(by evaporating, laser ablation, sputtering a target, etc.). They are directed towards a substrate which is, however, covered at the same time with a film so that they are included into this film.

Quantum dots can also be manufactured by means of lithography. Since a high resolution is required, electron beam lithography must be employed. The procedure takes place mostly in such a way that the material which is to be converted into quantum dots is deposited on a substrate like GaAs via MBE. The size of the electron beam spot determines the smallest possible size of the quantum dot. Therefore, if the wafer is etched, islands of this size remain. They are further reduced by etching so that quantum dots of a few 10 nm can be produced. Very often, the layers are covered in order to improve the results.

MODULE III
CHARACTERIZATION OF NANOSTRUCTURES

3.1 INTRODUCTION

Optical microscopes are generally used for observing micron level materials with reasonable resolution. Further magnification cannot be achieved through optical microscopes due to aberrations and limit in wavelength of light. In the field of characterization electron microscopy (EM) takes a vital role. Electron microscopes uses electrons instead of photons, because electron have a much shorter wavelength than photons and allow observe matter with atomic resolution.

There are two general types of electron microscopes:

- ***The scanning electron microscope (SEM)*** that scans an electron beam over the surface of an object and measures how many electrons are scattered back, and the transmission electron microscopy (TEM) that shoots electrons through the sample and it measures the electron beam changes. Because of electron scattering in the sample, it works in a vacuum environment and makes the instrument considerably larger and expensive. All electron microscopes work at low pressure and usually in high vacuum chamber to avoid scattering electron in the gas. In environmental electron microscopes, differential pumping systems are used to actually have gases present by the sample together with the electron beam.
- ***Scanning probe microscope (SPM)*** covers the methods where a sharp tip is scanned over a surface in a raster pattern and the interaction with the surface is recorded in each pixel to form an image of the interaction. There are a multitude of methods and interactions in SPM. Broadly speaking, there are three main categories.
  - In ***Scanning Tunneling Microscopy*** (STM), one uses an atomically sharp matellic tip and records the minute tunneling current between the tip and the surface, when the tip is hovering so close to the surface that electrons can move between the surface and the tip.

- In ***Atomic Force Microscopy*** (AFM), a cantilever with a sharp tip-somewhat like the needle of an old record player is scanned over the surface and the topography or surface softness can be recorded.
- In ***Scanning Near- Field Optical Microscopy*** (SNOM) a probe with a small aperture is scanned over the surface collecting the light coming from regions much smaller than the wavelength of the light used.

3.2 SCANNING TUNNELING MICROSCOPY (STM)

The scaning tunneling microscopy (STM) is a type of electron microscope that shows three-dimensional images of a sample. In the STM, the structure of a surface is studied using a stylus that scans the surface at a fixed distance from it. Fig.3.1 shows Schematic view of an STM.

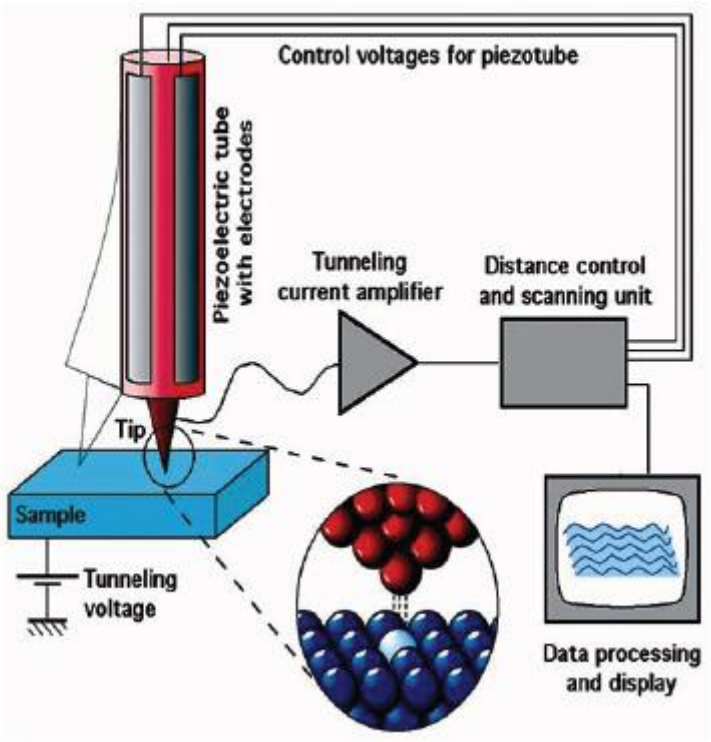


Fig.3.1 Schematic view of a STM

Scanning tunneling microscopy is an instrument for producing surface images with atomic scale lateral resolution, in which a fine probe tip is scanned over the surface of a conducting specimen with the help of a piezoelectric crystal at a distance of 0.5–1 nm, and the resulting tunneling current or the position of the tip required to maintain a constant tunneling current is monitoredThe principle of STM is based on the concept of quantum tunneling. When a conducting tip is brought very near to a metallic or semiconducting surface, a bias between the

two can allow electrons to tunnel through the vacuum between them. For low voltages this tunneling current is a function of the local density of states at the Fermi level of the sample. Variations in current as the probe passes over the surface are translated into an image.

For STM, good resolution is considered to be 0.1 nm lateral resolution and 0.01 nm depth resolution. They normally generate images by holding the current between the tip of the electrode and the specimen at some constant value by using a piezoelectric crystal to adjust the distance between the tip and the specimen surface, while the tip is piezoelectrically scanned in a raster pattern over the region of specimen surface being imaged by holding the force, rather than the electric current, between tip and specimen at a set point value. When the height of the tip is plotted as a function of its lateral position over the specimen, an image that looks very much like the surface topography results. STM can be operated in two modes. In constant current imaging, a feedback mechanism is enabled and a constant current is maintained while a constant bias is applied between the sample and tip. As the tip scans over the sample, the vertical position of the tip is altered to maintain the constant separation. An alternating imaging mode is the constant height operation in which constant height and bias are simultaneously maintained.

A variation in current results as the tip scans the sample surface because a topographic structure varies in the tip sample separation. The constant current mode produces a constant directly related to electron charge density profiles, whereas the constant height mode permits faster scan rates. The STM can be used not only in ultra high vacuum but also in air and various other liquid or gas, at ambient and wide range of temperatures. STM can be a challenging technique, as it requires extremely clean surfaces and sharp tips.

The Scanning tunneling microscopy (STM) is capable of acquiring remarkable images on the most extreme scale, easily resolving atomic structure in the right environments. It is apowerful tool in nanotechnology and nanoscience providing facilities for characterization and modification of a variety of materials. STM is successfully used to detect and characterize materials like carbon nanotubes (single-walled and multi-walled) and graphine layer . The instrument has also been used to move single nanotubes or metal atoms and molecules on smooth surfaces with high precision.

3.3 ATOMIC FORCE MICROSCOPE

The atomic force microscope (AFM) is a very powerful microscope invented by Binnig, Quate and Gerber in 1986. Besides imaging it is also one of the foremost tools for the manipulation of matter at the nanoscale. Fig.3.2 shows schematic diagram of AFM

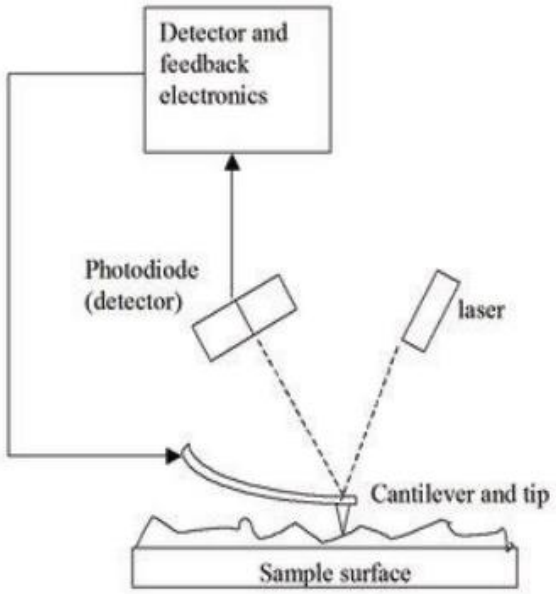


Fig.3.2 Schematic diagram of AFM

The AFM consists of a cantilever with a sharp tip at its end. The tip is brought into close proximity of a sample surface. The AFM measures the forces acting between a fine tip and a sample. The tip is attached to the free end of a cantilever and is brought very close to a surface. Attractive or repulsive forces resulting from interactions between the tip and the surface will cause a positive or negative bending of the cantilever. The force between the tip and the sample leads to a deflection of the cantilever according to Hooke's law.

Typically, the deflection is measured using a laser spot reflected from the top of the cantilever. In this procedure, a flexible force sensing cantilever stylus is scanned in a raster pattern over the surface of the sample. During the scan, it provides the topographic information.

Atomic force microscopy is currently applied to various environments (air, liquid, vacuum) and types of materials such as metal semiconductors, soft biological samples and conductive and non-conductive materials. With this technique size measurements or even manipulations of nano-objects may be performed. AFM provides a number of advantages over conventional microscopy techniques. AFMs probe the sample and make measurements in three dimensions, *x,*

y, and *z* (normal to the sample surface), thus enabling the presentation of three-dimensional images of a sample surface. This provides a great advantage over any microscope available previously. With good samples (clean, with no excessively large surface features), resolution in the *x-y* plane ranges from 0.1 to 1.0 nm and in the *z* direction is 0.01 nm (atomic resolution).

AFMs require neither a vacuum environment nor any special sample preparation, and they can be used in either an ambient or liquid environment.

In atomic force microscopy a probe consisting of a sharp tip (nominal tip radius is in the order of 10 nm) located near the end of a cantilever beam is raster scanned across the surface of a specimen using piezoelectric scanners. Changes in the tip specimen interaction are often monitored using an optical lever detection system, in which a laser is reflected off of the cantilever and onto a position sensitive photodiode. During scanning, a particular operating parameter is maintained at a constant level, and images are generated through a feedback loop between the optical detection system and the piezoelectric scanners.
There are three scan modes for AFM, depending on the application.
The imaging modes are namely ***contact mode, non contact mode and tapping mode***.

❖ *contact mode*

Fig.3.3. shows schematic diagram for operation of AFM in the contact mode . Contact mode AFM is one of the more widely used scanning probe modes, and operates by rastering a sharp tip, made either of silicon or Si_3N_4 attached to a low spring constant cantilever across the sample.

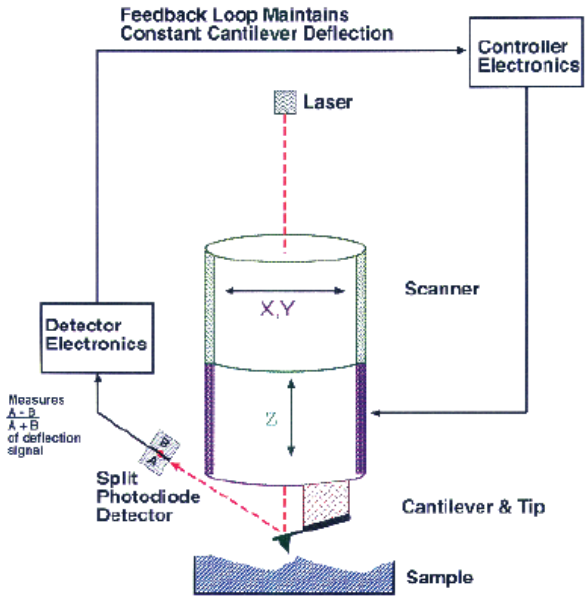


Fig.3.3 Schematic diagram for operation of AFM in the contact mode

An extremely low force approximately 10-9 N, is maintained on the cantilever, thereby pushing the tip against the sample as it raster's. Either the repulsive force between the tip and sample or

the actual tip deflection is recorded relative to spatial variation and then converted into an analogue image of the sample surface.
The AFM tip is first brought close to the sample surface, and then the scanner makes a final adjustment in tip sample distance based on a set point determined by the user.
The tip, now in contact with the sample surface through any adsorbed gas layer, is then scanned across the sample under the action of a piezoelectric actuator, either by moving the sample or the tip relative to the other. A laser beam aimed at the back of the cantilever tip assembly reflects off the cantilever surface to a split photodiode, which detects the small cantilever deflections. By maintaining a constanttip sample separation and using Hooke's Law, the force between the tip and the sample iscalculated. Finally, the distance the scanner moves in the *z* direction is stored in thecomputer relative to spatial variation in the *x-y* plane to generate the topographic image of the sample surface.

❖ In the ***non contact mode***, the cantilever is externally oscillated close to its resonance frequency. The oscillation gets modified by the tip-sample interaction forces; these changes in oscillation with respect to the external reference oscillation provide information about the sample's characteristics. The dynamic mode generates lower lateral forces on the sample and is widely used to image biological sample. Non contact mode operation includes frequency modulation and the more common amplitude modulation. In frequency modulation, changes in the frequency of modulation provide information about a sample's characteristics. In amplitude modulation, changes in the oscillation amplitude yield topographic information about the sample. Additionally, changes in the phase of oscillation under tapping mode can be used to discriminate between different types of materials on the surface.

In ***tapping mode*** the cantilever tip is stimulated to vibrations near the resonant frequency (often hundreds of kilohertz) using a piezoelectric crystal. The piezo motion causes the cantilever to oscillate when the tip is not in contact with the surface of a material. The oscillating tip is then moved towards the surface until it begins to tap the surface. During scanning, the vertically oscillating tip alternately contacts the surface and lifts off, generally at a frequency of 50,000–500,000 cycles/s. As the oscillating cantilever begins to intermittently contact the surface, the cantilever oscillation is reduced due to energy loss caused by the tip contacting the surface. The reduction in oscillation amplitude is used to measure the surface characteristics. The tapping mode is less destructive than the contact mode, because the exerted forces are in the pico Newton (pN) range. Tapping mode AFM can be performed on both wet and dry sample surfaces.

Atomic force microscopy mainly used in fundamental and practical research and has important role in the development of nanotechnology. AFM can be used to explore the nanostructures, properties, and surfaces and interfaces of fibers and fabrics. For example, structural characteristics of nanofibre materials, nanolevel surface modification of textile surfaces by plasma can be easily accessed by this sophisticated technique.
The AFM has several advantages over the electron microscope. Unlike the electron microscope which provides a two-dimensional projection or a two-dimensional image of a sample, the AFM provides a true three-dimensional surface profile. Additionally, samples viewed by an AFM do not require any special treatment that would actually destroy the sample and prevent its reuse. While an electron microscope needs an expensive vacuum environment for proper operation, the AFM can work perfectly well in an ambient or even liquid environment. The main disadvantage of AFM as compared to the electron microscope is the image size. The electron microscope can show an area on the order of millimeters by millimeters and a depth of field on

the order of millimeters. The AFM can only show a maximum height on the order of micrometres and a maximum area of around 100 by 100 micrometers.

3.4 SCANNING ELECTRON MICROSCOPY (SEM)

Scanning electron microscopy (SEM) is used to inspect the topographies of specimens at very high magnifications. The SEM is a microscope that uses electrons instead of light to form an image.

The scanning electron microscope (SEM) uses a focused beam of high energy electrons to generate a variety of signals at the surface of solid specimens. The signals that derive from electron sample interactions reveal information about the sample including external morphology (texture), chemical composition, crystalline structure and orientation of materials making up the sample.

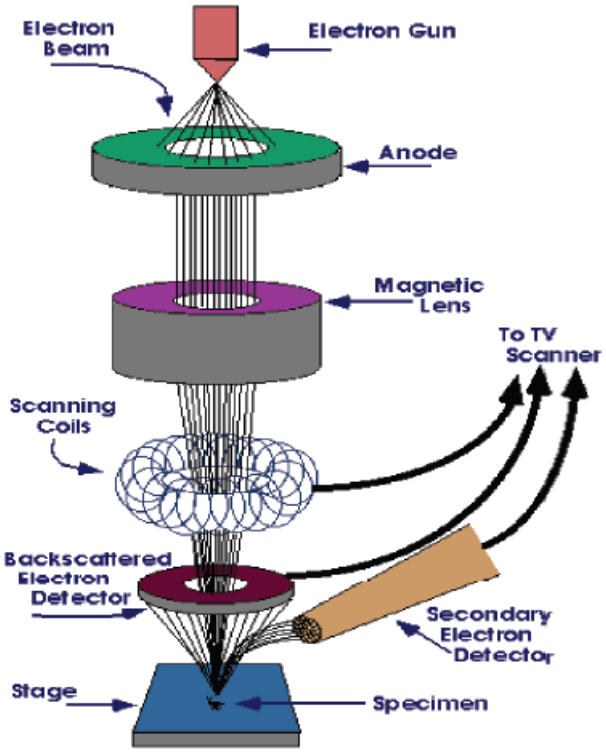


Fig.3.4 Schematic diagram of typical scanning electron microscope

In scanning electron microscope a beam of electrons is produced at the top of the microscope by an electron gun. The electron beam follows a vertical path through the microscope, which is held within a vacuum. The beam travels through electromagnetic fields and lenses, which focus the beam down toward the sample. When the primary electron beam strikes the surface of the sample it is scattered by atoms in the sample. Through this scattering event, the primary beam spreads effectively and fills a teardrop shaped volume, known as the interaction volume, and this extends about 100 nm to 5 μm into the surface. The size of interaction volume depends on the beam accelerating voltage, the atomic number of the specimen and the specimen's density.

Secondary electrons are emitted from interactions in this region, they are then detected, converted to a voltage and amplified to produce an image.

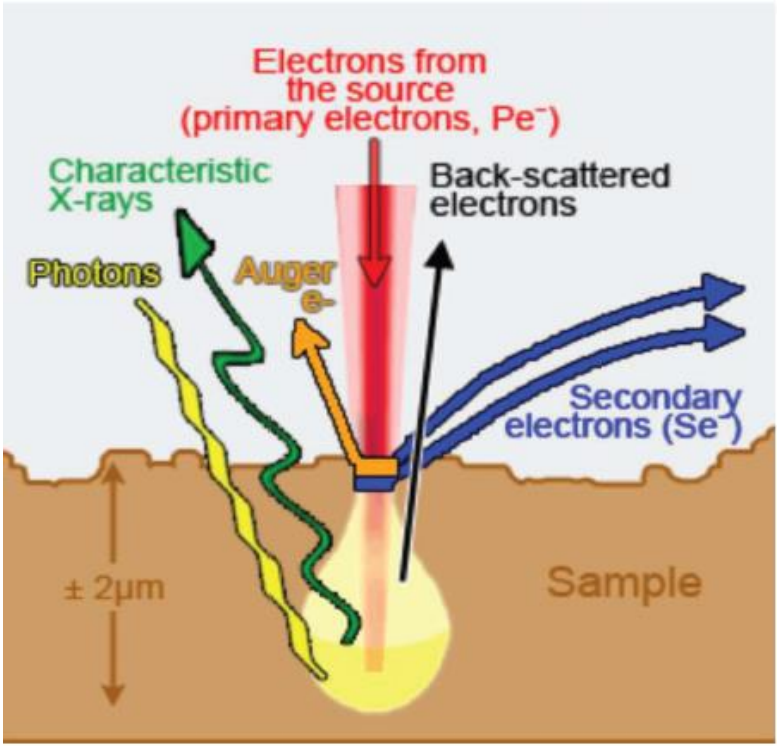


Fig.3.5 Electrons produced in SEM

In SEM, more than one type of signal can be produced and they include secondary electrons, back scattered electrons (BSEs), characteristic X-rays, specimen current and transmitted electrons. Specialized detectors are required for the detection of these signals and they are not usually all present on a single machine. Fig.3.5. shows electrons produced in SEM. The SEM can produce very high resolution images of a sample surface, revealing details about 1-5 nm in size in its primary detection mode i.e. secondary electron imaging. An additional electron interaction in the SEM is that the primary electron collide with an ejected electron from an atom will decay to its ground state by emitting either a characteristics X-ray photon or Auger electron. This characteristics X-ray or Auger electrons have been used for chemical characterization i.e to identify the elemental composition of the sample by a technique known as energy dispersive X-ray .

Back scattered electron images can be used for rapid discrimination of phases in multiphase samples. SEM required with diffracted backscattered electron detectors can be used to examine micro fabric and crystallographic orientation in many materials . SEM requires that the

specimens should be conductive for the electron beam to scan the surface and that the electrons have a path to ground for conventional imaging. Non conductive solid specimens are generally coated with a layer of conductive material by low vacuum sputter coating or high vacuum evaporation. This is done to prevent the accumulation of static electric charge on the specimen during electron irradiation. Non conducting specimens may also be imaged uncoated using specialized SEM instrumentation such as the "Environmental SEM" (ESEM) or in field emission gun (FEG) SEM operated at low voltage, high vacuum or at low vacuum, high voltage.

The scanning electron microscope has many advantages over traditional microscopes. The SEM has a large depth of field that allows more of a specimen to be in focus at one time. The SEM also has much higher resolution, so closely spaced specimens can be magnified at much higher levels. As the SEM uses electromagnets rather than lenses, the researcher has much more control in the degree of magnification. All of these advantages, as well as the actual, strikingly clear images, make the scanning electron microscope one of the most useful instruments in research today. The SEM shows very detailed three dimensional images at much high magnifications (up to × 300000) as compared to light microscope (up to × 10000) and can also be used to view dispersion of nanoparticles with a seemingly three dimensional structure. But as the images are created without light waves, they are black and white.

3.5 TRANSMISSION ELECTRON MICROSCOPY (TEM)

The transmission electron microscope (TEM) operates on the same basic principles as the light microscope but uses electrons instead of light. TEMs use electrons as “light source” and their much lower wavelength make it possible to get a resolution a thousand times better than that with a light microscope. With the help of TEM objects can be seen in the order of a few angstroms (10^{-10} m). It can be used to study small details in the cell or different materials down to near atomic levels.

At smaller magnifications TEM image contrast is due to absorption of electrons in the material, due to the thickness and composition of the material. At higher magnifications complex wave interactions modulate the intensity of the image, requiring expert analysis of observed images. Alternate modes of use allow TEM to observe modulations in chemical identity, crystal orientation, electronic structure and sample induced electron phase shift as well as the regular absorption based imaging.

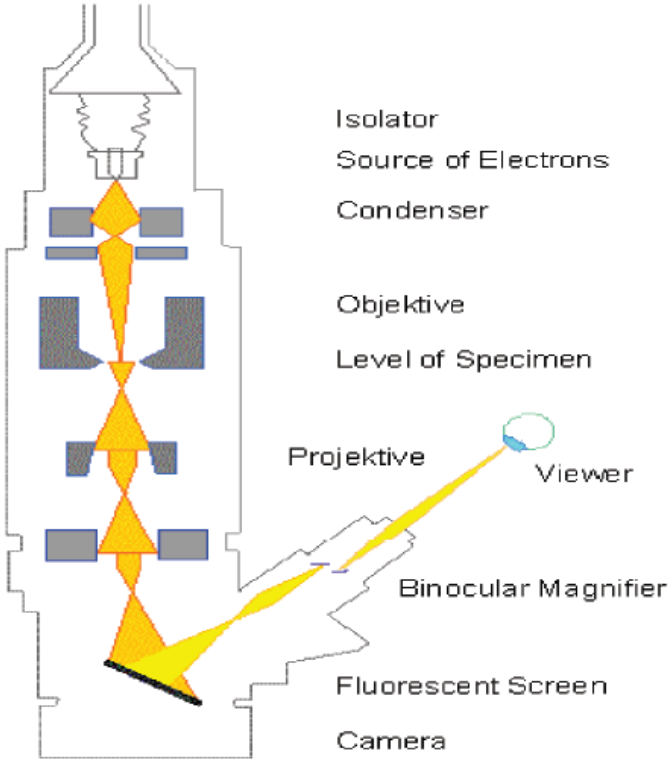


Fig.3.6 Schematic diagram of TEM

Transmission electron microscopy is a microscopy technique whereby a beam of electrons is transmitted through a very thin sample and interacting with the sample as it passes through. The interaction of the electrons are transmitted, magnified and then focused onto a fluorescent screen or it can be detected by a CCD camera. Fig.3.6. shows a schematic diagram of a TEM. The operation of the TEM requires an ultra high vacuum and a high voltage. The electrons are emitted by a source and then it is focused and magnified by a set of magnetic lenses. The image that is formed is either shown on a fluorescent screen or on a monitor and it is printed on photographic film.

In TEM, the crystalline sample interacts with the electron beam mostly by diffraction rather than by absorption. The intensity of the diffraction depends on the orientation of the planes of atoms in a crystal relative to the electron beam; at certain angles the electron beam is diffracted strongly from the axis of the incoming beam, while at other angles the beam is largely transmitted. TEM exploits three different interactions of electron beam specimen like unscattered electrons (transmitted beam), elastically scattered electrons (diffracted beam) and inelastically scattered electrons.

Different types of images are obtained in TEM with the different types of electrons using the apertures properly. As a result, diffraction patterns are shown because of the scattered electrons. If the unscattered beam is selected, we obtain the bright field image or light field images. It is also possible to produce an image from electrons deflected by a particular crystal plane which is known as a dark field image.

TEM forms a major analysis method in a range of scientific fields, in both physical and biological sciences. TEMs find application in cancer research, virology, materials science as well as pollution, nanotechnology and semiconductor research. TEM images reveal the distribution and dispersion of nanoparticles in polymer matrices of nanocomposite fibers, nanocoatings and etc. The extent of exfoliation, intercalation and orientation of nanoparticles can also be visualized using the TEM micrograph. Today's transmission electron microscopes offer resolutions up to 0.1nm at 300kV and probe diameters up to 0.34nm.

3.6 X - RAY DIFFRACTION (XRD)

X - ray diffraction technique is the most common and efficient method for the determination of structure and crystallainity and material identification. XRD is an apt method to examine whether a resultant material has amorphous or crystalline nature. Crystalline phases can be identified by just comparing the interplanar distance 'd' values obtained from XRD data with the fundamental data in Joint Committee on Powder Diffraction Standards (JCPDS).

X - ray diffraction is based on constructive interference of monochromatic X - rays from a crystalline sample. The X - rays, generated by a cathode ray tube are filtered to produce monochromatic radiation, collimated and directed towards the sample. X - ray primarily interact with electrons in atoms, collide and some photons from the incident beam are deflected away from original. The X - rays interfere constructively and destructively producing a diffraction pattern on the detector. The incident X - ray radiation produces a Bragg peak if their reflections from the various planes interfered constructively. The interference is constructive, when the phase shift is a multiple of 2π, this condition can be expressed by Bragg's law.

$$n\lambda = 2d\ sin\theta$$

where n is an integer, λ is the wavelength of incident wave, d is the spacing between the planes in the atomic lattice and θ is the angle between the incident ray and the scattering planes. Schematic diagram of X - ray diffraction is shown in Figure 3.7

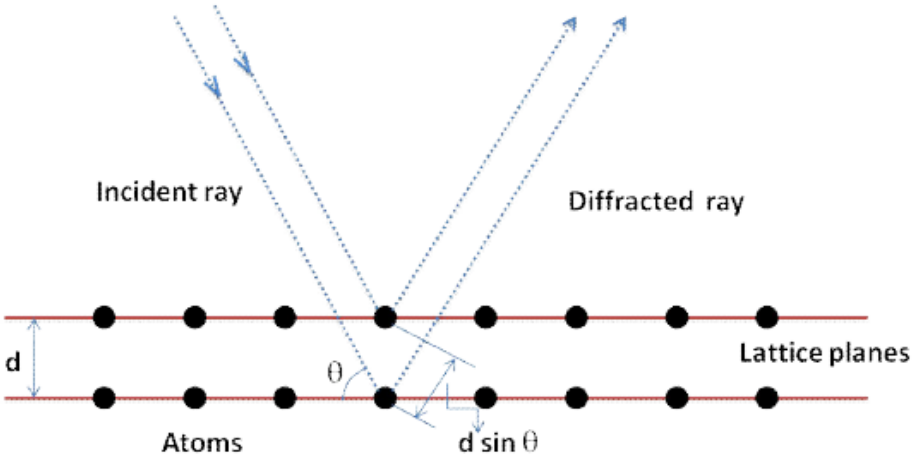


Fig 3.7 Bragg's law

A typical powder X - ray diffractometer consists of a source of radiation, a monochromator to choose the wavelength, slits to adjust the shape of the beam, a sample and a detector. A goniometer is used for fine adjustment of the sample and the detector positions. The goniometer mechanism supports the sample and detector, allowing precise movement. The source X - rays contains several components; the most common being Kα and Kβ. The specific wavelengths are characteristic of the target material (Cu, Fe, Mo, Cr). Monochromators and filters are used to absorb the unwanted emission with wavelength Kβ, while allowing the desired wavelength, Kα to pass through. The X - ray radiation most commonly used is that emitted by copper, whose characteristic wavelength for the Kα radiation is equal to 1.5418 Å. The filtered X - rays are collimated and directed onto the sample as shown in the Figure 3.8 .

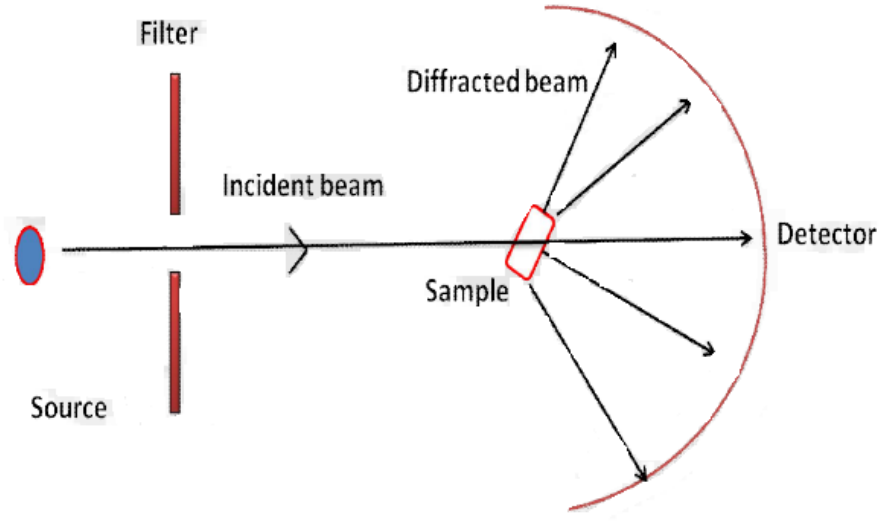


Figure 3.8 Schematic diagram of the diffractometer

When the incident beam strikes a powder sample, diffraction occurs in every possible orientation of 2θ. The diffracted beam may be detected by using a moveable detector such as a Geiger counter, which is connected to a chart recorder.

The counter is set to scan over a range of 2 θ values at a constant angular velocity. Routinely, a 2 θ range of 5 to 70 degrees is sufficient to cover the most useful part of the powder pattern. The scanning speed of the counter is usually 2 θ of 2^{o} min^{-1}. A detector records and processes this X - ray signal and converts the signal to a count rate which is then fed to a device such as a printer or computer monitor. The sample must be ground to fine powder before loading it in the glass sample holder. Sample should completely occupy the square glass well.

3.7 UV - VISIBLE SPECTROSCOPY (UV - Vis)

This refers to absorption spectroscopy in the ultra - violet and visible spectral region. In this region of the electromagnetic spectrum, molecules undergo electronic transition. When sample molecules are exposed to light having an energy ($E = hv$ where *'E'* is energy in joules, 'h' is Planck's constant 6.62×10^{-34} J s and 'v' is frequency in Hertz), that matches a possible electronic transition within the molecule, some of the light energy will be absorbed as the electron is promoted to a higher energy orbital. An optical spectrometer records the wavelengths at which absorption occurs, together with the degree of absorption at each wavelength. The resulting spectrum is presented as a graph of absorbance (A) versus wavelength (λ). The optical properties of materials can be studied with the help of UV - Vis spectra.

Principle

The absorbance of light by molecules in the solution is based on the Beer - Lambert law,

$$A = log\frac{I_o}{I} = \varepsilon \times b \times c$$

where, *I0* is the intensity of the reference beam and *I* is the intensity of the sample beam, ε is the molar absorbtivity with units of L mol^{-1} cm^{-1}, b = path length of the sample in centimeters and c = concentration given solution expressed in mol L^{-1}

The main components of the UV - Vis spectrometers are a light source, double beams (reference and sample beam), a monochromator, a detector and a recording device. The source is usually a tungsten filament lamp for visible and deuterium discharge lamp for UV measurements. The light coming out of the source is split into two beams - the reference and the sample beam as shown in the Figure 3.9.

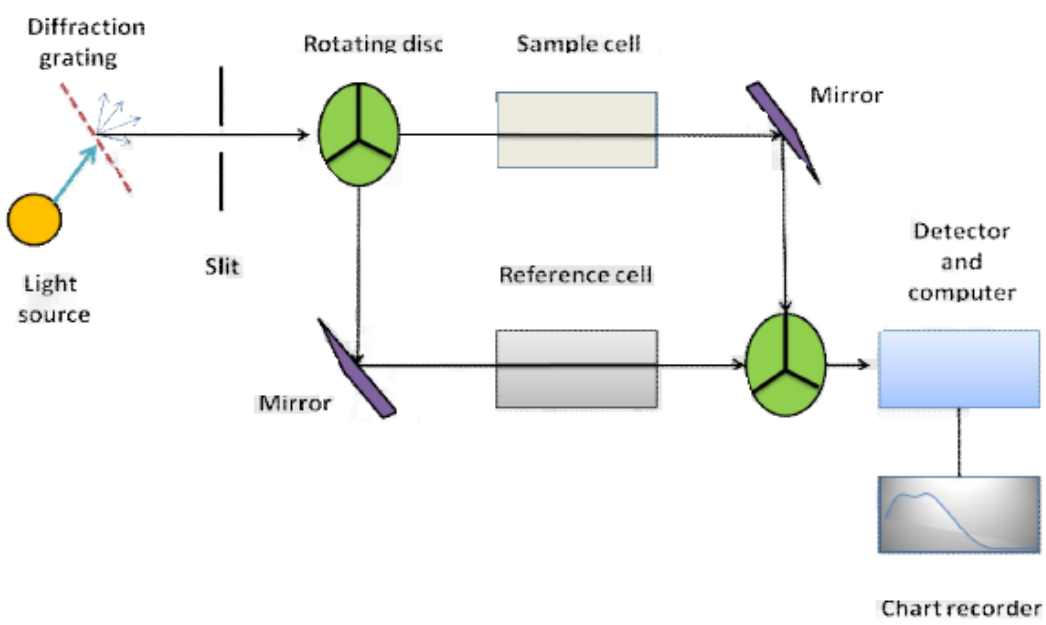


Fig 3.9: Functional Block diagram of a double beam UV – Visible Spectrophotometer

The sample and reference cells are rectangular quartz / glass containers; they contain the solution (to be tested) and pure solvent, respectively. The spectrometer records the ratio between the reference and sample beam intensities. The recorder plots the absorbance (A) against the wavelength (λ).

3.8 PHOTOLUMINESCENCE SPECTROSCOPY (PL)

Photoluminescence spectroscopy is a contactless, non - destructive method to probe the electronic structure of materials. The intensity and spectral content of the emitted photoluminescence is a direct measure of various important material properties, including band gap determination, impurity levels and defect detection, recombination mechanisms.

Principle

Light is directed onto a sample, where it is absorbed and imparts excess energy into the material in a process called photo - excitation. Photo – excitation causes electrons within a material to move into permissible excited states. These electrons return to their equilibrium states, by a radiative process (the emission of light) or by a non - radiative process as shown in Figure 3.10. The quantity of the emitted light is related to the relative contribution of the radiative process.

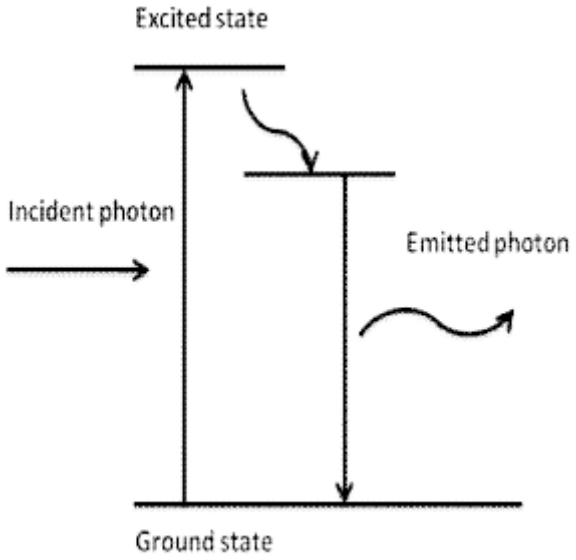


Fig 3.10 : Principle of photoluminescence

The fluorescence instruments contain three basic items: a source of light, a sample holder and a detector. A schematic representation of a fluorimeter is shown in Figure 3.11. The light source produces light photons over a broad energy spectrum, typically ranging from 200 to 900 nm. Photons impinge on the excitation monochromator, which selectively transmits light in a narrow range centered about the specified excitation wavelength. The transmitted light passes through adjustable slits that control magnitude and resolution by further limiting the range of transmitted light. The filtered light passes into the sample cell causing fluorescent emission by fluorphors within the sample. Emitted light enters the emission monochromator, which is positioned at a 90 º angle from the excitation light path to eliminate background signal and minimize noise due to stray light. Again, emitted light is transmitted in a narrow range centered about the specified emission wavelength and exits through adjustable slits, finally entering the photomultiplier tube (PMT). The signal is amplified and creates a voltage that is proportional to the measured emitted intensity. Noise in the counting process arises primarily in the PMT. Therefore, spectral resolution and signal to noise is directly related to the selected slit widths.

Sample preparation process is the same as that of UV - Visible spectroscopy. In both the cases, the sample cell (cuvette) must be free from contaminants.

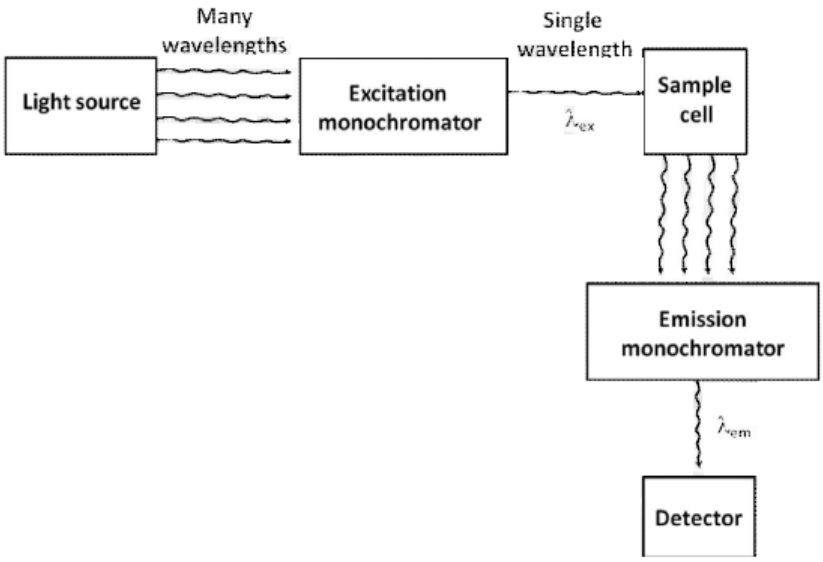


Figure 3.11 Schematic representation of a fluorescence spectrophotometer

MODULE IV
SEMICONDUCTOR QUANTUM NANOSTRUCTURES AND SUPERLATTICES

4.1 INTRODUCTION

Modern microelectronic and optoelectmnic devices can be constructed from a few basic structures, a mong them heterojunctions, or junctions between two semiconductors of different gaps, and metal-oxide-semioonductor (MOS) structures. These two kinds of nanostructures provide electrons with a potential well for electrons of nanometric size at the junction. MOS structure is the basic block of the most important device in microelectronics, the metal-oxide-semiconductor field-effect-transistor (MOSFET. In fact, the discovery of the quantum Hall effect (QHE) by IL von Klitzing in 1980 was based on the study of the transport properties of electrons in the channel of a MOSFET under the influence of simultaneous electric and magnetic fields [1]. However, these new quantum effects (QHE, Aharonov-Bohm effect and Shubnikovde Haas oscillations) are better observed in III-V heterojunctions, since the electron effective mass is much lower in materials such as GaAs than in Si.

In this chapter we shall study first the behaviour of electrons confined in 2D wells at the semiconductor-oxide interface in MOSFET transistors. Next, we shall proceed with the III-V modulation-doped heterojunction used in high-frequency transistors. Next. we will focus on the modulation-doped. square potential, quantum wells. This simple building block is used as a single unit or more often as a multiple quantum well structure in devices. ***When the thickness of the barriers separating the wells is small, tunnelling of elections between neighbouring wells takes place and the resulting device called a superlattice***, displays a band diagram similar to that of electron bands in crystals. However, the allowed energy bands and gaps correspond to much smaller energy intervals, since the SL has a spatial periodicity (equal to the sum of well and barrier thicknesses) which is much larger than the lattice constant. The band structure of SL can be engineered by a proper choice of the well and barrier widths. In this sense, SLs can be considered as artificial solids since their electron energy band structure is similar to the ones in crystals, but they do not exist in nature.

4.2 MOSFET STRUCTURES

The MOSFET is formed by a MOS structure and two p-n"' junctions in which the n material is heavily doped (Figure 4.1(a)), which act as the source and drain of the FET. The gate of the transistor is formed by the MOS structure. The semiconductor is usually p-type silicon over which a thin oxide layer (gate oxide)is grown by thermal oxidation.

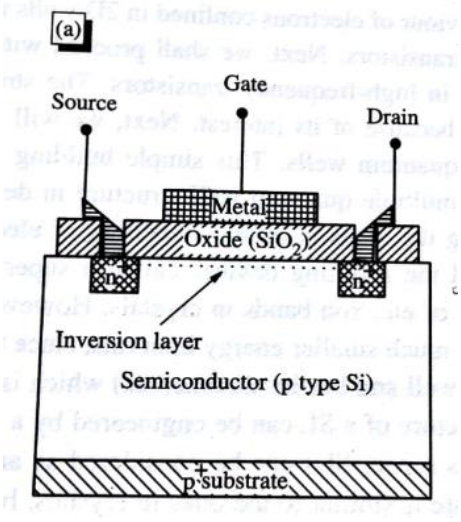


Figure 4.1(a)Structure of MOS Transistor

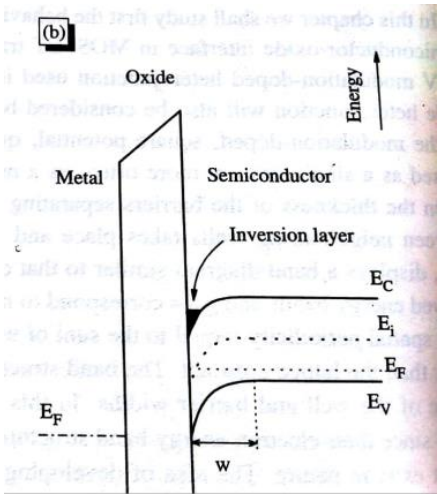


Figure 4.1(b) : Band diagram of MOS structure

Figure 4.1(b) shows the band diagram of the MOS structure for a p-type silicon semiconductor under a fairly strong positive bias. When a positive potential is applied to the gate, the electrons coming from the n+ regions and some from the bulk p-silicon are accumulated at the Si-Si0_2 interface. These electrons form the so-called inversion layer or channel and are located in an almost triangular-shaped potential well of nanometric dimensions. The shape of the well is due to the space charge of ionized acceptors in the p-type silicon, whose corresponding holes are repelled by the electric held across the dielectric oxide produced by the positive gate potential. If a positive potential is applied between drain and source, the electrons in the channel will create a current. The current can be modulated by changes at the potential gate, since the amount of electrons in the inversion layer depends on the magnitude of the electric field across the insulator. The name “Field Effect Transistor” is due to the control effect of the electric field applied to the gate.

In order to observe quantum effects at the Si-SiO_2 interface, several conditions must be met: (a) the SiO_2 insulator, of amorphous nature, should have neither high concentration of impurities (Na+ ions) nor of trapped charge; (b) the smoothness of the Si-Si02 interface should be controlled

at the atomic size level, since a rough surface on top of the inversion channel would greatly decrease the electron mobility in the inversion layer.

In order to study the behaviour of electrons in the potential well, it should be recognized that the electron inversion layer can be considered a 2D system of electrons immersed in a triangular-shaped quantum well ,located in the semiconductor, close to the interface with the oxide. In the MOS structure, although the electrons are confined along the perpendicular direction, they are practically free to move in the plane of the interface. Therefore, the quantized values for the energy of confinement should be given by

$$E = En + \frac{h^2}{2m_{x^*}} {k_x}^2 + \frac{h^2}{2m_{y^*}} {k_y}^2 \quad \text{.....(1)}$$

Where En corresponds to quantized energy for triangular well.

$$En \approx \left(\frac{3}{2}\pi\left(n - \frac{1}{4}\right)\right)^{\frac{2}{3}} \left(\frac{e^2F^2h^2}{2m_{Z^*}}\right)^{\frac{1}{3}}, \quad n = 1,2,3.. \quad \text{.....(2)}$$

Eqn (1) represents parabolas in reciprocal space , the bottoms having values of En.

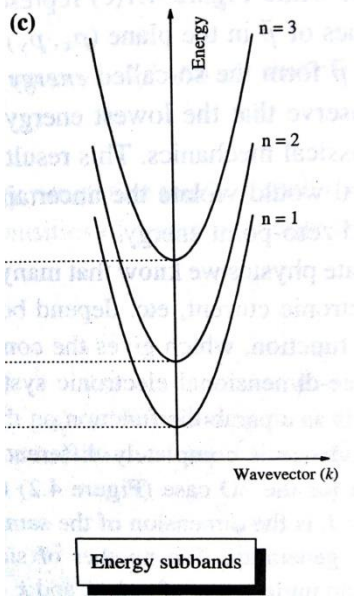


Figure 4.2 : Energy subbands

Similarly the density of states (DOS) function corresponds to the 2D case and is given by

$$g(E) = g_v \frac{m_T^*}{\pi h^2} \quad \text{.....(3)}$$

where we have added the factor g_v which takes into account the conduction band valley degeneracy. This degeneracy arises from the fact that constant energy surfaces of the silicon conduction band are formed by six ellipsoids in the <001> direction of momentum space.

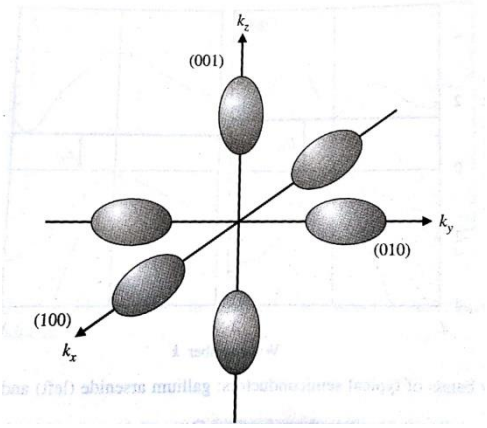


Figure 4.3 : Constant energy surfaces in silicon around the minima of conduction band

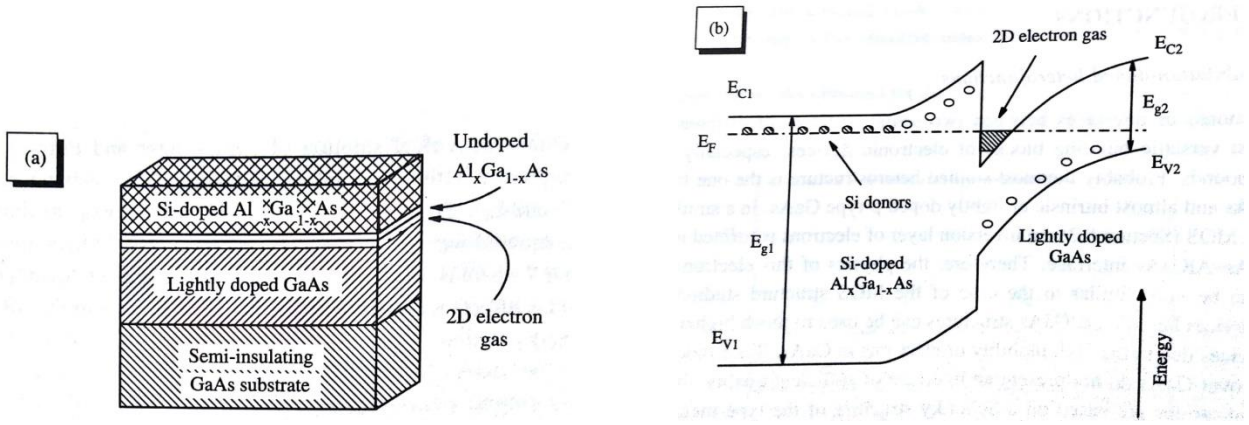
The long axis of the ellipsoid corresponds to the longitudinal effective mass $m_T{}^*$=0.91m_o and the two equal short axes to the transversal effective mass $m_L{}^*$ = 0.19mo, which is the one that appears in Eq. (3). Therefore, there are two conduction band minima corresponding to the heavy effective mass and four to the light one. As a consequence, after solving the Schrodinger equation, neglecting coupling of the electrons in the various conduction band minima, one should expect two different subband values or subband ladders for the Si <001>. First consider the electrons in valleys perpendicular to the interfaces. The effective mass which enters Eq. (2) is $m_L{}^*$ Also, $m_x{}^*$ =$m_y{}^*$; = $m_T{}^*$ in Eq. (1) and g_v = 2. This results in subbands of lower energy (higher value of effective mass). Evidently the second subband ladder is originated when the parallel valleys are considered. In this case, m_T is the effective mass for the expression of En in Eq. (2) and gv = 4.

4.2 HETEROJUNCTIONS

4.2.1 MODULATION-DOPED HETEROJUNCTIONS /HIGH ELECTRON MOBILITY TRANSISTOR

Heterojunctions, or interfaces between two semiconductors of different gaps, are one of the most versatile building blocks of electronic devices, especially those based on III-V compounds. Probably the most studied heterostructure is the one formed by n-type Al_x Ga_{1-x} As and almost intrinsic or lightly doped p-type GaAs. In a similar fashion to the case of the MOS, an inversion layer of electrons is formed in the GaAs close to the GaAs-AlGaAs interface. Devices based on AlGaAs structures can be used to much higher frequencies than silicon devices due to the high mobility of electrons in GaAs. Since oxides and insulators deposited over GaAs do not present an interface of sufficient quality, the most important device applications are based on a Schottky

structure of the type metal-AlGaAs-GaAs (Figure 4.4 (a)). In this section we will focus on the properties and band diagram of the Alea1_xAs-GaAs heterojunction shown m Figure 4.4(b).



4.4 (a) : Structure of a AlGaAs-GaAs Modulation dopped Hetero junction (b) Band diagram

Let us first consider, from a qualitative point of view, how an electron well of nanometric size is formed at the AlGaAs-GaAs interface.

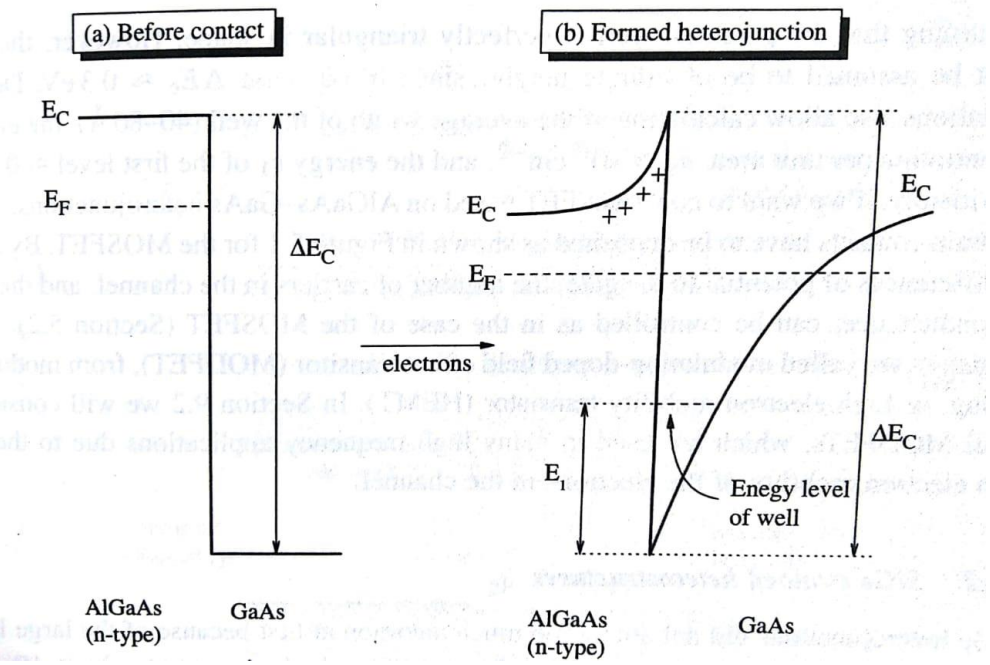


Figure 4.5 Formation of band diagram in modulation doped heterojunction AlGa As-GaAs when both semiconductors enter in contact

Suppose, as in Figure 4.5, that we have an AlGaAs-GaAs heterojunction, where the left material is gallium arsenide doped with aluminium and the right one is near-intrinsic GaAs. This structm‘e is called a modulation-doped heterojunction and the method to produce it is known as modulation doping. First consider the hypothetical situation of Figure 4.5(a), before the two semiconductors enter in contact. In the figure, for simplicity, we only draw the bottom of the conduction bands and the Fermi level, which in the case of n-type AlGaAs is close to the conduction band, and for lightly p-doped GaAs is located close to the middle of the gap. Evidently the bands are flat because the materials are electrically neutral and have uniform doping. The barrier between them in the conduction band, ΔEc, can be approximately found following Anderson's rule. According to this rule, when we join two materials, the vacuum levels should line up. If χ_AA and χ_B are the electron afiinities of the AlGaAs and GaAs, respectively, we should have ΔEc ≡ $\chi_A - \chi_B$, since the electron aflinity of a semiconductor is defined as the energy required for an electron located at the bottom Ec of the conduction band, to get out of the solid, i.e. χ = E_{vac} -EC. According to this rule, one gets a value of ΔEc of 0.35 eV for a doping x in Al_x Ga_{1-x} As around 0.3.

When both materials, AlGaAs and GaAs, enter in contact, some of the electrons from the donors of the n-material will cross the interface reaching the undoped GaAs. Therefore, as in the p-n junction, an internal electric field will be created and directed from the non-neutralized donors in the AlGaAs to the additional electronic charges in the GaAs. This field is the one that causes the band bending shown in Figure 4.4(b). At equilibrium, the two Fermi levels line up, the bands are bent like in the case of the p-n junction, with the only difference that the barrier ΔEc is created. Note also that far from the interface the bottom EC of the conduction bands is flat and at the same distance from the Fermi level E_F as in the case of Figure 4.5(a). Therefore, it is relatively easy to sketch the band diagram of Figure 4.5(b).As it can be appreciated, a quantum well for the electrons has been formed which is limited by a potential well of height ΔEc in the left and a potential energy curve at the right, within the GaAs. The quantum well for the electrons produced at the AlGaAs-GaAs interface has a shape close to a triangle as in the case of the MOS structure. Therefore, if we call Z the direction perpendicular to the interface, the electrons forming the 2D inversion layer are free to move along the (x, y) plane, but their energy for the motion along Z is quantized as in a potential well.

The most important aspect of this heterojunction is that the charge carriers are located in a region (mainly in the GaAs), spatially separated from the AlGaAs semiconductor which originates the free electrons. The electrons in the well should have very high mobility for their motion along the (x, y) plane, since they move within the GaAs which is free of dopent impurities and it is well known that impurity scattering is one of the main factors which limit carrier mobility, especially at low temperatures. Evidently, the electron mobilities are also much higher than in the case of the MOS structure.

4.2.2 SiGe STRAINED HETEROSTRUCTURES

Figure 4.6 shows two typical examples of SiGe heterostructures. In Figure 4.6(a) the substrate is <001> Si (Ex = 1.17 eV) and the strained active layer $Si_{0.7}Ge_{0.3}$ (Eg = 0.78 eV). In this case the conduction band offset is rather small. in contrast to the valence band offset. This situation allows the formation of a 2D hole gas in the SiGe alloy, with electron mobilities around 2 $m^2V^{-1}S^{-1}$, i.e. about half the value found for electrons in a typical MOSFET .In Figure 4.6(b) the situation is reversed and the strained layer is Si. In this case the discontinuity in the conduction band is fairly large and the electrons form a 2D gas, with free motion in the plane of the interface The silicon effective mass corresponding to this motion is the low transversal one ($m_T{}^* \approx 0.19m_0$)therefore yielding a high mobility of around 20 $m^2V^{-1}S^{-1}$,several times higher than the one corresponding to the MOSFET.

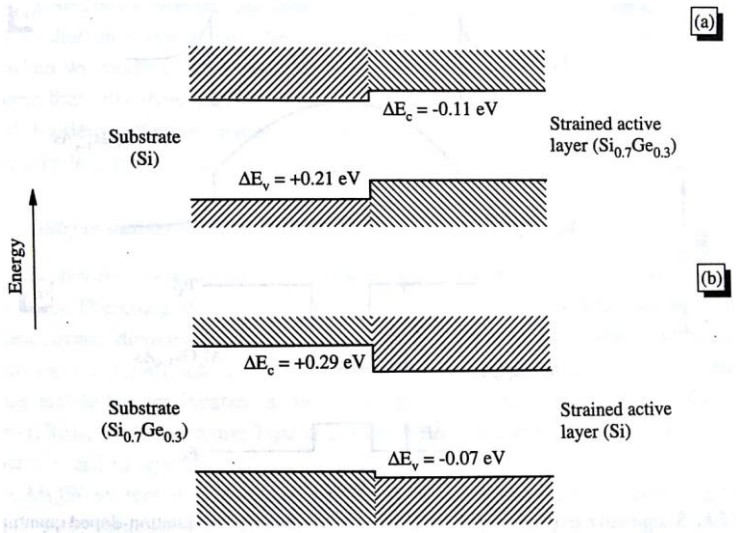


Figure 4.6: Band diagram of two strained SiGe heterostructures
(a). active layer is SiGe (b). Substrate layer is SiGe

SiGe heterostructures have also found an important application in the field of bipolar silicon transistors. One way to improve the efficiency of a bipolar transistors is to use a narrow-bandgap material for the base region, which improves the efficiency of the Si emitter region. In this case the advantage stems from the reduced values of the SiGe bandgap alloys, in comparison to silicon. It is important to mention that the strain that appears in the heterojunction also contributes to the decrease of the bandgap. In addition, the large bandgap offset allows the fabrication of a highly doped, low resistivity, base material, which extends the performance of silicon transistors to much higher frequencies.

4.3 QUANTUM WELLS

4.3.1 MODULATION DOPED QUANTUM WELLS

Modulation doping is a technique for fabricating semiconductors such that the free charge carriers are spatially separated from the donors. Because this eliminates scattering from the donors, modulation-doped semiconductors have very high carrier mobilities.

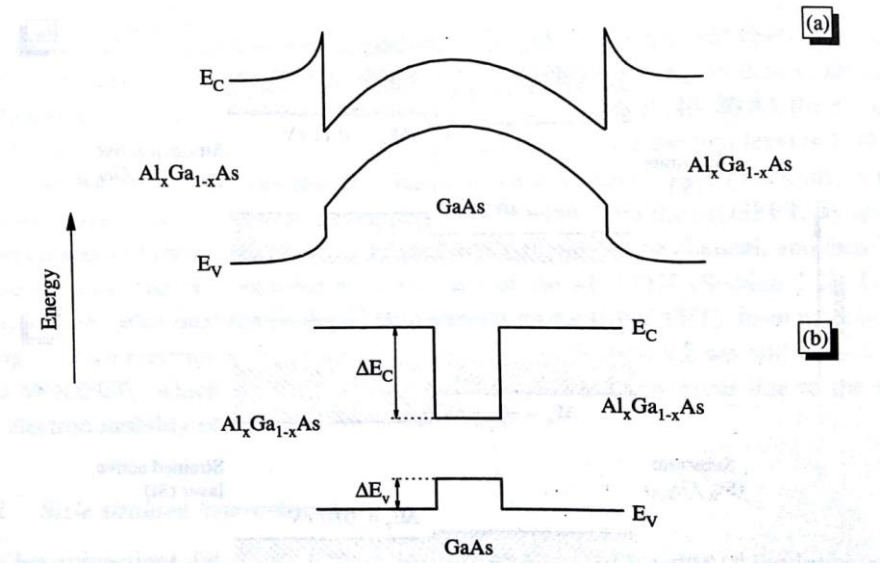


Figure 4.7 Band diagram of AlGaAs-GaAs modulation dopped quantum well (a). wide well (b). Narrow well

To build a symmetric well by facing two AlGaAs-GaAs heterojunctions opposing each other like in Figure 4.7 (a). The wide gap semiconductor material AI_xGa_{1-x} As is located at the ends and the GaAs in the middle. Assuming the distance between the two interfaces is sufficiently small. Then the resulting well (Figure 4.7(b)) for electrons and holes would be almost square with a barrier on each side of the same height as in Figure 4.7(a).

The volume of material inside the well is free of ionized donors (located in the AlGaAs material). Therefore the electrons inside the well, which originated at the neighbouring AlGaAs donor-material, can move into the GaAs region or channel with very high mobility. As in the case of the modulation-doped heterojunctions, MODFET high-frequency transistors can be fabricated if appropriate source and drain contacts are deposited.

Quantum well structures with either high or low mobility for electrons can be fabricated by introducing a controlled amount of impurities. A double quantum well structure with high and low mobilities constitutes the base of the velocity-modulation transistors. In these transistors, the switching from one state to the other is controlled by an electric field transverse to the layers which redistributes the amount of charged electrons and therefore the current in either wells. Velocity-modulation transistors can be operated at very high frequencies.

4.3.2 MULTIPLE QUANTUM WELLS (MQW)

The signal provided by a single quantum well is usually too small to be used in the solid state devices. Therefore, it is often necessary to use an array of quantum wells, especially in optoelectronic devices, such as photodetectors. These structures are called multiple quantum wells (MQW) and are formed by several single quantum wells. If the wells for electrons and holes are located in the same space location, the MQW is called Type I (Figure 4.8(a)), while the name Type II is used when the corresponding wells are located alternatively as in Figure 4.8 (b).

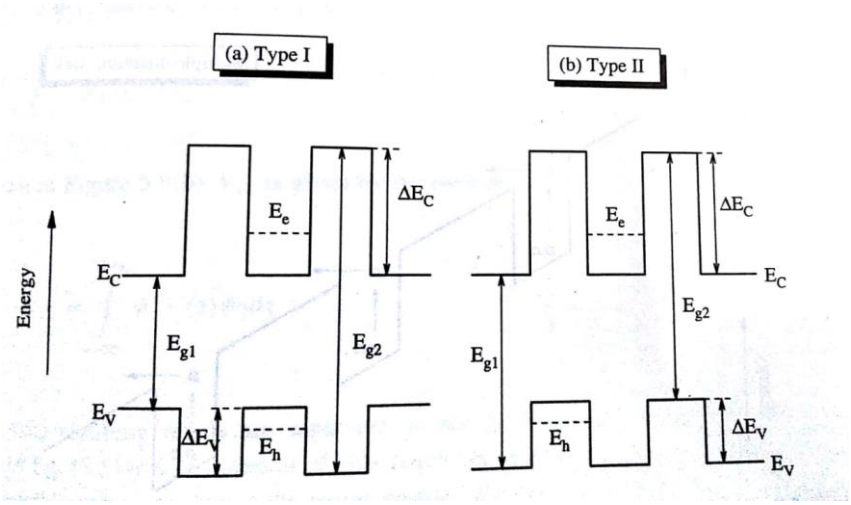


Figure 4.8: (a) Type I MQW (b) Type II MQW

In a MQW system it is assumed that there is no interaction between neighbouring quantum wells, because the barriers separating the wells are thick enough, usually more than about 10 nm.

However, if the energy barriers between consecutive wells are thin enough, the wells will be coupled to each other by tunnelling effects. The discrete energy levels of the quantum wells are then transformed into energy bands. In this case, the system of MQWs is called a ***super lattice*** and the energy spectrum shows very interesting new features.

4.4 THE CONCEPT OF SUPER LATTICES

A superlattice consists of a periodic set of MQW in which the thickness of the energy barriers separating the individual wells is made sufficiently small. As the barriers become thinner, the electron wave functions corresponding to the wells overlap due to the tunnelling effect. As a consequence, the discrete energy levels of the wells broaden and produce energy bands, in a similar way as happens with the states of the individual atoms when they are arranged in a crystal lattice. The most singular aspect of a superlattice consists of introducing at will a new periodicity ***d*** in the material, which is equal to the breadth of the well ***a***, plus the thickness of the barrier ***b.*** Typical thicknesses for a and b could be 4 and 2 nm, respectively. An accurate control over these small thicknesses can only be achieved by techniques for thin film deposition such as molecular beam epitaxy or metal organic chemical vapour deposition.

In order to study the origin of the band structure of superlattices, let us consider first the overlapping between the electron states for a simple two-well system. This is already a familiar problem, because from a quantum mechanical point of view it is formally similar to the case of the diatomic molecule.

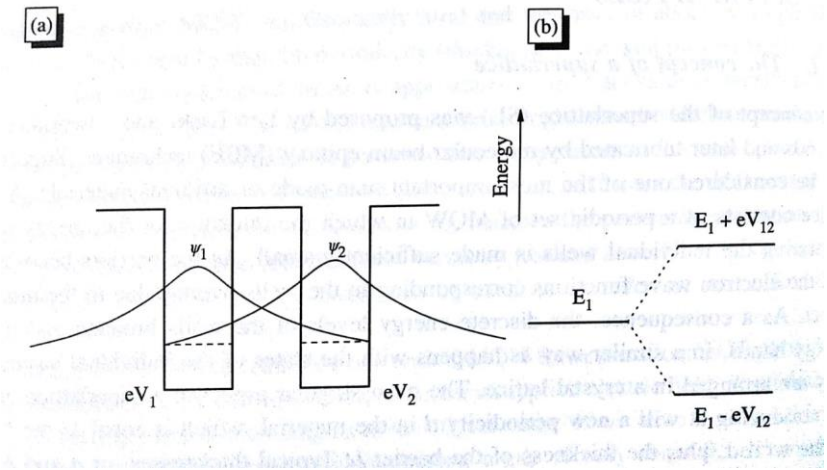


Figure 4.9: (a) Double coupled quantum well (b) Splitting of energy levels

Figure 4.9(a) shows two neighbouring identical quantum wells and corresponding wave functions of what is known as the double coupled quantum well system. The solution for this problem is based on perturbation theory in quantum mechanics. According to it, each original level, say E_1, of the isolated wells splits into two, with energies

$$E = E_1 \pm |V_{12}| \quad \text{.....(1)}$$

$$V_{12} = \int_{-\infty}^{\infty} \psi_1^* V(z) \psi_2 dz \quad \text{.....(2)}$$

The two resulting levels are separated in energy by $2|V_{12}|$ where the magnitude of V_{12} of Eq. (2) is an indication of how much one well can iniiuence the energy states of the neighbouring one, hence the name overlap integral.

4.5 KRONIG - PENNEY MODEL OF SUPER LATTICE

The fundamental nature of insulators, conductors and semiconductors can be functionally explained based on band theory. The recent development in semiconductor physics, the semiconductor hetero-structures are also analyzed using the concept of band theory. In the Kronig Penney model, instead of experiencing a gradual variation in the strength of the potential electrons experience a maximum potential (potential well) and minimum value (potential barrier) in the presence of the lattice planes. Kronig and Penney established a model for a solid in which the periodic potential seen by the electrons was precisely that of the square type shown in Figure 4.10 for a superlattice potential.

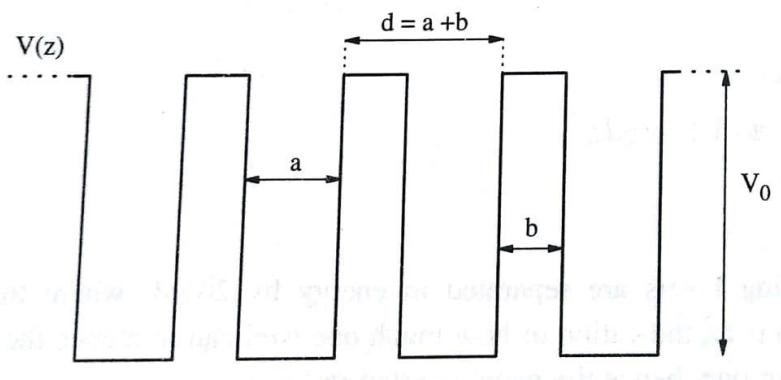


Figure 4.10: Scheme of the periodic potential of a superlattice

This periodic one-dimensional potential is characterized by the following parameters: well thickness a, barrier thickness b, and barrier height V_0. The spatial periodicity is d = a + b. Even if the periodic potential of Figure 4.10 is much simpler than the real one, the Kronig-Penney model yields very interesting results related to the structure of the bands, the forbidden zones, size of the gaps, etc. In the well region (0 < z < a), V = 0, and the wave function is

$$\psi(z) = Ae^{ik_0 z} + Be^{-ik_0 z} \quad \text{.....(1)}$$

With

$$k_0{}^2 = \frac{2mE}{h^2} \quad \text{.....(2)}$$

Due to tunnelling ,the wave function extends inside the energy barrier of height Vo and thickness b.Therefore if -b<z<0;

$$\psi(z) = Ce^{qz} + De^{-qz} \quad \text{.....(3)}$$

Where the wave vector and the energy are related by

$$Vo - E = \frac{h^2 q^2}{2m} \quad \text{.....(4)}$$

From the condition that both wave functions and their derivatives are continuous at z=0 and z= a

We get after operating

A+B = C+D

$$ik_0(A - B) = q(C - D) \quad \text{.....(5)}$$

According to the Bloch theorem, we can relate the wave functions at two different locations by

$$\psi(a) = \psi(-b)e^{ik(a+b)} \quad \text{.....(6)}$$

Where k is the wave vector corresponding to wave functions .Applying this equation to wave functions corresponding to the well and the barrier regions respectively, then

$$Ae^{ik_0 a} + Be^{-ik_0 a} = (Ce^{-qb} + De^{qb})e^{ik_0(a+b)} \quad \text{.....(7)}$$

$$iko\left(Ae^{ik_0 a} - Be^{-ik_0 a}\right) = q(Ce^{-qb} - De^{qb})e^{ik_0(a+b)} \quad \text{.....(8)}$$

The equations 5,6,7,8 for the amplitudes A,B,C and D have a solution only if the determinant of coefficients equals zero. After some calculations ,we get the relation

$\frac{q-{k_0}^2}{2qk_0}\sin k_0 a \sinh qb + \cos qa \cosh qb = \cos q(a+b)$ …….(9).

Let us assume the simple case of a = b and further that the effective mass of the electron is the same in the well and barrier materials. For the case of GaAs-AlGaAs and E < V_0, solution of Eq. (9) gives the values of allowed and forbidden energies.

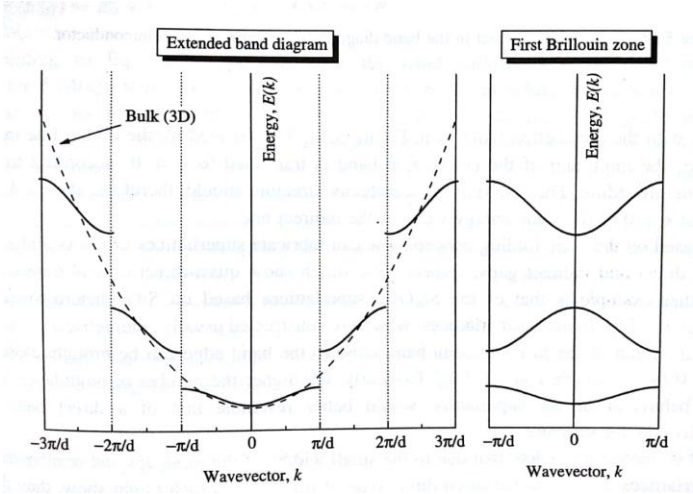


Figure 4.11: Extended band diagram of superlattice and band diagram in reduced first Brillouin zone

It is evident from Figure 4.11 that the general features of the E-k relationship in a superlattice, within the Kronig-Penney model are very similar to the case of electrons in solids if we replace the bulk lattice constant by the larger period of a superlattice. The free electron parabola, therefore, breaks down into several bands and gaps at the edges of the Brillouin zones$k = \pm\frac{n\pi}{d}$, as shown in the extended E-k diagram of Figme 4.11. Next, the portions of the bands can be translated to the reduced zone$(-\frac{\pi}{d} \le z \le +\frac{\pi}{d})$. Notice that everything occurs as if the superlattice potential folds the quasi-free energy band of a solid into the centre of the reduced zone, as it can be appreciated if we make the representation in the first Brillouin zone. Since, usually $d \gg a$, the breadth of the bands and gaps in a superlattice are much smaller, often receiving the names of minibands and minigaps. This band folding procedure is typical of superlattices and is called ***zone folding*** since it implies that the pieces of the band in the extended representation are zone-folded into the smaller zone with values of k: smaller than $\frac{2\pi}{d}$

The zone-folding effect has important consequences in the direct or indirect character of semiconductor structures.

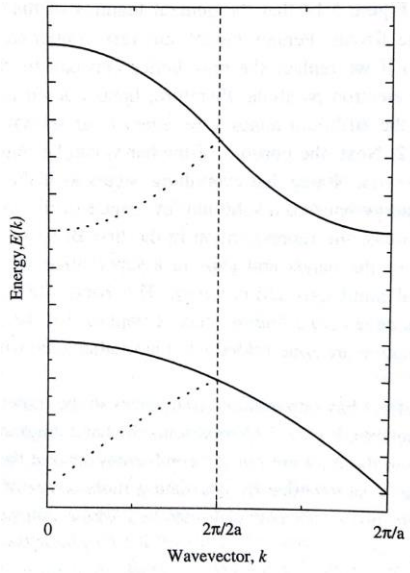


Figure 4.12: Zone folding effect in the band diagram of an indirect gap semiconductor

Figure 4.12 represents the band diagram of a typical indirect gap semiconductor with the minimum of the conduction band at the zone edge. Suppose next that we construct a superlattice by alternating monolayers of two semiconductors with similar electronic properties and well-matched lattice constants, but with one of them with the conduction band as in Figure 4.12. As indicated by the dotted line in the figure, the minimum of the conduction band is translated to k = 0, according to the folding procedure. The resulting semiconductor structure should, therefore, show a direct gap at k = 0 of the same energy value as the indirect one.

Based on the zone folding concept, one can fabricate superlattices of GaAs and AlAs with direct and indirect gaps, respectively, which show quasi-direct optical transitions. Another example is that of the Si_nGe_m superlattices based on SiGe heterostructures .In these superlattices, which are constructed usually symmetrical (n = m), the minimum of the Si conduction band close to the band edge can be brought close to k = 0 by successive zone-folding. Evidently, the higher the number of monolayers (n), the behaviour of the superlattice would better resemble that of a direct bandgap semiconductor structure.