

6

Black-body radiation

The subjects for consideration in this chapter are the black-body model, which is of primary importance in thermal radiation theory and practice, and the fundamental laws of radiation of such a system. Natural and artificial physical objects, which are close in their characteristics to black bodies, are considered here. The quantitative black-body radiation laws and their corollaries are analysed in detail. The notions of emissivity and absorptivity of physical bodies of grey-body radiation character are also introduced. The Kirchhoff law, its various forms and corollaries are analysed on this basis.

6.1 THE IDEAL BLACK-BODY MODEL: HISTORICAL ASPECTS

The ideal black-body notion (hereafter the black-body notion) is of primary importance in studying thermal radiation and electromagnetic radiation energy transfer in all wavelength bands. Being an ideal radiation absorber, the black body is used as a standard with which the absorption of real bodies is compared. As we shall see later, the black body also emits the maximum amount of radiation and, consequently, it is used as a standard for comparison with the radiation of real physical bodies. This notion, introduced by G. Kirchhoff in 1860, is so important that it is actively used in studying not only the intrinsic thermal radiation of natural media, but also the radiations caused by different physical nature. Moreover, this notion and its characteristics are sometimes used in describing and studying artificial, quasi-deterministic electromagnetic radiation (in radio- and TV-broadcasting and communications). The emissive properties of a black body are determined by means of quantum theory and are confirmed by experiment.

The black body is so called because those bodies that absorb incident visible light well seem black to the human eye. The term is, certainly, purely conventional and has, basically, historical roots. For example, we can hardly characterize our Sun,

which is, indeed, almost a black body within a very wide band of electromagnetic radiation wavelengths, as a black physical object in optics. Though, it is namely the bright-white sunlight, which represents the equilibrium black-body radiation. In this sense, we should treat the subjective human recognition of colours extremely cautiously. So, in the optical band a lot of surfaces really approach an ideal black body in their ability to absorb radiation (examples of such surfaces are: soot, silicon carbide, platinum and golden niellos). However, outside the visible light region, in the wavelength band of IR thermal radiation and in the radio-frequency bands, the situation is different. So, the majority of the Earth's surfaces (the water surface, ice, land) absorb infrared radiation well, and, for this reason, in the thermal IR band these physical objects are ideal black bodies. At the same time, in the radio-frequency band the absorptive properties of the same media differ both from a black body and from each other, which, generally speaking, just indicates the high information capacity of microwave remote measurements.

6.1.1 Definition of a black body

A black body is an ideal body which allows the whole of the incident radiation to pass into itself (without reflecting the energy) and absorbs within itself this whole incident radiation (without passing on the energy). This property is valid for radiation corresponding to all wavelengths and to all angles of incidence. Therefore, the black body is an ideal absorber of incident radiation. All other qualitative characteristics determining the behaviour of a black body follow from this definition (see, for example, Siegel and Howell, 1972; Ozisik, 1973).

6.1.2 Properties of a black body

A black body not only absorbs radiation ideally, but possesses other important properties which will be considered below.

Consider a black body at constant temperature, placed inside a fully insulated cavity of arbitrary shape, whose walls are also formed by ideal black bodies at constant temperature, which initially differs from the temperature of the body inside. After some time the black body and the closed cavity will have a common equilibrium temperature. Under equilibrium conditions the black body must emit exactly the same amount of radiation as it absorbs. To prove this, we shall consider what would happen if the incoming and outgoing radiation energies were not equal. In this case the temperature of a body placed inside a cavity would begin to increase or decrease, which would correspond to heat transfer from a cold to a heated body. But this situation contradicts the second law of thermodynamics (the question is, certainly, on the stationary state of an object and ambient radiation). Since, by definition, the black body absorbs a maximum possible amount of radiation that comes in any direction from a closed cavity at any wavelength, it should also emit a maximum possible amount of radiation (*as an ideal emitter*). This situation becomes clear if we consider any less perfectly absorbing body (a grey body), which should

emit a lower amount of radiation as compared to the black body, in order that equilibrium be maintained.

Let us now consider an isothermal closed cavity of arbitrary shape with black walls. We move the black body inside the cavity into another position and change its orientation. The black body should keep the same temperature, since the whole closed system remains isothermal. Therefore, the black body should emit the same amount of radiation as before. Being at equilibrium, it should receive the same amount of radiation from the cavity walls. Thus, the total radiation received by the black body does not depend on its orientation and position inside the cavity; therefore, the radiation passing through any point inside a cavity does not depend on its position or on the direction of emission. This implies that the equilibrium thermal radiation filling a cavity is isotropic (the property of isotropy of black-body radiation). And, thus, the net radiation flux (see equation (5.7)) through any plane, placed inside a cavity in any arbitrary manner, will be strictly zero.

Consider now an element of the surface of a black isothermal closed cavity and the elementary black body inside this cavity. Some part of the surface element's radiation falls on a black body at some angle to its surface. All this radiation is absorbed, by definition. In order that the thermal equilibrium and radiation isotropy be kept throughout the closed cavity, the radiation emitted by a body in the direction opposite to the incident beam direction should be equal to the absorbed radiation. Since the body absorbs maximum radiation from any direction, it should also emit maximum radiation in any direction. Moreover, since the equilibrium thermal radiation filling the cavity is isotropic, the radiation absorbed or emitted in any direction by the ideal black surface encased in the closed cavity, and related to the unit area of surface projection on a plane normal to the beam direction, should be equal.

Let us consider a system comprising a black body inside a closed cavity which is at thermal equilibrium. The wall of the cavity possesses a peculiar property: it can emit and absorb radiation within a narrow wavelength band only. The black body, being an ideal energy absorber, absorbs the whole incident radiation in this wavelength band. In order that the thermal equilibrium be kept in a closed cavity, the black body should emit radiation within the aforementioned wavelength band; and this radiation can then be absorbed by the cavity wall, which absorbs in the given wavelength band only. Since the black body absorbs maximum radiation in a certain wavelength band, it should emit maximum radiation in the same band. The black body should also emit maximum radiation at the given wavelength. Thus, the black body is an ideal emitter at any wavelength. However, this in no way implies uniformity in the intensity of black-body emission at different wavelengths (the 'white noise' property). The peculiar spectral (and, accordingly, correlation) properties of black-body radiation could only be revealed by means of quantum mechanics.

The peculiar properties of a closed cavity have no relation to the black body in the reasoning given, since the emission properties of a body depend on its nature only and do not depend on the properties of a cavity. The walls of a cavity can even be fully reflecting (mirroring).

If the temperature of a closed cavity changes, then, accordingly, the temperature of a black body enclosed inside it should also change and become equal to the new temperature of a cavity (i.e. a fully insulated system should tend to thermodynamic equilibrium). The system will again become isothermal, and the energy of radiation absorbed by a black body will again be equal to the energy of radiation emitted by it, but it will slightly differ in magnitude from the energy corresponding to the former temperature. Since, by definition, the body absorbs (and, hence, emits) the maximum radiation corresponding to the given temperature, the characteristics of an enclosing system have no influence on the emission properties of a black body. Therefore, the total radiation energy of a black body is a function of its temperature only.

In addition, according to the second law of thermodynamics, energy transfer from a cold surface to a hot one is impossible without doing some work at a system. If the energy of radiation emitted by a black body increased with decreasing temperature, then the reasoning could easily be constructed (see, for example, Siegel and Howell, 1972), which would lead us to a violation of this law. As an example, two infinite parallel ideal black plates are usually considered. The upper plate is maintained at temperature higher than the temperature of the lower plate. If the energy of emitted radiation decreased with increasing temperature, then the energy of radiation, emitted by the lower plate per unit time, would be greater than the energy of radiation emitted by the upper plate per unit time. Since both plates are black, each of them absorbs the whole radiation emitted by the other plate. For maintaining the temperatures of plates the energy should be rejected from the upper plate per unit time and added to the lower plate in the same amount. Thus, it happens, that the energy transfers from a less heated plate to more heated one without any external work being done. According to the second law of thermodynamics, this situation is impossible. Therefore, the energy of radiation emitted by a black body, should increase with temperature. On the basis of these considerations we come to the conclusion, that the total energy of radiation emitted by a black body is proportional to a monotonously increasing function of thermodynamic temperature only.

All the reasoning we set forth above proceeding from thermodynamic considerations represents quite important, but, nevertheless, only qualitative, laws of black-body radiation. As was ascertained, classical thermodynamics is not capable of formulating the quantitative laws of black-body radiation in principle.

6.1.3 Historical aspects

Until the middle of the nineteenth century a great volume of diverse experimental data on the radiation of heated bodies was accumulated. The time had come to comprehend the data theoretically. And it was Kirchhoff who took two important steps in this direction. At the first step Kirchhoff, together with Bunsen, established the fact that a quite specific spectrum (the set of wavelengths, or frequencies) of the light emitted and absorbed by a substance corresponds to that particular substance. This discovery served as a basis for the spectral analysis of substances. The second step consisted in finding the conditions, under which the radiation spectrum of

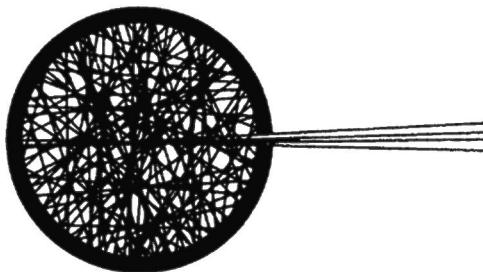


Figure 6.1. Classic experimental model of black-body source.

heated bodies depends only on their temperature and does not depend on the chemical composition of the emitting substance. Kirchhoff considered theoretically the radiation inside a closed cavity in a rigid body, whose walls possess some particular temperature. In such a cavity the walls emit as much energy as they absorb. It was found that under these conditions the energy distribution in the radiation spectrum does not depend on the material the walls are made of. Such a radiation was called ‘absolutely (or ideally) black’.

For a long time, however, black-body radiation was, so to speak, a ‘thing-in-itself’. Only 35 years later, in 1895, W. Wien and O. Lummer suggested the development of a test model of an ideal black body to verify Kirchhoff’s theory experimentally. This model was manufactured as a hollow sphere with internal reflecting walls and a narrow hole in the wall, the hole diameter being small as compared to the sphere diameter. The authors proposed to investigate the spectrum of radiation issuing through this hole (Figure 6.1). Any light beam undergoes multiple reflections inside a cavity and, actually, cannot exit through the hole. At the same time, if the walls are at a high temperature the hole will brightly shine (if the process occurs in the optical band) owing to the electromagnetic radiation issuing from inside the cavity. It was this particular test model of a black body on which the experimental investigations to verify thermal radiation laws were carried out, and, first of all, the fundamental spectral dependence of black-body radiation on frequency and temperature (the Planck formula) was established quantitatively. The success of these experimental (and, a little bit later, theoretical) quantum-approach-based investigations was so significant that for a long time, up until now, this famous reflecting cavity has been considered in general physics textbooks as a unique black-body example. And, thus, some illusion of black body exclusiveness with respect to natural objects arises. In reality, however (as we well know both from the radio-astronomical and remote sensing data, and from the data of physical (laboratory) experiments), the natural world around us, is virtually saturated with physical objects which are very close to black-body models in their characteristics.

First of all, we should mention here *the cosmic microwave background (CMB) of the universe* – the fluctuation electromagnetic radiation that fills the part of the universe known to us. The radiation possesses nearly isotropic spatial-angular field with an intensity that can be characterized by the radiobrightness temperature

of 2.73 K. The microwave background is, in essence, some kind of ‘absolute ether at rest’ that physicists intensively sought at the beginning of the twentieth century. A small dipole component in the spatial-angular field of the microwave background allowed the researchers to determine, to a surprising accuracy, the direction and velocity of motion of the solar system. The contribution of the microwave background as a re-reflected radiation should certainly be taken into account in performing fine investigations of the emissive characteristics of terrestrial surfaces from spacecraft.

The second (but not less important) source of black-body radiation is the star nearest to the Earth – *the Sun* (see section 1.4). The direct radar experiments, performed in the 1950s and 1960s, have indicated a complete absence of a radio-echo (within the limits of the receiving equipment capability) within the wide wavelength band – in centimetre, millimetre and decimetre ranges. Detailed spectral studies of solar radiation in the optical and IR bands have indicated the presence of thermal black-body radiation with a brightness temperature of 5800 K at the Sun. In other bands of the electromagnetic field the situation is essentially more complicated – along with black-body radiation there exist powerful, non-stationary quasi-noise radiations (flares, storms), which are described, nevertheless, in thermal radiation terms.

The third space object is our home planet, – *the Earth*, which possesses radiation close to black-body radiation with a thermodynamic temperature of 287 K. The basic radiation energy is concentrated in the 8–12 micrometre band, in which almost all terrestrial surfaces possess black-body radiation properties. Just that small portion of radiation energy which falls in the radio-frequency band is of interest for microwave sensing. The detailed characteristics of radiation from terrestrial surfaces in this band have shown serious distinctions of many terrestrial media from the black-body model.

In experimental measurements of the radiation properties of real physical bodies it is necessary to have an ideally black surface or a black emitter as a standard. Since ideal black sources do not exist, some special technological approaches are applied to develop a realistic black-body model. So, in optics these models represent hollow metal cylinders having a small orifice and cone at the end, which are immersed in a thermostat with fixed (or reconstructed) temperature (Siegel and Howell, 1972). In the radio-frequency band segments of waveguides or coaxial lines, filled with absorbing substance (such as carbon-containing fillers), are applied. Multilayer absorbing covers, which are widely used in the military-technological area (for instance, Stealth technology), are applied as standard black surfaces in this band. It is clear, that objects covered with such an absorbing coat are strong emitters of the fluctuation electromagnetic field. It is important also to note that in the radio-frequency band a closed space with well-absorbing walls (such as a concrete with various fillers) represents a black-body cavity to a good approximation. For these reasons the performance of fine radiothermal investigations in closed rooms (indoors) makes no sense. (Of interest is the fact that it was in a closed laboratory room that in 1888 Hertz managed to measure for the first time the wavelength of electromagnetic radiation.)

6.2 BLACK-BODY RADIATION LAWS

But now we return to the quantitative laws of black-body radiation. The general thermodynamic considerations allowed Kirchhoff, Boltzmann and Wien to derive rigorously a series of important laws controlling the emission of heated bodies. However, these general considerations were insufficient for deriving a particular law of energy distribution in the ideal black-body radiation spectrum. It was W. Wien who advanced in this direction more than the others. In 1893 he spread the notions of temperature and entropy to thermal radiation and showed, that the maximum radiation in the black-body spectrum displaces to the side of shorter wavelengths with increasing temperature (the Wien displacement law); and at a given frequency the radiation intensity can depend on temperature only, as the parameter appeared in the (ν/T) ratio. In other words, the spectral intensity should depend on some function $f(\nu/T)$. The particular form of this function has remained unknown.

In 1896, proceeding from classical concepts, Wien derived the law of energy distribution in the black-body spectrum (the Wien radiation law). However, as was soon made clear, the formula of Wien's radiation law was correct only in the case of short (in relation to the intensity maximum) waves. Nevertheless, these two laws of Wien have played a considerable part in the development of quantum theory (the Nobel Prize, 1911).

J. Rayleigh (1900) and J. Jeans (1905) derived the spectral distribution of thermal radiation on the basis of the assumption that the classical idea on the uniform distribution of energy is valid. However, the temperature and frequency dependencies obtained basically differed from Wien's relationships.

According to the results of fairly accurate measurements, carried out before that time, and to some theoretical investigations, Wien's expression for spectral energy distribution was invalid at high temperatures and long wavelengths. This circumstance forced Planck to turn to consideration of harmonic oscillators, which have been taken as the sources and absorbers of radiation energy. Using some further assumptions on the mean energy of oscillators, Planck derived Wien's and the Rayleigh–Jeans' laws of radiation. Finally, Planck obtained the empirical equation, which very soon was reliably confirmed experimentally on the basis, first of all, of the Wien–Lummer black-body model. Searching for the theory modifications which would allow this empirical equation to be derived, Planck arrived at the assumptions constituting the quantum theory basis (the Nobel Prize, 1918).

6.2.1 The Planck law (formula)

According to quantum statistics principles, the spectral volume density of radiation energy can be determined (see relation (5.10)) by calculating the equilibrium distribution of photons, for which the radiation field entropy is maximum, and taking into consideration that the photon energy with frequency ν is equal to $h\nu$, where h is the Planck constant (Table A.4). If the radiation field is considered to be a gas obeying the Einstein–Bose statistics, then we obtain the Planck formula for the

volume density of radiation (see, for example, Schilling, 1972; Amit and Verbin, 1999):

$$u_\nu(T) d\nu = \frac{8\pi h\nu^3}{c^3} \frac{1}{[\exp(h\nu/kT) - 1]} d\nu, \quad (6.1)$$

where k is the Boltzmann constant (Table A.4).

Apart from a rigorous quantum derivation of Planck's formula, there exists a spectrum of heuristic approaches (see, for example, Penner, 1959).

From the remote sensing point of view, of principal significance is the other radiation field characteristic, namely, the spectral radiation intensity, which is measured at once by remote sensing devices. With allowance for relation (5.12), the spectral intensity of black-body radiation into the transparent medium with refractive index n will be specified by the following expression:

$$I_\nu(T, \nu) = \frac{2h\nu^3 n^2}{c_0^2} \frac{1}{[\exp(h\nu/kT) - 1]}. \quad (6.2)$$

It can easily be seen from this relation that the black-body radiation into the transparent medium is n^2 times greater than when emitting into a vacuum (the Clausius law).

In many practical applications in determining the spectral intensity of radiation the wavelength is used instead of frequency. It is impossible to transfer from frequency to wavelength by simply replacing the frequency with the wavelength in expression (6.2), because this expression includes the differential quantity. However, this expression can be transformed taking into account that the energy of radiation, emitted within the frequency band $d\nu$, that includes frequency ν , is equal to the energy of radiation, emitted within the wavelength band $d\lambda$ that includes the working wavelength λ ,

$$I_\nu(T, \nu) |d\nu| = I_\lambda(T, \lambda) |d\lambda|. \quad (6.3)$$

The wavelength depends on the medium, in which the radiation propagates (see section 1.6). Subscript 0 denotes that the considered medium is the vacuum. At the same time, the electromagnetic radiation frequency does not depend on the medium. The frequency and wavelength in a transparent dielectric medium (λ) are related by the equation:

$$\nu = \frac{c_0}{n\lambda}. \quad (6.4)$$

Supposing the refractive index of a transparent medium to be independent of the frequency, we shall obtain, after appropriate differentiation, the expression of Planck's formula for the intensity of black-body radiation into the transparent medium, expressed in terms of the wavelength in a medium, as:

$$I_\lambda(T, \lambda) = \frac{2hc_0^2}{n^2 \lambda^5} \frac{1}{[\exp(hc_0/n\lambda kT) - 1]} \quad (6.5)$$

In the SI system the intensity, presented in such a form, is measured in $\text{W}/(\text{m}^3 \text{sr})$. Often, in the IR band especially, the wavelength is measured in micrometres; then the

dimension of radiation intensity will be $\text{W}/(\text{m}^2 \text{sr } \mu\text{m})$. However, it is convenient to use the frequency presentation of Planck's formula (6.2) in the cases where the radiation propagates from one medium into another, since in this case the frequency remains constant and the wavelength changes.

In many practical applications (remote sensing, heat transfer, radio-astronomy) of interest is the surface density (per unit of the surface) of a spectral flux of black-body radiation determined in the form of equation (5.5) and (5.9). Substituting the spectral density value from (6.5), we have

$$q_\lambda(T) = \frac{C_1}{n^2 \lambda^5} \frac{1}{[\exp(C_2/n\lambda T) - 1]}, \quad (6.6)$$

where the quantities

$$C_1 = 2\pi h c_0^2; C_2 = \frac{hc_0}{k} \quad (6.7)$$

were called the first and second radiation constants (see Table A.4).

Note that $q_\lambda(T)$ represents the amount of radiation energy emitted by the unit area of the black-body surface at temperature T per unit time, in the wavelength band unit, in all directions within the limits of the hemispherical solid angle. In the SI system this quantity is measured in W/m^3 , and if the wavelength is measured in micrometres then this quantity is measured in $\text{W}/(\text{m}^2 \mu\text{m})$.

Figure 6.2 presents the spectral distribution of the surface density of a monochromatic black-body radiation flux $q_\lambda(T)$, calculated by formula (6.6) for $n = 1$. In order to understand better the implication of this equation, Figure 6.2 gives the

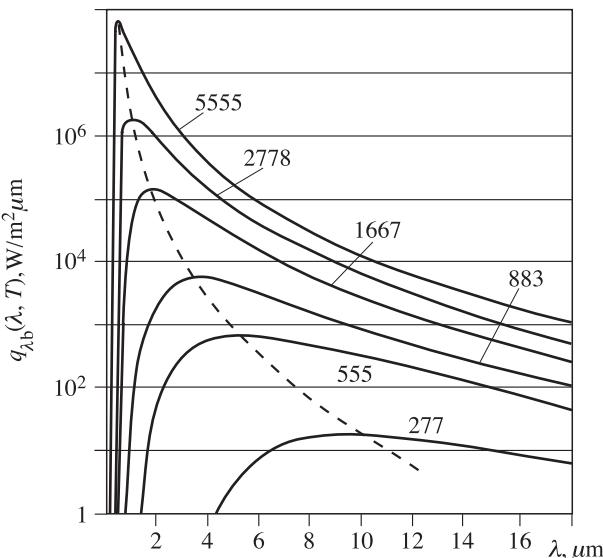


Figure 6.2. Hemispherical spectral radiation flux of black bodies for some values of temperatures versus wavelengths. Black-body temperatures are shown by figures next to the curves. Positions of spectral radiation flux maxima are shown by a dotted line.

wavelength dependencies of hemispherical spectral surface density of radiation flux for several values of absolute temperature. A peculiarity of Planck's curves is the increase of the energy of radiation, corresponding to all wavelengths, with increasing temperature. As was shown in section 6.1, qualitative thermodynamic considerations and everyday experience indicate that the energy of total radiation (including all wavelengths) should increase with temperature. It also follows from Figure 6.2 that this conclusion is also valid for the energy of radiation corresponding to each wavelength. Another peculiarity is the displacement of maxima of the spectral surface density of radiation flux to the side of shorter wavelengths with increasing temperature. The cross-sections of the plot in Figure 6.2 at fixed wavelengths, which determine the radiation energy as a function of temperature, allow us to state that the energy of radiation, emitted at the short-wave extremity of the spectrum, increases with temperature faster than the energy of radiation corresponding to greater wavelengths. Figure 6.2 indicates the position of the wavelength band in the visible spectrum region. For a body at temperature of 555 K only a very small fraction of energy falls on the visible spectrum range, which is virtually imperceptible by the human eye. Since the curves at lower temperatures are dropping from the red section toward the violet extremity of the spectrum, then, at first, the red light becomes visible with increasing temperature (the so-called Driper point, corresponding to 525°C). At sufficiently high temperature the emitted light becomes white and consists of a set of all wavelengths of the visible spectrum. The radiation spectrum of the Sun is similar to the radiation spectrum of a black body at a temperature of 5800 K, and a considerable portion of released energy falls on the visible spectrum range. (This type of radiation is sometimes called 'white' noise – as we see, quite wrongly.) More likely, owing to very long biological evolution, the human eye became most sensitive precisely in the spectrum region with maximum energy.

Equation (6.6) can be presented in a more convenient form that allows us to avoid constructing the curves for each value of temperature; for this purpose equation (6.6) is divided by temperature to the fifth power:

$$(q_\lambda(T, \lambda)/T^5) = \frac{\pi I_\nu(T, \lambda)}{T^5} = \frac{C_1}{(\lambda T)^5} \frac{1}{[\exp(C_2/\lambda T) - 1]}. \quad (6.8)$$

This equation determines quantity $q_\lambda(T, \lambda)/T^5$ as a function of single variable λT . The plot of such a dependence is presented in Figure 6.3; it substitutes a set of curves in Figure 6.2.

The Planck law for energy distribution in the black-body spectrum gives a maximum value of the intensity of radiation that can be emitted by any body at the given temperature and wavelength. This intensity plays a part of an optimum standard, with which the characteristics of real surfaces can be compared.

But, more simply, approximate forms of the Planck law are sometimes applied. However, it is necessary to bear in mind that they can be used only in that range, where they provide acceptable accuracy.

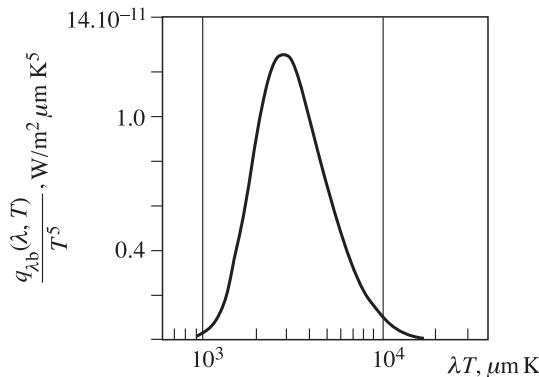


Figure 6.3. Hemispherical of black-body spectral radiation flux distribution versus generalized coordinates.

6.2.2 The Wien radiation law

If the term $\exp(C_2/\lambda T) > 1$, then equation (6.8) is reduced to the expression

$$\frac{I_\lambda(T, \lambda)}{T^5} = \frac{C_1}{\pi(\lambda T)^5 \exp(C_2/\lambda T)}, \quad (6.9)$$

which is known as the Wien radiation law. For the values of $\lambda T < 3000 \mu\text{m K}$ this formula gives an error within the limits of 1%.

6.2.3 The Rayleigh–Jeans radiation law

Another approximate expression can be obtained by expanding the denominator in equation (6.8) into the Taylor series. If λT is essentially greater than C_2 , then the series can be restricted by the second term of expansion, and equation (6.8) takes the form:

$$\frac{I_\lambda(T, \lambda)}{T^5} = \frac{C_1}{\pi C_2} \frac{1}{(\lambda T)^4}. \quad (6.10)$$

This equation is known as the Rayleigh–Jeans radiation law. This formula gives an error within the limits of 1% for the values of $\lambda T > 7.8 \times 10^5 \mu\text{m K}$. These values are outside the range usually considered in IR thermal radiation, but they are of principal importance for the radio-frequency band. The frequency presentation of the Planck formula is usually applied in this band, and then the Rayleigh–Jeans law takes the widely used form:

$$I_\nu(T, \nu) = \frac{2\nu^2}{c_0^2} nkT = \frac{2f^2}{c_0^2} nkT. \quad (6.11)$$

6.2.4 The Wien displacement law

Another quantity of interest, which relates to the black-body radiation spectrum, is the wavelength λ_m , to which corresponds the maximum of surface density of an emitted energy flux. As is shown by the dotted curve in Figure 6.2, this maximum displaces to the side of shorter wavelengths as the temperature increases. Quantity λ_m can be found by differentiating the Planck function from equation (2.12) and by equating the obtained expression to zero. As a result, the transcendental equation is obtained

$$\lambda_m T = \frac{C_2}{5} \frac{1}{1 - \exp(-C_2/\lambda_m T)}, \quad (6.12)$$

whose solution is as follows:

$$\lambda_m T = C_3 \quad (6.13)$$

and represents one of expressions of the Wien displacement law. The values of constant C_3 are given in Table A.4. According to equation (6.13), as the temperature increases, the maxima of surface density of the radiation flux and its intensity displace to the side of shorter wavelengths in inverse proportion to T . If we consider the black-body radiation into a transparent medium (with refractive index n), then the Wien law takes the form of

$$n\lambda_{m,n} T = C_3, \quad (6.14)$$

where $\lambda_{m,n}$ is the wavelength corresponding to the maximum of radiation in the transparent medium.

Of interest is the fact that the substitution of the wavelength from the Wien displacement law (6.13) into equation (6.8) results in the following expression:

$$I_\lambda(T, \lambda_m) = T^5 \frac{C_1}{\pi C_3^5 [\exp(C_2/C_3) - 1]}. \quad (6.15)$$

It follows from this relation, that the maximum value of radiation intensity increases in proportion to temperature to the fifth power. Generally speaking, it is this relation that was obtained by Wien in 1893.

It can easily be seen from the expression obtained that the maximum of spectral intensity of the microwave background of the universe at radiation temperature of 2.73 K will be approximately equal to 1 mm.

6.2.5 The Stefan–Boltzmann law

Integrating $q_\lambda(T)$ over all wavelengths from zero to infinity (or, accordingly, $q_\nu(T)$ in the frequency presentation), we obtain by means of expressions for determinate integrals (Gradshteyn and Ryzhik, 2000) the surface density of the total black-body radiation flux $q(T)$ as:

$$q(T) = \int_0^\infty q_\lambda(T, \lambda) d\lambda = \int_0^\infty q_\nu(T, \nu) d\nu = \pi \int_0^\infty I_\nu(T, \nu) d\nu = n^2 \sigma T^4, \quad (6.16)$$

where the Stefan–Boltzmann constant σ is equal (see Table A.4) to:

$$\sigma = \frac{2\pi^5 k^4}{15c_0^2 h^3}. \quad (6.17)$$

Similar expressions can also be obtained for the total radiation intensity:

$$I(T) = \int_0^\infty I_\nu(T, \nu) d\nu = n^2 \frac{\sigma}{\pi} T^4 \quad (6.18)$$

and for the total volume density of radiation (for vacuum):

$$u = \int_0^\infty u_\nu(T, \nu) d\nu = a T^4, \quad (6.19)$$

where a is called the radiation density constant (see Table A.4).

Let us consider now the instructive example, associated with the relation of the amount of energy, emitted from the unit of black body's surface into vacuum within the whole frequency band and in the radio-frequency band separately. Using relations (6.16), we obtain the total power, emitted by a black body from 1 square metre at room temperature (300 K), which is equal to 450 W. Now, using the Rayleigh–Jeans law (6.11), we obtain the expression for the Stefan–Boltzmann law in the long-wavelength approximation, as follows:

$$q(T) = \frac{2}{3} \frac{\pi k}{c_0^2} T \nu^3. \quad (6.20)$$

From this expression we can easily obtain the estimate for the total power emitted by a black body from 1 square metre at $T = 300$ K throughout the radio-frequency band from zero frequency up to 10^{11} Hz (the wavelength is 3 mm). It is equal to 10^{-4} W. Thus, the amount of energy falling on the whole radio-frequency band is 10^{-7} times lower than the total power of black-body radiation. In this case an even smaller part (10^{-9}) of the total power will fall on the whole, for example, centimetre band. And, in spite of such small values of radiation power in the radio-frequency band, modern microwave remote radio systems successfully record such low levels of a thermal signal (see Chapter 3).

6.2.6 Correlation properties of black-body radiation

From the viewpoint of the theory of random processes (Chapter 2), the spectral volume density of radiation energy $u_\nu(\nu)$ represents a spectral density of fluctuating strengths $E(t)$ and $H(t)$ of the thermal radiation field. This can easily be seen, taking into consideration relations (1.17), (2.27), (5.13). In each of the planar waves into which this field can be decomposed, the relation between the vectors of running planar waves is given by expression (1.11), all directions of strengths being equiprobable. As a result of small transformations in (1.17), it can be seen that the electric and magnetic energies are equal, and E and H components in any arbitrary direction have identical correlation functions but are not correlated among themselves. Thus,

the correlation properties can be considered with respect to any component of the electromagnetic field strength.

Let us find the correlation coefficient corresponding to the spectral density (6.1), i.e. the quantity

$$R_u(\tau) = \frac{B_u(\tau)}{B_u(0)}, \quad (6.21)$$

where

$$B_u(\tau) = \frac{2}{\pi} \int_0^\infty u_\nu(T, \nu) \cos 2\pi\nu d\nu \quad (6.22)$$

Substituting here expression (6.1) for the spectral density and calculating the integral, we obtain (Rytov, 1966):

$$R_u(\tau) = 15 \left[\frac{3}{\operatorname{sh}^4 \beta} - \frac{3}{\beta^3} + \frac{2}{\operatorname{sh}^2 \beta} \right] = \frac{15}{2} \frac{d^3}{d\beta^3} L(\beta), \quad (6.23)$$

where $L(\beta) = \operatorname{cth} \beta - (1/\beta)$ is the Langevin function, and $\beta = 2\pi^2 k T \tau / h$ ($\operatorname{sh} x$, $\operatorname{cth} x$ are hyperbolic sine and cotangent).

The form of the correlation coefficient from the temporary lag is shown in Figure 6.4 and, as should be expected, it certainly does not look like the delta-function. First of all, we note that for $\beta \cong 1.37$ (which corresponds to $\tau_0 = h/2\pi k T$) the positive correlation is changed to a negative one. This implies that for temporary shifts $\tau < \tau_0$ the values of component $E_p(t)$ in some fixed direction p will more frequently have at instants t and $t + \tau$ the same sign, and for $\tau > \tau_0$ the opposite sign. The temporary lag τ_0 can be put in correspondence to the spatial correlation radius $\lambda_0 = c\tau_0$, which to an accuracy of numerical coefficient coincides with the wavelength $\lambda_m = 0.2(hc/kT)$ in the Wien displacement law. From the comparison of these expressions we can obtain the following important relation:

$$\lambda_0 = 0.35\lambda_m. \quad (6.24)$$

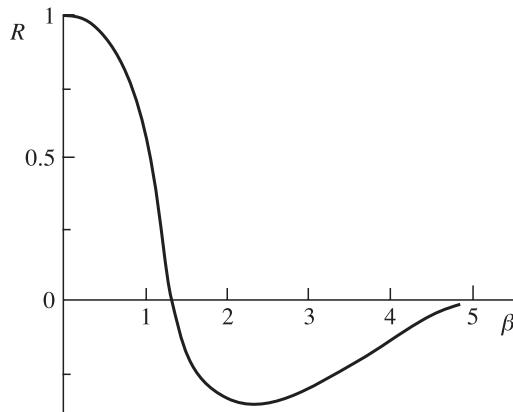


Figure 6.4. Correlation coefficient of spectral black-body radiant energy density versus generalized lag $\beta = 2\pi^2 \frac{k T \tau}{h}$.

It can easily be found from this relation that the spatial correlation radius of the microwave background of the universe equals a quite macroscopic value, namely, $\lambda_0 = 0.35$ mm. It is interesting to mention, that earlier (in 1971) it was proposed to measure the velocity of solar system motion relative to the microwave ('absolute ether at rest') background by recording the variable part of the interferogram (in other words, the correlation coefficient (6.24)) just at the place where it changes its sign (near λ_0) (Soglasnova and Sholomitskii, 1971).

At spatial distances greater than $(4\text{--}5)\lambda_0$ the correlation sharply drops, and the statistical process of emission at such scales can be represented as a non-correlated random (white) noise. Generally speaking, it is this circumstance which is often used in analysing thermal radiation.

6.3 THE KIRCHHOFF LAW

As we have noted above (Chapter 4), the fluctuation–dissipation theorem, which represents one of the fundamental laws of statistical physics, establishes for an arbitrary dissipative physical system the relationship between the spectral density of spontaneous equilibrium fluctuations and its nonequilibrium properties and, in particular, the energy dissipation in a system. For the wave field of an absorbing half-space, i.e. for the field of radiation which can be recorded by an external (relative to the emitting medium) instrument, the solution of the fluctuation electrodynamic problem directly results in the Kirchhoff law in the form of (4.20).

Before describing the properties of non-black physical bodies, it is useful to introduce the definitions of emissive ability and absorbing ability and also to consider the Kirchhoff law forms that are often used for analysing the emitting half-space (i.e. when there are two material media with a sharp boundary between them), as well as for analysing the radiation transfer processes in a transparent infinite medium (the atmosphere). In the first case (the planar version) the measuring instrument is inside one of media and measures the radiation of the other one. In the second case (the solid version) the instrument is directly inside the medium, whose radiation it just measures. Below we shall consider the first version in detail. As far as the solid (three-dimensional) version is concerned, we shall postpone the detailed study of radiation transfer processes for this version until Chapter 9.

6.3.1 Emissive ability

This characteristic, which is sometimes called the emissivity, indicates what portion of black-body radiation energy constitutes the radiation energy of a given body. The emissive ability of a real physical body depends on such factors as its temperature, its physical and chemical composition, its intrinsic geometrical structure, its degree of surface roughness, the wavelength to which the emitted radiation corresponds, and the angle at which the radiation is emitted. For remote microwave sensing problems it is necessary to know the emissive ability both in any required direction (the

angular characteristics) and at various wavelengths (the spectral characteristics). In this case the degree of remote information capacity of angular and spectral characteristics is strongly distinguished, generally speaking, depending on the type of a physical object under study. This radiation characteristic is called the directional emissive ability (or the directional emissivity).

In calculating a body's total energy losses through radiation (as in heat-and-power engineering problems) it is necessary to know the radiation energy in all directions and, for this reason, the emissivity, averaged over all directions and wavelengths, is used in such calculations. For calculating a complicated heat exchange through radiation between surfaces, the emissivities can be required, which are averaged only over the wavelengths and not over the directions. So, the researcher should possess the emissivity values, averaged in different ways, and they should be obtained, most frequently, from the available experimental data.

In this book we shall keep to the definition of directional spectral emissivity. If necessary, this emissivity can then be averaged over the wavelengths and the directions, and, finally, over wavelengths and directions simultaneously. Averaged over the wavelengths, they are called total (integral) quantities, and the quantities averaged over the directions are called hemispherical quantities (Siegel and Howell, 1972).

Recalling the definitions of spectral intensity of emission from the unit of a physical body's surface (see section 5.1), we shall define the directional emissivity as the ratio of the spectral intensity of a real surface $I_\nu(\mathbf{r}, \Omega, T, \dots)$, which depends on body's temperature, physical and chemical composition, intrinsic geometrical structure and degree of surface roughness, as well as on the observation angle and working wavelength (frequency), to the black-body intensity $I_{\nu B}(\nu, T)$ at the same temperature and at the same wavelength (frequency) (6.2):

$$\kappa_\nu(\mathbf{r}, T, \nu, \Omega, \dots) = \frac{I_\nu(\mathbf{r}, T, \nu, \Omega, \dots)}{I_{\nu B}(T, \nu)} \quad (6.25)$$

This expression for emissivity is most general, since it includes the dependencies on the wavelength, direction and temperature. The total and hemispherical characteristics can be obtained by appropriate integration (Siegel and Howell, 1972).

As far as the volumetric version is concerned, here we should note that the directional spectral emissivity of a unit homogeneous volume of the medium is equal to the ratio of intensity of radiation, emitted by this volume in the given direction, to the intensity of radiation emitted by a black body at the same temperature and wavelength.

6.3.2 Absorbing ability

The absorbing ability of a body is the ratio of the radiation flux absorbed by the body to the radiation flux falling (incident) on the body. The incident radiation possesses the properties inherent in a particular power source. The spectral distribution of the incident radiation energy does not depend on temperature or on the physical nature of an absorbing surface (so long as the radiation, emitted by the

surface, is not partially reflected back onto this surface). In this connection, in defining the absorbing ability (as compared to emissivity), additional difficulties arise, which are related to the necessity of taking into account the directional and spectral characteristics of incident radiation.

By the directional spectral absorptivity $\alpha(\mathbf{r}, \Omega, T, \dots)$ we shall mean the ratio of the spectral intensity of absorbed radiation $I_{\nu a}(\mathbf{r}, \Omega, \nu, T, \dots)$ to the spectral intensity of incident radiation at the given wavelength and from the given direction $I_{\nu 0}(\mathbf{r}, \Omega, \nu, T, \dots)$:

$$\alpha_\nu(\mathbf{r}, T, \nu, \Omega, \dots) = \frac{I_{\nu a}(\mathbf{r}, T, \nu, \Omega, \dots)}{I_{\nu 0}(\mathbf{r}, T, \nu, \Omega, \dots)}. \quad (6.26)$$

In addition to the incident radiation dependence on the wavelength and direction, the directional spectral absorptivity is also a function of temperature, physical and physico-chemical properties of an absorbing surface.

6.3.3 The Kirchhoff law forms

This law establishes the relation between the abilities of emitting and absorbing the electromagnetic energy by any physical body. This law can be presented, to an equal degree of certainty, in terms of spectral, integral, directional or hemispherical quantities. In the case of microwave sensing it is expedient for us to dwell on the directional properties. From equations (5.1) and (6.25), the energy of radiation, emitted by a surface element from dA in the frequency band $d\nu$, within the limits of solid angle $d\Omega$ and during time dt , is equal to

$$dE_\nu = \kappa_\nu(\mathbf{r}, T, \nu, \Omega, \dots) I_{\nu B}(T, \nu) dA \cos \theta d\Omega d\nu dt. \quad (6.27)$$

If we assume the element dA at temperature T to be inside the isothermal, ideally black, closed cavity, also at temperature T , then the intensity of radiation, falling on the element dA in the direction Ω , will be equal to $I_{\nu B}(T, \nu)$ (remember the property of isotropy of radiation intensity of an ideal black cavity) (section 6.1). For maintaining the isotropy of radiation inside an ideal black closed cavity, the fluxes of absorbed and emitted radiation, determined by equations (6.26) and (6.27), should be equal and, therefore, the following relation should be met:

$$\kappa_\nu(\mathbf{r}, T, \nu, \Omega, \dots) = \alpha_\nu(\mathbf{r}, T, \nu, \Omega, \dots). \quad (6.28)$$

Equality (6.28) sets the relationship between the fundamental properties of physical substances and is valid, without limitations, for all media in a state of thermodynamic equilibrium. It represents the most general form of the Kirchhoff law. It is just this form of the law that was presented by G. Kirchhoff in his famous work published in 1860 (see Schopf (1978) for more details).

The following important corollary follows from (6.28). Since in its physical sense quantity α is always less than unity, the emissive ability of any physical body is concluded between zero and unity, i.e. $0 < \kappa < 1$. This characteristic is used very widely in microwave sensing, since it allows us to estimate and compare the emission

properties of investigated substances without resorting to the measurement of radiation energy values.

Another formulation of Kirchhoff's law, also set forth by him, is also possible. The ratio of the radiation intensity of a physical body, heated up to temperature T , to its absorptivity is a universal function of temperature and frequency, which does not depend on the physical and geometrical properties of a body. Proceeding from (6.25), (6.26) and (6.28), we have:

$$\frac{I_\nu(\mathbf{r}, T, \nu, \Omega, \dots)}{\alpha_\nu(\mathbf{r}, T, \nu, \Omega, \dots)} = I_{\nu B}(T, \nu). \quad (6.29)$$

Kirchhoff himself considered the finding of an explicit form of this universal function to be 'the problem of fundamental importance' for physics (Schopf, 1978).

One further form of Kirchhoff's law is used in microwave sensing (and we shall use it later). It follows from the relations presented above:

$$I_\nu(\mathbf{r}, T, \nu, \Omega, \dots) = \kappa(\mathbf{r}, T, \nu, \Omega, \dots) I_{\nu B}(T, \nu). \quad (6.30)$$

As we have seen, in section 1.4, electromagnetic waves propagate in free space, where there exist two components of a wave, which oscillate at right angles with respect to each other and with respect to the wave propagation direction. In the particular case of equilibrium thermal radiation these two components of polarization are equal. Strictly speaking, relations (6.28)–(6.31) are fulfilled for each polarization component, and, in order that they be valid for the total incident radiation, the radiation should have equal polarization components. Thus, the original equilibrium radiation is non-polarized (which, however, is invalid for grey bodies) (see Chapter 7).

The Kirchhoff law was proved for the case of thermodynamic equilibrium in an isothermal closed cavity and, hence, it is strictly valid only in the absence of a resulting thermal flux directed towards the surface or away from it. Under real conditions, as a rule, there exists a resulting flux of electromagnetic radiation, so that relations (6.28) and (6.30) are approximate, strictly speaking. The validity of this approximation is confirmed by reliable experimental data, according to which in the majority of practical cases the ambient radiation field does not have any significant influence on the values of emissive and absorbing abilities. Another confirmation of this approximation is a substance's ability to be at the state of local thermodynamic equilibrium (section 4.4), in which the set of energy states during absorption and emission processes corresponds, to a very close approximation, to their equilibrium distributions (corresponding to the local temperature in this case). Thus, the spreading of Kirchhoff's law to natural nonequilibrium systems is not the result of simple thermodynamic considerations, but, most likely, it is a result of the physical nature of substances. Owing to this circumstance, in the majority of cases the substance is capable of independently maintaining a local thermodynamic equilibrium and, thus, to possesses 'independence' of its properties from the ambient radiation field.

In conclusion, we note that, as astrophysical investigations have shown, the Kirchhoff law can actually also be applied in cases where the radiation is not in

full equilibrium with the substance, and its distribution over frequencies essentially differs from Planck's one. However, the Kirchhoff law is not applicable in cases where thermodynamic equilibrium conditions are strongly violated (nuclear explosions, shock waves, the interplanetary medium). This law is not suitable for determining the emissivities of sources of non-thermal radiation (synchrotron, maser radiation, thunderstorm activity) and sources of quasi-deterministic radiation (radio- and TV-broadcasting, communications).

00.blackbody_radiation_28-Dec-2020

Engineering Physics

PHY 1701

Module-1

M. Ummal Momeen

School of Advanced Sciences,
VIT University, Vellore Campus.

Email: ummal.momeen@vit.ac.in

Office: SJT Annexe G01A13

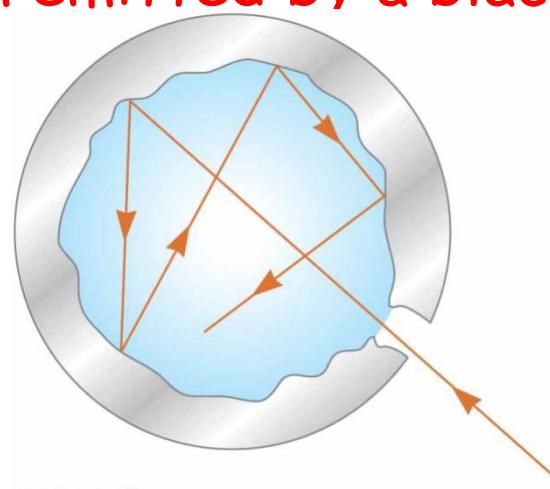
Course plan for this Lecture

Introduction of particle properties of waves:
QUALITATIVE ideas about blackbody radiation
(blackbody spectrum); Ultraviolet catastrophe

Introduction of Planck's hypothesis to resolve the UV catastrophe.

What is black body radiation?

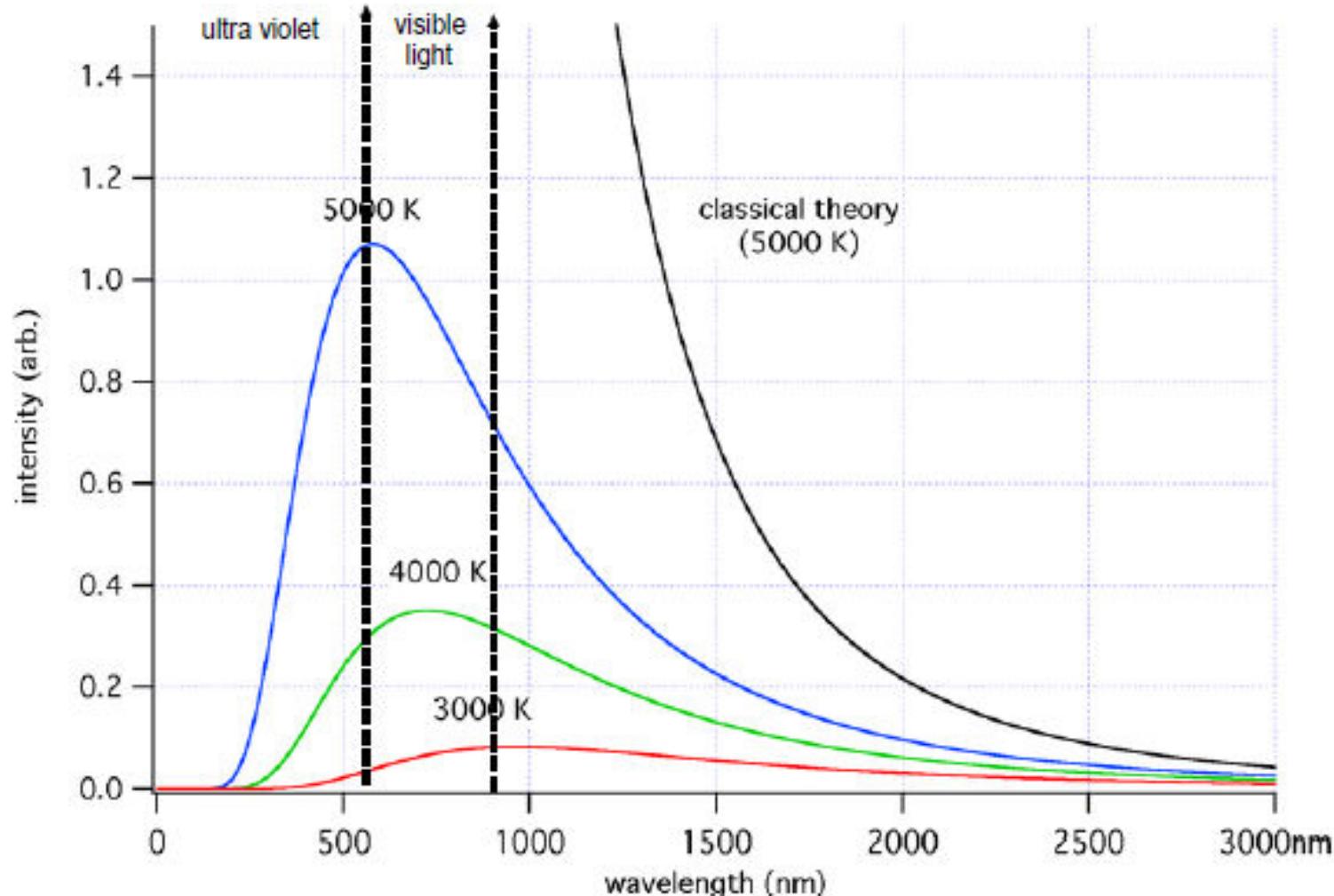
- One that absorbs all radiation incident upon it, regardless of frequency. Such a body is called a black body.
- A **black body** is an ideal system that absorbs all radiation incident on it.
- The electromagnetic radiation emitted by a black body is called **blackbody radiation**



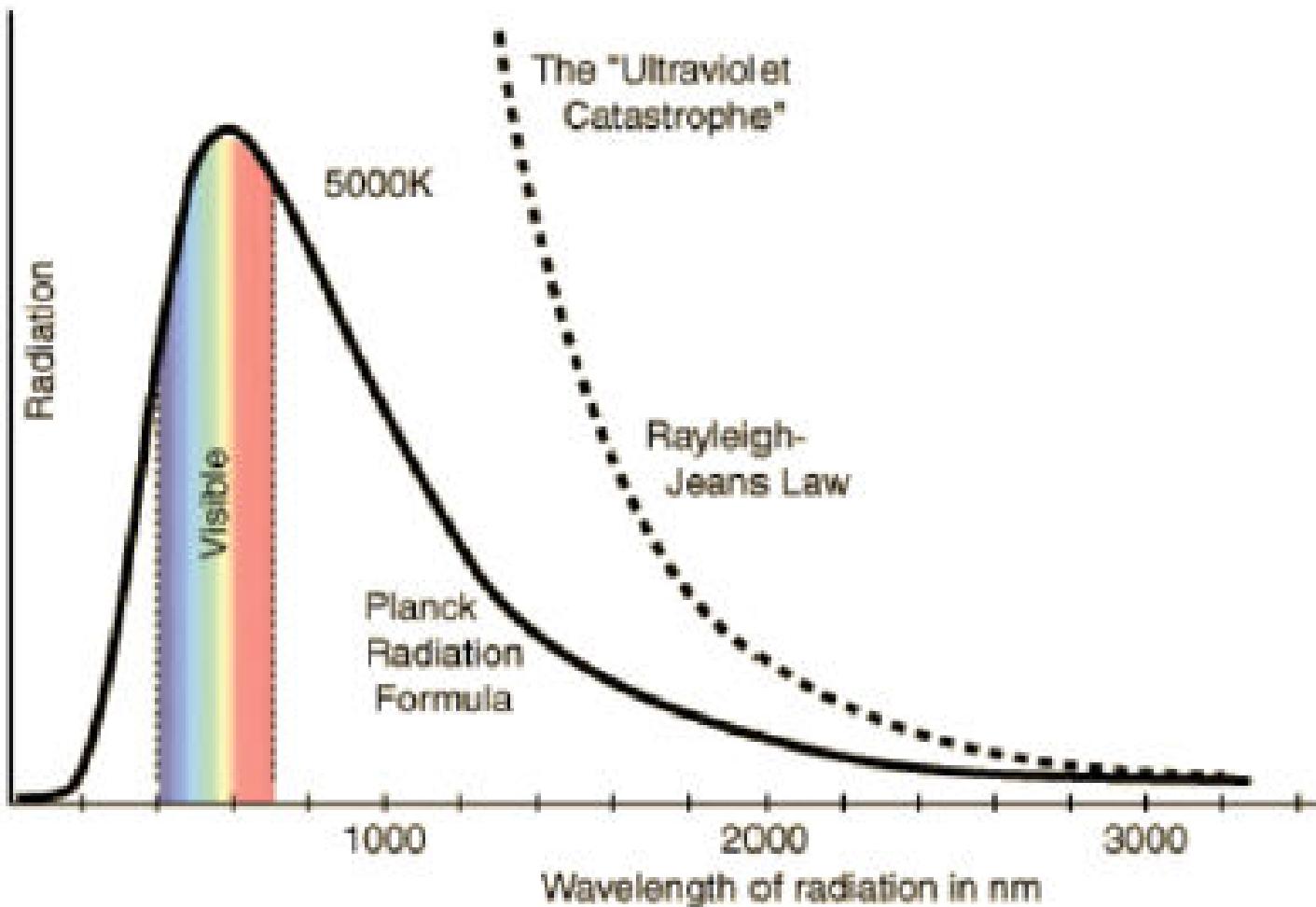
Example: Black Holes

- A good approximation of a black body is a small hole leading to the inside of a hollow object
- The hole acts as a perfect absorber
- The nature of the radiation leaving the cavity through the hole depends only on the temperature of the cavity

Black body spectrum



As the temperature decreases, the peak of the black-body radiation curve moves to lower intensities and longer wavelengths. The black-body radiation graph is also compared with the classical model



Classical approach of explaining black body radiation-1

Wien's displacement law

$$\lambda_{\max} T = 2.898 \times 10^{-3} \text{ m}\cdot\text{K}$$

λ_{\max} is the wavelength at which the curve peaks

T is the absolute temperature

The wavelength is inversely proportional to the absolute temperature

As the temperature increases, the peak is "displaced" to shorter wavelengths

Wien's law works well only for short wavelengths

Classical approach of explaining black body radiation-2

Rayleigh jeans law

An early classical attempt to explain blackbody radiation was the **Rayleigh-Jeans law**

At long wavelengths, the law matched experimental results fairly well

At short wavelengths, there was a major disagreement between the Rayleigh-Jeans law and experiment

This mismatch known as the ultraviolet catastrophe; You would have infinite energy as the wavelength approaches zero

It is explained by considering the radiation inside a cavity of absolute temperature T whose walls are perfect reflectors to be a series of standing em waves.

This is a three-dimensional generalization of standing waves in a stretched string.

The number of independent standing waves $G(\nu)d\nu$

In the frequency interval between ν and $d\nu$ per unit volume in the cavity turned out to be density of standing waves in cavity

$$G(\nu)d\nu = \frac{8\pi\nu^2d\nu}{c^3}$$

According to the theorem of equipartition of energy, the average energy per degree of freedom of an entity

$$\frac{1}{2}K_B T$$

Classical average energy per standing wave

$$\bar{\varepsilon} = K_B T$$

The total energy $U(\nu)d\nu$ per unit volume in the cavity in the frequency interval from ν to $\nu+d\nu$ is therefore

$$U(\nu)d\nu = \bar{\varepsilon}G(\nu)d\nu$$

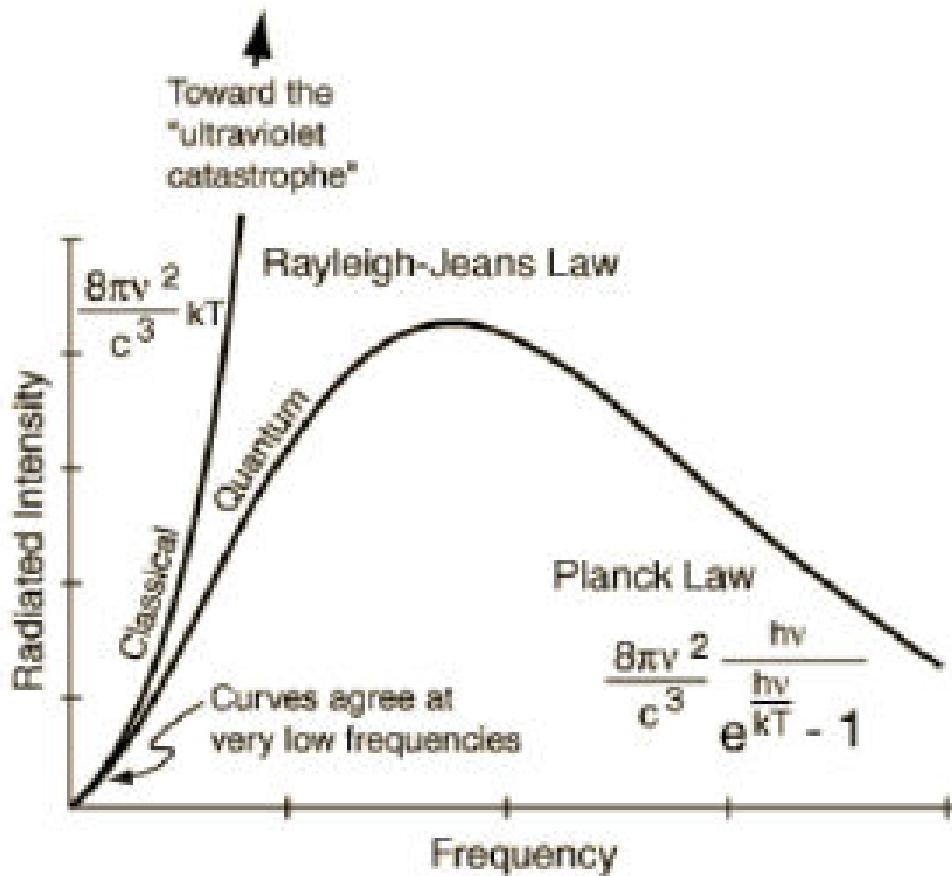
$$U(\nu)d\nu = \frac{8\pi kT}{c^3} \nu^2 d\nu$$

This is called Rayleigh Jeans formula.

It contains everything that classical physics can say about the spectrum of black body Radiation

Energy density is proportional to ν^2 ; at high frequencies $U(\nu)d\nu \rightarrow \infty$

In the black body spectrum $U(\nu)d\nu \rightarrow 0$ and $\nu \rightarrow 0$



The Rayleigh-Jeans curve agrees with the Planck radiation formula for long wavelengths, low frequencies

Planck's Hypothesis

Molecules can only have discrete values of energy E_n , given by:

$$E_n = nh\nu$$

where n is a positive integer called a quantum number and ν is the natural frequency of oscillation of the molecules.

This is quite different from the classical model of the harmonic oscillator, in which energy of identical oscillators is related to the amplitude of the motion and unrelated to the frequency.

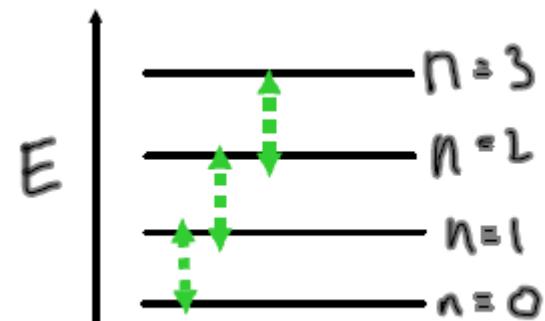
Because the energy of a molecule can only have discrete values, we say the energy is quantized. Each discrete energy level represents a specific quantum state.

The molecules emit or absorb energy in discrete packets that later became known as photons. The molecules emit or absorb photons by 'jumping' from one quantum state to another.

If the jump is downward from state to an adjacent lower state, the amount of energy radiated by the molecule in a single photon is $\hbar\nu$. A molecule emits or absorbs energy only when it changes quantum states.

The RayleighJeans curve agrees with the Planck radiation formula for longer wavelengths, (low frequencies)

Planck's constant $h = 6.626 \times 10^{-34} \text{ J s}$



Average energy of an oscillator whose frequency of vibration is ν .

$$\bar{E} = \frac{h\nu}{(e^{h\nu/kT} - 1)}$$

Energy radiated per unit volume in the frequency interval ν and $\nu + d\nu$

$$U(\nu) d\nu = \frac{8\pi h \nu^3}{c^3} \frac{d\nu}{e^{h\nu/k_B T} - 1}$$

This is called Planck's radiation formula.

h - Planck's constant

c - velocity of light

k_B - Boltzmann constant 1.38×10^{-23} J/K

T- Absolute temperature

At high frequencies $h\nu >> kT$ and $e^{h\nu/kT} \rightarrow \infty$

Which means that $U(\nu)d\nu \rightarrow 0$ as observed.

At low frequencies

$$\frac{h\nu}{kT} \ll 1$$

$$e^x = 1 + x + \frac{x^2}{2!} + \dots$$

If x is small

$$e^x = 1 + x$$

$$\frac{1}{e^{h\nu/kT} - 1} = \frac{1}{1 + \frac{h\nu}{kT} - 1} = \frac{kT}{h\nu}$$

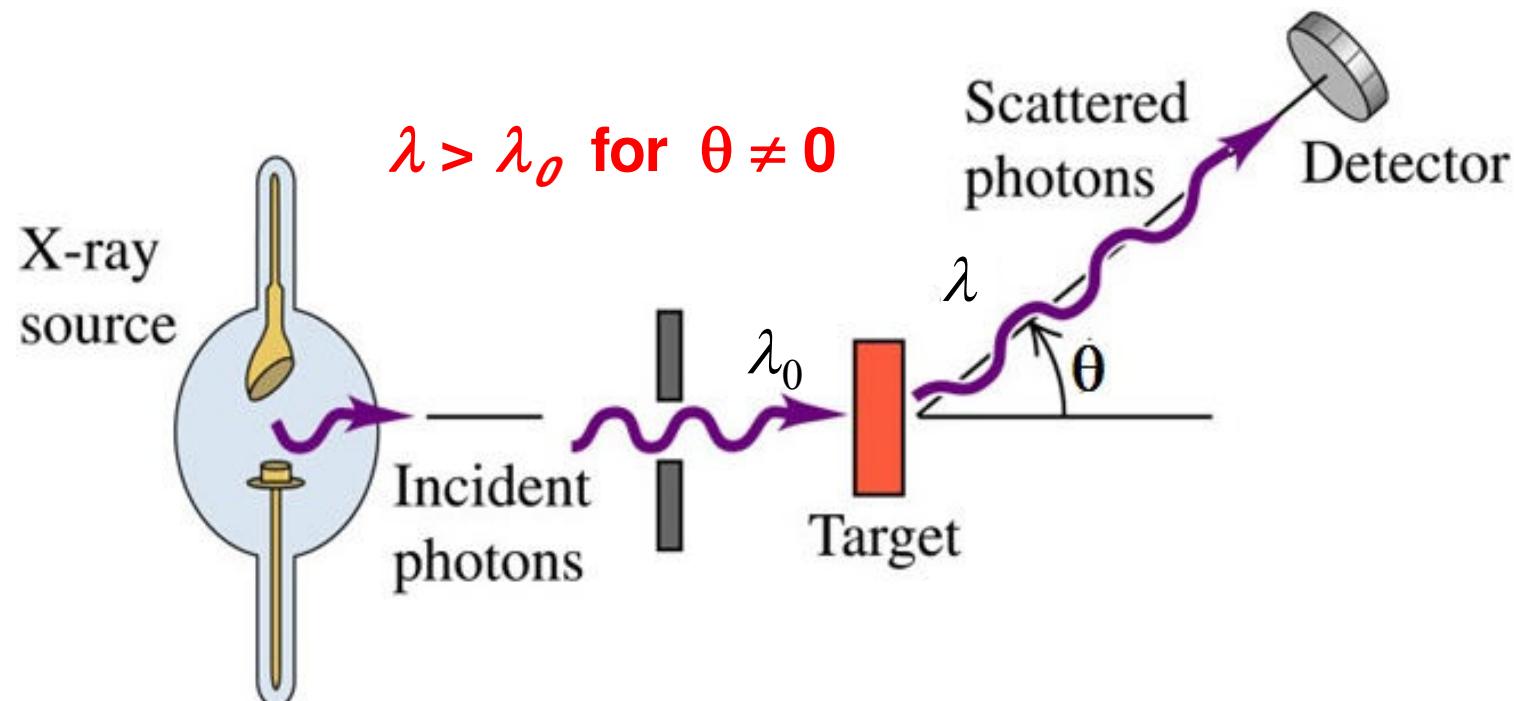
At low frequencies Planck's formula becomes

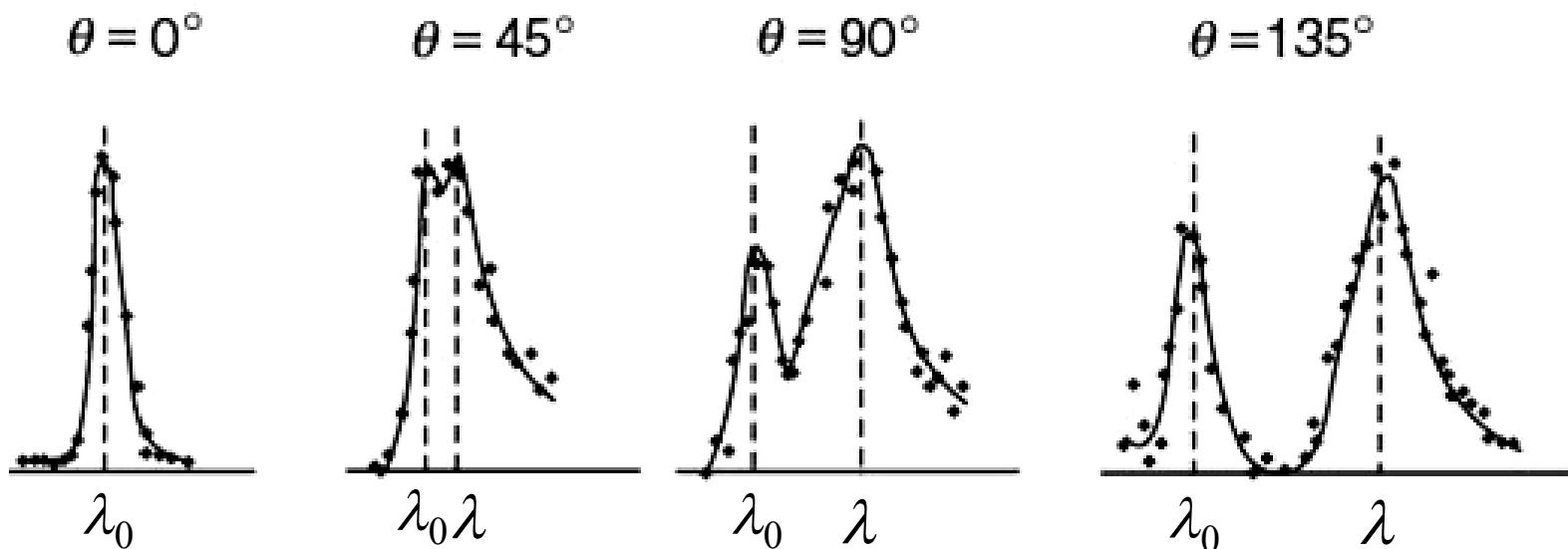
$$\begin{aligned} U(v)dv &= \frac{8\pi h v^3}{c^3} dv \left(\frac{kT}{hv} \right) \\ &= \frac{8\pi k T v^2}{c^3} dv \end{aligned}$$

which is Rayleigh Jeans formula.

Compton Effect

Compton discovered that when a beam of X-rays is scattered from a target, the wavelengths of the scattered X-rays are slightly greater than the wavelength of incident beam.



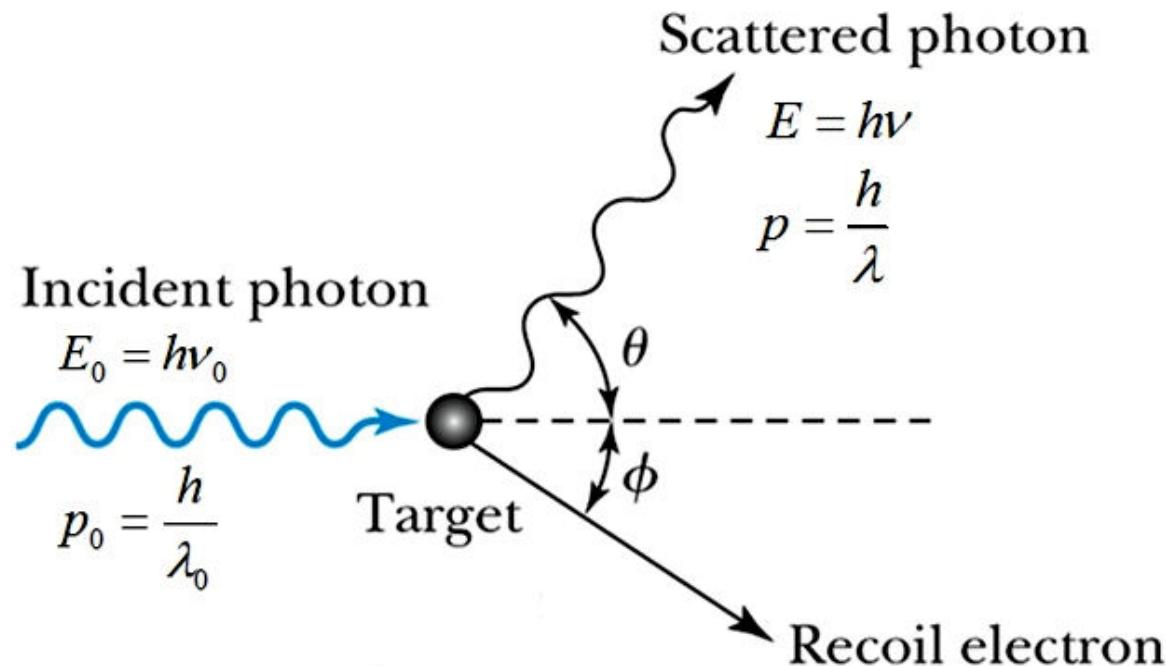


Compton experimental results for four different values of the scattering angle θ

- The peak at λ_0 is due to the scattering of the incident radiation from the tightly bound inner electrons of the atom.
- The second peak represents the radiation scattered from the loosely bound, nearly free outer electrons

$$\text{Compton Shift} \quad \Delta\lambda = \lambda - \lambda_0 = \frac{h}{m_e c} (1 - \cos \theta)$$

- The scattering of X-rays by electrons is treated as the collision between photons and electrons.
- The electrons which are loosely bound can be assumed as almost free particles at rest.



Before Collision

Energy of incident photons $E_0 = h\nu_0$

Linear momentum of incident photons $p_0 = \frac{E_0}{c} = \frac{h\nu_0}{c}$

Relativistic relation between energy and momentum $E = \sqrt{p^2 c^2 + m_0^2 c^4}$
 Rest mass of photon, $m_0 = 0$. So $E_0 = p_0 c$

Energy of electrons = $m_e c^2$ (Free electron at rest... No momentum)

After Collision

Energy of scattered photons $E = h\nu$

Linear momentum of scattered photons $p = \frac{h\nu}{c}$

Energy of electrons $E_e = \sqrt{p_e^2 c^2 + m_e^2 c^4}$

Linear momentum of recoiled electron = p_e

Conservation of momentum

$$\vec{p}_0 + 0 = \vec{p} + \vec{p}_e \quad \text{or} \quad \vec{p}_e = \vec{p}_0 - \vec{p}$$

$$p_e^2 = (\vec{p}_0 - \vec{p}) \cdot (\vec{p}_0 - \vec{p}) = p_0^2 + p^2 - 2 \vec{p}_0 \bullet \vec{p}$$

$$p_e^2 = p_0^2 + p^2 - 2 p_0 p \cos \theta$$

Conservation of Energy

$$E_0 + m_e c^2 = E + \sqrt{p_e^2 c^2 + m_e^2 c^4}$$

$$[(E_0 - E) + m_e c^2] = \sqrt{p_e^2 c^2 + m_e^2 c^4}$$

$$[(E_0 - E) + m_e c^2]^2 = p_e^2 c^2 + m_e^2 c^4$$

$$(E_0 - E)^2 + 2(E_0 - E)m_e c^2 = p_e^2 c^2$$

$$(E_0 - E)^2 + 2(E_0 - E)m_e c^2 = p_0^2 c^2 + p^2 c^2 - 2p_0 p c^2 \cos \theta$$

$$= E_0^2 + E^2 - 2E_0 E \cos \theta$$

~~$$(E_0 - E)^2 + 2(E_0 - E)m_e c^2 = E_0^2 + E^2 - 2E_0 E + 2E_0 E - 2E_0 E \cos \theta$$~~

$$2(E_0 - E)m_e c^2 = 2E_0 E(1 - \cos \theta)$$

$$\left(\frac{1}{E} - \frac{1}{E_0} \right) = \frac{1}{m_e c^2} (1 - \cos \theta)$$

$$\left(\frac{1}{\nu} - \frac{1}{\nu_0} \right) = \frac{h}{m_e c^2} (1 - \cos \theta)$$

$$\lambda - \lambda_0 = \frac{h}{m_e c} (1 - \cos \theta)$$

Compton wavelength

$$\lambda_c = \frac{h}{m_e c} = 0.00243 \text{ nm}$$

Compton Scattering

The photoelectric effect and Einstein's theories about light having a particle nature caused a lot of scientists to start to reexamine some basic ideas, as well as come up with some new ones

- Based on a lot of Einstein's work (including his Special Theory of Relativity), physicists predicted that these photons should have momentum, just like other particles do.
- The momentum that the light photons have must be very small, and not based on the common way of calculating momentum using $p = mv$ (since light has no rest mass).
- Instead the formula was based on the wavelength and frequency of the light, just like Planck's formula.

$$p = \frac{h}{\lambda} \quad \text{or} \quad p = \frac{h}{c}$$

p = momentum (kgm/s)

h = Planck's Constant

λ = wavelength (m)

v = frequency (Hz)

c = speed of light

In 1923 A.H. Compton started shooting high frequency X-rays at various materials and found that his results seemed to support the idea of photons having momentum. In one setup he shot the high frequency X-rays at a piece of graphite.

- If light was a wave, we would expect the X-rays to come out the other side with their wavelength smaller.

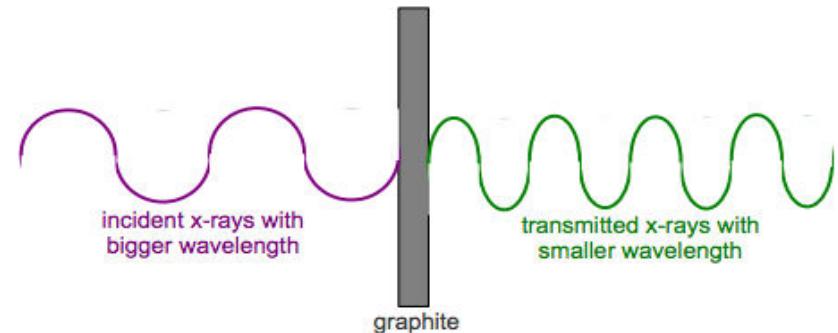


Illustration 1: If light was a wave, we'd expect results like this.

- Instead, Compton found that the X-rays scattered after hitting the target, changing the direction they were moving and actually getting a longer wavelength.
 - Remember, longer wavelength means smaller frequency.
 - Since $E = hv$, the scattered photons had less energy! Somehow, the X-ray photons were losing energy going through the graphite. So where'd the energy go?

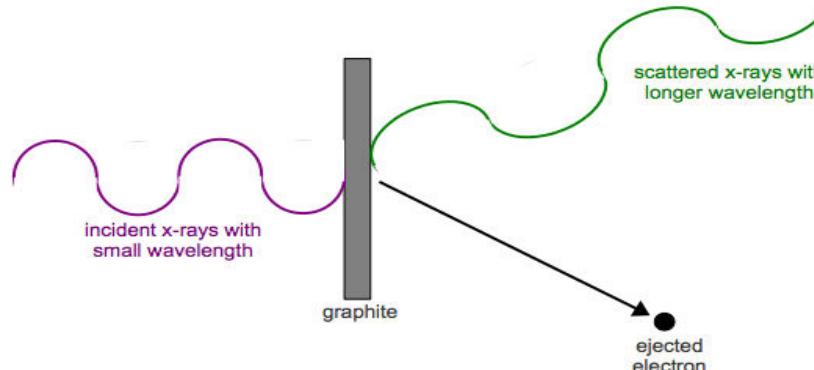
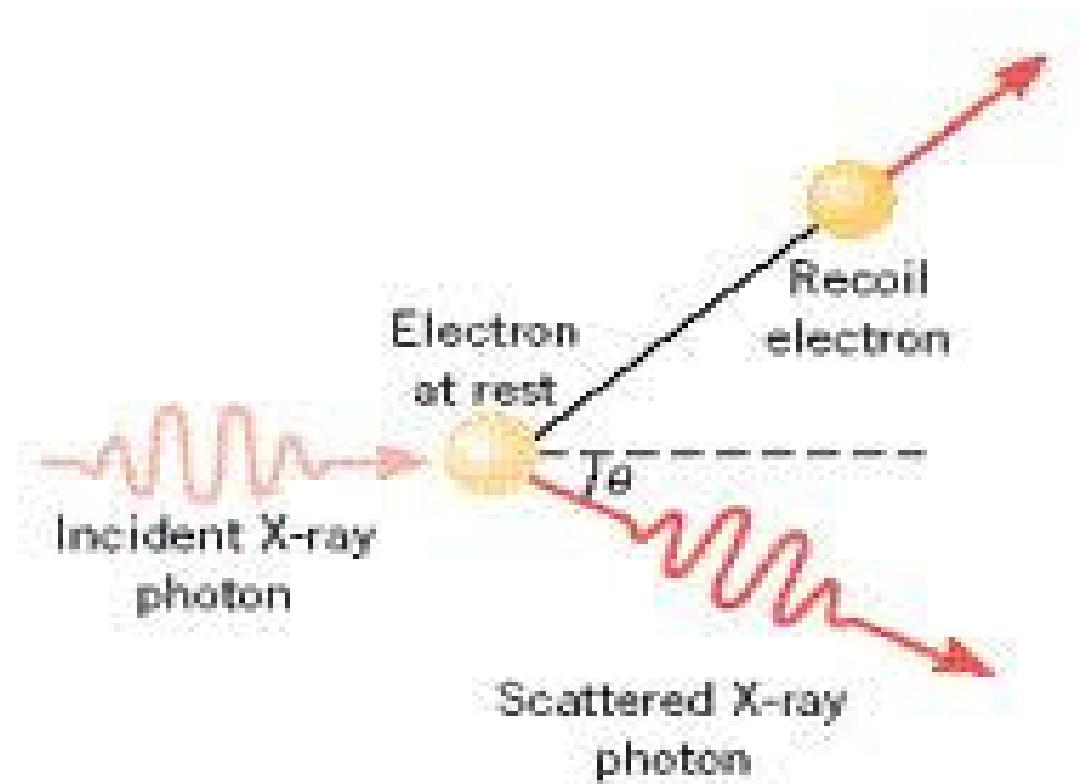


Illustration 2: What Compton actually observed.

Compton scattering is an inelastic scattering of a photon by a free charged particle, usually an electron. It results in a decrease in energy (increase in wavelength) of the photon (which may be an X-ray or gamma ray photon), called the Compton effect. Part of the energy of the photon is transferred to the recoiling electron.



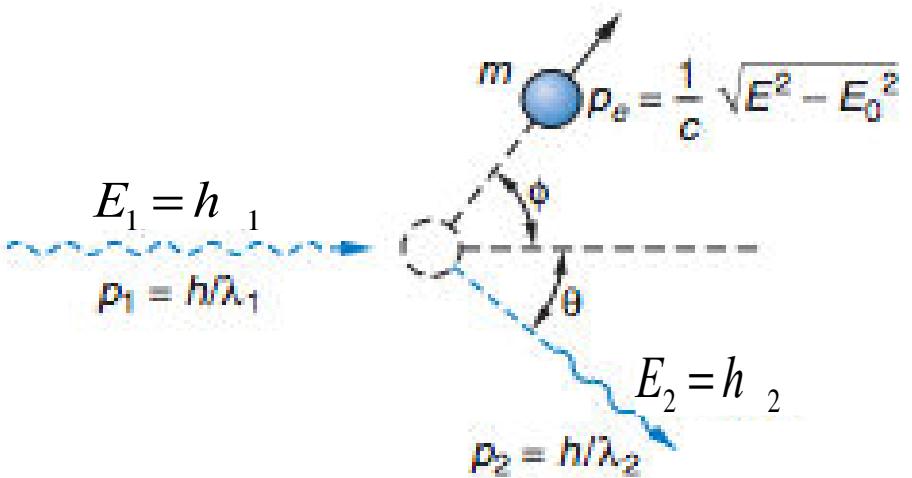
Let $v_1(\lambda_1)$ and $v_2(\lambda_2)$ be the frequencies (wavelengths) of the incident and scattered x rays, respectively, as shown in Figure. The corresponding momenta are

For incoming photon

$$p_1 = \frac{h}{c} v_1 = \frac{E_1}{c} = \frac{h}{\lambda_1}$$

For scattering photon

$$p_2 = \frac{h}{c} v_2 = \frac{E_2}{c} = \frac{h}{\lambda_2}$$



Energy of electron at rest

$$E_0 = mc^2$$

Energy of electron after scattering

$$E_e = \sqrt{m^2 c^4 + p_e^2 c^2}$$

From the conservation of *Momentum*

03.compton_scattering_04-Jan-2021_Reference_Material_II

$$\vec{p}_1 + \vec{0} = \vec{p}_2 + \vec{p}_e$$

$$\Rightarrow \vec{p}_e = \vec{p}_1 - \vec{p}_2$$

$$\Rightarrow (\vec{p}_e)^2 = (\vec{p}_1 - \vec{p}_2)^2$$

$$\Rightarrow \vec{p}_e^2 = p_1^2 + p_2^2 - 2\vec{p}_1 \bullet \vec{p}_2$$

$$\Rightarrow \vec{p}_e^2 = p_1^2 + p_2^2 - 2\vec{p}_1 \vec{p}_2 \cos\theta \rightarrow (1)$$

From the conservation of *Energy*

$$E_1 + E_0 = E_2 + E_e$$

$$\Rightarrow h_1 + mc^2 = h_2 + \sqrt{m^2 c^4 + p_e^2 c^2}$$

$$\Rightarrow h_1 - h_2 + mc^2 = \sqrt{m^2 c^4 + p_e^2 c^2}$$

$$\Rightarrow p_1 c - p_2 c + mc^2 = \sqrt{m^2 c^4 + p_e^2 c^2}$$

$$\Rightarrow (p_1 - p_2)c + mc^2 = \sqrt{m^2 c^4 + p_e^2 c^2}$$

$$\Rightarrow [(p_1 - p_2)c + E_0]^2 = E_0^2 + p_e^2 c^2$$

$$\Rightarrow p_e^2 = (p_1 - p_2)^2 + \frac{2E_0(p_1 - p_2)}{c} \rightarrow (2)$$

$$\begin{aligned}
 p_1^2 + p_2^2 - 2p_1p_2 \cos\theta &= (p_1 - p_2)^2 + \frac{2E_0(p_1 - p_2)}{c} \\
 \Rightarrow p_1^2 + p_2^2 - 2p_1p_2 \cos\theta &= p_1^2 + p_2^2 - 2p_1p_2 + \frac{2E_0(p_1 - p_2)}{c} \\
 \Rightarrow \frac{E_0(p_1 - p_2)}{c} &= p_1p_2(1 - \cos\theta)
 \end{aligned}$$

Again $E_0 = mc^2$ & $p_1 = \frac{h}{1}$ & $p_2 = \frac{h}{2}$

$$\Rightarrow \frac{mc^2}{c} \left[\frac{h}{1} - \frac{h}{2} \right] = \frac{h}{1} \frac{h}{2} (1 - \cos\theta)$$

$$\frac{h}{2} - \frac{h}{1} = \frac{h}{mc} (1 - \cos\theta)$$

\rightarrow Compton's Equation

$$= \frac{h}{mc} (1 - \cos \theta)$$

$\Delta\lambda$ is change in wavelength

$$\frac{h}{mc} = \frac{6.626 \times 10^{-34}}{(9.11 \times 10^{-31})(3 \times 10^8)}$$

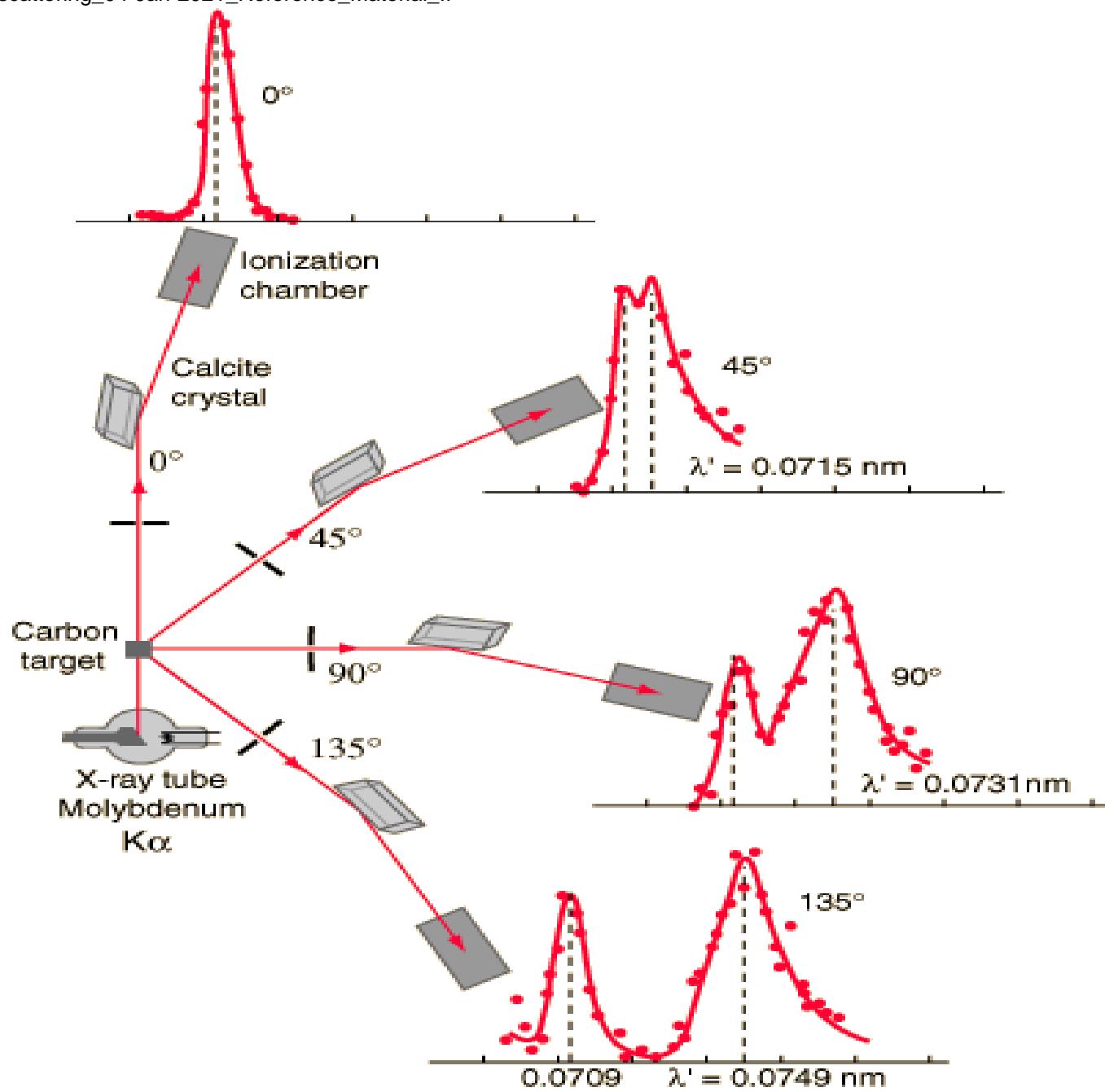
$$c = 2.426 \times 10^{-12} m = 0.02426 \text{ \AA}^\circ \longrightarrow \text{Compton Wavelength}$$

If $\theta=0$, $\Delta\lambda=0$, i.e. no wavelength shift along the direction of incident radiation.

The shift increases with increase of angle of scattering.

If $\theta=\pi/2$, $\Delta\lambda=\lambda_c$ & If $\theta=\pi$, $\Delta\lambda=2\lambda_c$

Thus Compton shift $\Delta\lambda$ and its dependence on the angle of scattering can be explained by treating X-rays as particles rather than waves.



Recalling Einstein relativistic equation

$$\text{Total Energy, } E = \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}}$$

$$E^2 = \frac{m^2 c^4}{1 - \frac{v^2}{c^2}}$$

$$\text{momentum, } P = \frac{mv}{\sqrt{1 - \frac{v^2}{c^2}}}$$

$$P^2 c^2 = \frac{m^2 v^2 c^2}{1 - \frac{v^2}{c^2}}$$

$$E^2 - p^2 c^2 = \frac{m^2 c^4}{\left[1 - \frac{v^2}{c^2}\right]} - \frac{m^2 v^2 c^2}{\left[1 - \frac{v^2}{c^2}\right]}$$

$$E^2 - p^2 c^2 = \frac{m^2 c^4 - m^2 v^2 c^2}{\left[1 - \frac{v^2}{c^2}\right]}$$

$$E^2 - p^2 c^2 = \frac{m^2 c^4 \left[1 - \frac{v^2}{c^2}\right]}{\left[1 - \frac{v^2}{c^2}\right]} = m^2 c^4$$

$$E^2 = m^2 c^4 + p^2 c^2$$

$$E = \sqrt{m^2 c^4 + p^2 c^2}$$

5

Matter Waves

Chapter Outline

- | | |
|---|---|
| <p>5.1 The Pilot Waves of de Broglie
<i>De Broglie's Explanation of Quantization in the Bohr Model</i></p> <p>5.2 The Davisson–Germer Experiment
<i>The Electron Microscope</i></p> <p>5.3 Wave Groups and Dispersion
<i>Matter Wave Packets</i></p> <p>5.4 Fourier Integrals (Optional)
<i>Constructing Moving Wave Packets</i></p> <p>5.5 The Heisenberg Uncertainty Principle
<i>A Different View of the Uncertainty Principle</i></p> <p>5.6 If Electrons Are Waves,
What's Waving?</p> | <p>5.7 The Wave–Particle Duality
<i>The Description of Electron Diffraction in Terms of Ψ</i>
<i>A Thought Experiment: Measuring Through Which Slit the Electron Passes</i></p> <p>5.8 A Final Note
Summary</p> |
|---|---|

In the previous chapter we discussed some important discoveries and theoretical concepts concerning the *particle nature of matter*. We now point out some of the shortcomings of these theories and introduce the fascinating and bizarre *wave properties of particles*. Especially notable are Count Louis de Broglie's remarkable ideas about how to represent electrons (and other particles) as waves and the experimental confirmation of de Broglie's hypothesis by the electron diffraction experiments of Davisson and Germer. We shall also see how the notion of representing a particle as a localized wave or wave group leads naturally to limitations on simultaneously measuring position and momentum of the particle. Finally, we discuss the passage of electrons through a double slit as a way of “understanding” the wave–particle duality of matter.



Figure 5.1 Louis de Broglie was a member of an aristocratic French family that produced marshals, ambassadors, foreign ministers, and at least one duke, his older brother Maurice de Broglie. Louis de Broglie came rather late to theoretical physics, as he first studied history. Only after serving as a radio operator in World War I did he follow the lead of his older brother and begin his studies of physics. Maurice de Broglie was an outstanding experimental physicist in his own right and conducted experiments in the palatial family mansion in Paris. (AIP Meggers Gallery of Nobel Laureates)

5.1 THE PILOT WAVES OF DE BROGLIE

By the early 1920s scientists recognized that the Bohr theory contained many inadequacies:

- It failed to predict the observed intensities of spectral lines.
- It had only limited success in predicting emission and absorption wavelengths for multielectron atoms.
- It failed to provide an equation of motion governing the time development of atomic systems starting from some initial state.
- It overemphasized the particle nature of matter and could not explain the newly discovered wave–particle duality of light.
- It did not supply a general scheme for “quantizing” other systems, especially those without periodic motion.

The first bold step toward a new mechanics of atomic systems was taken by Louis Victor de Broglie in 1923 (Fig. 5.1). In his doctoral dissertation he postulated that *because photons have wave and particle characteristics, perhaps all forms of matter have wave as well as particle properties*. This was a radical idea with no experimental confirmation at that time. According to de Broglie, electrons had a dual particle–wave nature. Accompanying every electron was a wave (not an electromagnetic wave!), which guided, or “piloted,” the electron through space. He explained the source of this assertion in his 1929 Nobel prize acceptance speech:

On the one hand the quantum theory of light cannot be considered satisfactory since it defines the energy of a light corpuscle by the equation $E = hf$ containing the frequency f . Now a purely corpuscular theory contains nothing that enables us to define a frequency; for this reason alone, therefore, we are compelled, in the case of light, to introduce the idea of a corpuscle and that of periodicity simultaneously. On the other hand, determination of the stable motion of electrons in the atom introduces integers, and up to this point the only phenomena involving integers in physics were those of interference and of normal modes of vibration. This fact suggested to me the idea that electrons too could not be considered simply as corpuscles, but that periodicity must be assigned to them also.

Let us look at de Broglie’s ideas in more detail. He concluded that the wavelength and frequency of a *matter wave* associated with any moving object were given by

$$\lambda = \frac{h}{p} \quad (5.1)$$

and

$$f = \frac{E}{h} \quad (5.2)$$

where h is Planck’s constant, p is the relativistic momentum, and E is the total relativistic energy of the object. Recall from Chapter 2 that p and E can be written as

$$p = \gamma mv \quad (5.3)$$

and

$$E^2 = p^2c^2 + m^2c^4 = \gamma^2m^2c^4 \quad (5.4)$$

where $\gamma = (1 - v^2/c^2)^{-1/2}$ and v is the object's speed. Equations 5.1 and 5.2 immediately suggest that it should be easy to calculate the speed of a de Broglie wave from the product λf . However, as we will show later, this is not the speed of the particle. Since the correct calculation is a bit complicated, we postpone it to Section 5.3. Before taking up the question of the speed of matter waves, we prefer first to give some introductory examples of the use of $\lambda = h/p$ and a brief description of how de Broglie waves provide a physical picture of the Bohr theory of atoms.

De Broglie's Explanation of Quantization in the Bohr Model

Bohr's model of the atom had many shortcomings and problems. For example, as the electrons revolve around the nucleus, how can one understand the fact that only certain electronic energies are allowed? Why do all atoms of a given element have precisely the same physical properties regardless of the infinite variety of starting velocities and positions of the electrons in each atom?

De Broglie's great insight was to recognize that although these are deep problems for particle theories, wave theories of matter handle these problems neatly by means of interference. For example, a plucked guitar string, although initially subjected to a wide range of wavelengths, supports only standing wave patterns that have nodes at each end. Thus only a discrete set of wavelengths is allowed for standing waves, while other wavelengths not included in this discrete set rapidly vanish by destructive interference. This same reasoning can be applied to electron matter waves bent into a circle around the nucleus. Although initially a continuous distribution of wavelengths may be present, corresponding to a distribution of initial electron velocities, most wavelengths and velocities rapidly die off. The residual standing wave patterns thus account for the identical nature of all atoms of a given element and show that atoms are more like vibrating drum heads with discrete modes of vibration than like miniature solar systems. This point of view is emphasized in Figure 5.2, which shows the standing wave pattern of the electron in the hydrogen atom corresponding to the $n = 3$ state of the Bohr theory.

Another aspect of the Bohr theory that is also easier to visualize physically by using de Broglie's hypothesis is the quantization of angular momentum. One simply assumes that **the allowed Bohr orbits arise because the electron matter waves interfere constructively when an integral number of wavelengths exactly fits into the circumference of a circular orbit**. Thus

$$n\lambda = 2\pi r \quad (5.5)$$

where r is the radius of the orbit. From Equation 5.1, we see that $\lambda = h/m_e v$. Substituting this into Equation 5.5, and solving for $m_e v r$, the angular momentum of the electron, gives

$$m_e v r = n\hbar \quad (5.6)$$

Note that this is precisely the Bohr condition for the quantization of angular momentum.

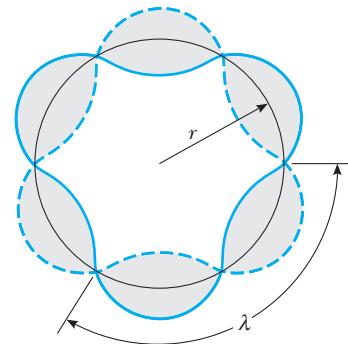


Figure 5.2 Standing waves fit to a circular Bohr orbit. In this particular diagram, three wavelengths are fit to the orbit, corresponding to the $n = 3$ energy state of the Bohr theory.

EXAMPLE 5.1 Why Don't We See the Wave Properties of a Baseball?

An object will appear “wavelike” if it exhibits interference or diffraction, both of which require scattering objects or apertures of about the same size as the wavelength. A baseball of mass 140 g traveling at a speed of 60 mi/h (27 m/s) has a de Broglie wavelength given by

$$\lambda = \frac{h}{p} = \frac{6.63 \times 10^{-34} \text{ J}\cdot\text{s}}{(0.14 \text{ kg})(27 \text{ m/s})} = 1.7 \times 10^{-34} \text{ m}$$

Even a nucleus (whose size is $\approx 10^{-15} \text{ m}$) is much too large to diffract this incredibly small wavelength! This explains why all macroscopic objects appear particle-like.

EXAMPLE 5.2 What Size “Particles” Do Exhibit Diffraction?

A particle of charge q and mass m is accelerated from rest through a small potential difference V . (a) Find its de Broglie wavelength, assuming that the particle is non-relativistic.

Solution When a charge is accelerated from rest through a potential difference V , its gain in kinetic energy, $\frac{1}{2}mv^2$, must equal the loss in potential energy qV . That is,

$$\frac{1}{2}mv^2 = qV$$

Because $p = mv$, we can express this in the form

$$\frac{p^2}{2m} = qV \quad \text{or} \quad p = \sqrt{2mqV}$$

Substituting this expression for p into the de Broglie relation $\lambda = h/p$ gives

$$\lambda = \frac{h}{p} = \frac{h}{\sqrt{2mqV}}$$

(b) Calculate λ if the particle is an electron and $V = 50 \text{ V}$.

Solution The de Broglie wavelength of an electron accelerated through 50 V is

$$\begin{aligned} \lambda &= \frac{h}{\sqrt{2m_e q V}} \\ &= \frac{6.63 \times 10^{-34} \text{ J}\cdot\text{s}}{\sqrt{2(9.11 \times 10^{-31} \text{ kg})(1.6 \times 10^{-19} \text{ C})(50 \text{ V})}} \\ &= 1.7 \times 10^{-10} \text{ m} = 1.7 \text{ Å} \end{aligned}$$

This wavelength is of the order of atomic dimensions and the spacing between atoms in a solid. Such low-energy electrons are routinely used in electron diffraction experiments to determine atomic positions on a surface.

Exercise 1 (a) Show that the de Broglie wavelength for an electron accelerated from rest through a *large* potential difference, V , is

$$\lambda = \frac{12.27}{V^{1/2}} \left(\frac{Ve}{2m_e c^2} + 1 \right)^{-1/2} \quad (5.7)$$

where λ is in angstroms (Å) and V is in volts. (b) Calculate the percent error introduced when $\lambda = 12.27/V^{1/2}$ is used instead of the correct relativistic expression for 10 MeV electrons.

Answer (b) 230%.

5.2 THE DAVISSON–GERMER EXPERIMENT

Direct experimental proof that electrons possess a wavelength $\lambda = h/p$ was furnished by the diffraction experiments of American physicists Clinton J. Davisson (1881–1958) and Lester H. Germer (1896–1971) at the Bell Laboratories in New York City in 1927 (Fig. 5.3).¹ In fact, de Broglie had already suggested in 1924 that a stream of electrons traversing a small aperture should exhibit diffraction phenomena. In 1925, Einstein was led to the necessity of postulating matter waves from an analysis of fluctuations of a molecular gas. In addition, he noted that a molecular beam should show small but measurable diffraction effects. In the same year, Walter Elsasser pointed out that the slow

¹C. J. Davisson and L. H. Germer, *Phys. Rev.* 30:705, 1927.



Figure 5.3 Clinton J. Davisson (left) and Lester H. Germer (center) at Bell Laboratories in New York City. (*Bell Laboratories, courtesy AIP Emilio Segrè Visual Archives*)

electron scattering experiments of C. J. Davisson and C. H. Kunsman at the Bell Labs could be explained by electron diffraction.

Clear-cut proof of the wave nature of electrons was obtained in 1927 by the work of Davisson and Germer in the United States and George P. Thomson (British physicist, 1892–1975, the son of J. J. Thomson) in England. Both cases are intriguing not only for their physics but also for their human interest. The first case was an accidental discovery, and the second involved the discovery of the particle properties of the electron by the father and the wave properties by the son.

The crucial experiment of Davisson and Germer was an offshoot of an attempt to understand the arrangement of atoms on the surface of a nickel sample by elastically scattering a beam of low-speed electrons from a polycrystalline nickel target. A schematic drawing of their apparatus is shown in Figure 5.4. Their device allowed for the variation of three experimental parameters—electron energy; nickel target orientation, α ; and scattering angle, ϕ . Before a fortunate accident occurred, the results seemed quite pedestrian. For constant electron energies of about 100 eV, the scattered intensity rapidly decreased as ϕ increased. But then someone dropped a flask of liquid air on the glass vacuum system, rupturing the vacuum and oxidizing the nickel target, which had been at high temperature. To remove the oxide, the sample was reduced by heating it cautiously² in a flowing stream of hydrogen. When the apparatus was reassembled, quite different results were found: Strong variations in the intensity of scattered electrons with angle were observed, as shown in Figure 5.5. The prolonged heating had evidently annealed the nickel target, causing large single-crystal regions to develop in the polycrystalline sample. These crystalline regions furnished the extended regular lattice needed to observe electron diffraction. Once Davisson and Germer realized that it was the elastic scattering from *single crystals* that produced such unusual results (1925), they initiated a thorough investigation of elastic scattering from large single crystals

²At present this can be done without the slightest fear of “stinks or bangs,” because 5% hydrogen–95% argon safety mixtures are commercially available.

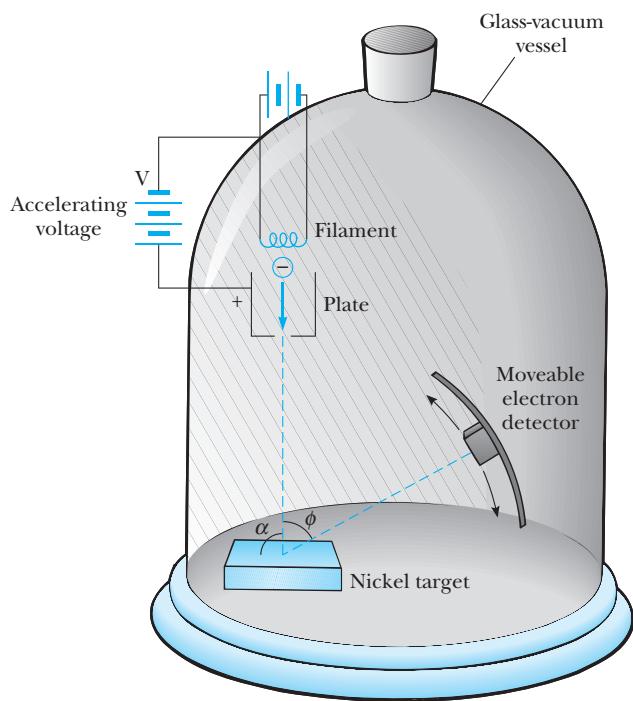


Figure 5.4 A schematic diagram of the Davisson–Germer apparatus.

with predetermined crystallographic orientation. Even these experiments were not conducted at first as a test of de Broglie’s wave theory, however. Following discussions with Richardson, Born, and Franck, the experiments and their analysis finally culminated in 1927 in the proof that electrons experience diffraction with an electron wavelength that is given by $\lambda = h/p$.

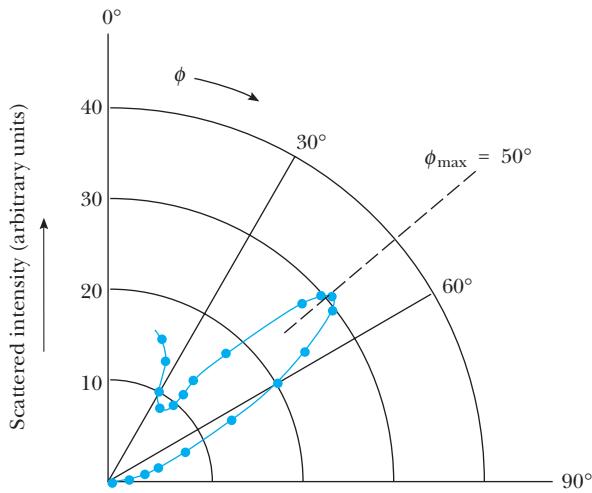


Figure 5.5 A polar plot of scattered intensity versus scattering angle for 54-eV electrons, based on the original work of Davisson and Germer. The scattered intensity is proportional to the distance of the point from the origin in this plot.

The idea that electrons behave like waves when interacting with the atoms of a crystal is so striking that Davisson and Germer's proof deserves closer scrutiny. In effect, they calculated the wavelength of electrons from a simple diffraction formula and compared this result with de Broglie's formula $\lambda = h/p$. Although they tested this result over a wide range of target orientations and electron energies, we consider in detail only the simple case shown in Figures 5.4 and 5.5 with $\alpha = 90.0^\circ$, $V = 54.0$ V, and $\phi = 50.0^\circ$, corresponding to the $n = 1$ diffraction maximum. In order to calculate the de Broglie wavelength for this case, we first obtain the velocity of a nonrelativistic electron accelerated through a potential difference V from the energy relation

$$\frac{1}{2}m_e v^2 = eV$$

Substituting $v = \sqrt{2Ve/m_e}$ into the de Broglie relation gives

$$\lambda = \frac{h}{m_e v} = \frac{h}{\sqrt{2Vem_e}} \quad (5.8)$$

Thus the wavelength of 54.0-V electrons is

$$\begin{aligned} \lambda &= \frac{6.63 \times 10^{-34} \text{ J}\cdot\text{s}}{\sqrt{2(54.0 \text{ V})(1.60 \times 10^{-19} \text{ C})(9.11 \times 10^{-31} \text{ kg})}} \\ &= 1.67 \times 10^{-10} \text{ m} = 1.67 \text{ \AA} \end{aligned}$$

The experimental wavelength may be obtained by considering the nickel atoms to be a reflection diffraction grating, as shown in Figure 5.6. Only the surface layer of atoms is considered because low-energy electrons, unlike x-rays, do not penetrate deeply into the crystal. Constructive interference occurs when the path length difference between two adjacent rays is an integral number of wavelengths or

$$d \sin \phi = n\lambda \quad (5.9)$$

As d was known to be 2.15 Å from x-ray diffraction measurements, Davisson and Germer calculated λ to be

$$\lambda = (2.15 \text{ \AA}) (\sin 50.0^\circ) = 1.65 \text{ \AA}$$

in excellent agreement with the de Broglie formula.

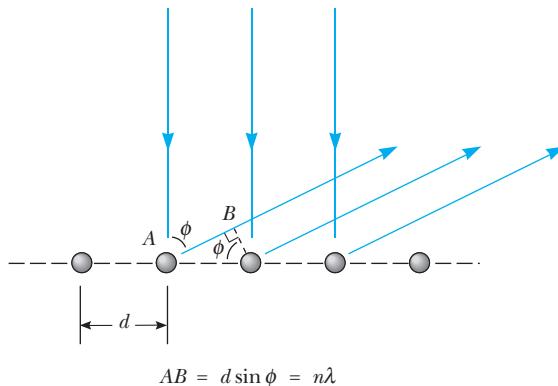


Figure 5.6 Constructive interference of electron matter waves scattered from a single layer of atoms at an angle ϕ .



Figure 5.7 Diffraction of 50-kV electrons from a film of Cu₃Au. The alloy film was 400 Å thick. (Courtesy of the late Dr. L. H. Germer)

It is interesting to note that while the diffraction lines from low-energy reflected electrons are quite broad (see Fig. 5.5), the lines from high-energy electrons transmitted through metal foils are quite sharp (see Fig. 5.7). This effect occurs because hundreds of atomic planes are penetrated by high-energy electrons, and consequently Equation 5.9, which treats diffraction from a surface layer, no longer holds. Instead, the Bragg law, $2d \sin \theta = n\lambda$, applies to high-energy electron diffraction. The maxima are extremely sharp in this case because if $2d \sin \theta$ is not exactly equal to $n\lambda$, there will be no diffracted wave. This occurs because there are scattering contributions from so many atomic planes that eventually the path length difference between the wave from the first plane and some deeply buried plane will be *an odd multiple of $\lambda/2$* , resulting in complete cancellation of these waves (see Problem 13).

Image not available due to copyright restrictions

If de Broglie's postulate is true for all matter, then any object of mass m has wavelike properties and a wavelength $\lambda = h/p$. In the years following Davisson and Germer's discovery, experimentalists tested the universal character of de Broglie's postulate by searching for diffraction of other "particle" beams. In subsequent experiments, diffraction was observed for helium atoms (Estermann and Stern in Germany) and hydrogen atoms (Johnson in the United States). Following the discovery of the neutron in 1932, it was shown that neutron beams of the appropriate energy also exhibit diffraction when incident on a crystalline target (Fig. 5.8).

EXAMPLE 5.3 Thermal Neutrons

What kinetic energy (in electron volts) should neutrons have if they are to be diffracted from crystals?

Solution Appreciable diffraction will occur if the de Broglie wavelength of the neutron is of the same order of magnitude as the interatomic distance. Taking $\lambda = 1.00 \text{ \AA}$, we find

$$p = \frac{h}{\lambda} = \frac{6.63 \times 10^{-34} \text{ J}\cdot\text{s}}{1.00 \times 10^{-10} \text{ m}} = 6.63 \times 10^{-24} \text{ kg}\cdot\text{m/s}$$

The kinetic energy is given by

$$\begin{aligned} K &= \frac{p^2}{2m_n} = \frac{(6.63 \times 10^{-24} \text{ J}\cdot\text{s})^2}{2(1.66 \times 10^{-27} \text{ kg})} \\ &= 1.32 \times 10^{-20} \text{ J} = 0.0825 \text{ eV} \end{aligned}$$

Note that these neutrons are nonrelativistic because K is much less than the neutron rest energy of 940 MeV, and so our use of the classical expression $K = p^2/2m_n$ is justified. Because the average thermal energy of a par-

ticle in thermal equilibrium is $\frac{1}{2}k_B T$ for each independent direction of motion, neutrons at room temperature (300 K) possess a kinetic energy of

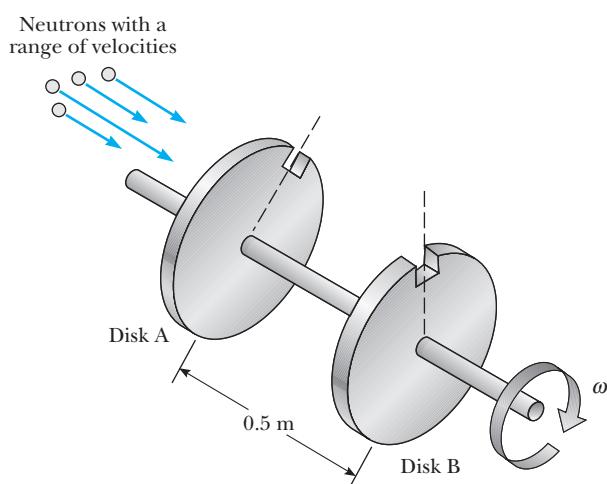
$$\begin{aligned} K &= \frac{3}{2}k_B T = (1.50)(8.62 \times 10^{-5} \text{ eV/K})(300 \text{ K}) \\ &= 0.0388 \text{ eV} \end{aligned}$$

Thus "thermal neutrons," or neutrons in thermal equilibrium with matter at room temperature, possess energies of the right order of magnitude to diffract appreciably from single crystals. Neutrons produced in a nuclear reactor are far too energetic to produce diffraction from crystals and must be slowed down in a graphite column as they leave the reactor. In the graphite moderator, repeated collisions with carbon atoms ultimately reduce the average neutron energies to the average thermal energy of the carbon atoms. When this occurs, these so-called thermalized neutrons possess a distribution of velocities and a corresponding distribution of de Broglie wavelengths with average wavelengths comparable to crystal spacings.

Exercise 2 *Monochromatic Neutrons.* A beam of neutrons with a single wavelength may be produced by means of a mechanical velocity selector of the type shown in Figure 5.9. (a) Calculate the speed of neutrons with a wavelength of 1.00 \AA . (b) What rotational speed (in rpm) should the shaft have in order to pass neutrons with wavelength of 1.00 \AA ?

Answers (a) $3.99 \times 10^3 \text{ m/s}$. (b) 13,300 rev/min.

Figure 5.9 A neutron velocity selector. The slot in disk B lags the slot in disk A by 10° .



The Electron Microscope

The idea that electrons have a controllable wavelength that can be made much shorter than visible light wavelengths and, accordingly, possess a much better ability to resolve fine details was only one of the factors that led to the development of the electron microscope. In fact, ideas of such a device were tossed about in the cafés and bars of Paris and Berlin as early as 1928. What really made the difference was the coming together of several lines of development—electron tubes and circuits, vacuum technology, and electron beam control—all pioneered in the development of the cathode ray tube (CRT). These factors led to the construction of the first transmission electron microscope (TEM) with magnetic lenses by electrical engineers Max Knoll and Ernst Ruska in Berlin in 1931. The testament to the fortitude and brilliance of Knoll and Ruska in overcoming the “cussedness of objects” and building and getting such a complicated experimental device to work for the first time is shown in Figure 5.10. It is remarkable that although the overall performance of the TEM has been improved thousands of times since its invention, it is basically the same in principle as that first designed by Knoll and Ruska: a device that focuses electron beams with magnetic lenses and creates a flat-looking two-dimensional shadow pattern on its screen, the result of varying degrees of electron transmission through the object. Figure 5.11a is a diagram showing this basic design and Figure 5.11b shows, for comparison, an optical projection microscope. The best optical microscopes using ultraviolet light have a magnification of about 2000 and can resolve two objects separated by 100 nm, but a TEM using electrons accelerated through 100 kV has a magnification of as much as 1,000,000 and a maximum resolution of 0.2 nm. In practice, magnifications of 10,000 to 100,000 are easier to use. Figure 5.12 shows typical TEM micrographs of microbes, Figure 5.12b showing a microbe and its DNA strands magnified 40,000 times. Although it would seem that increasing electron energy should lead to shorter electron wavelength and increased resolution, imperfections or aberrations in the magnetic lenses actually set the limit of resolution at about 0.2 nm. Increasing electron energy above 100 keV does not



Ernst Ruska played a major role in the invention of the TEM. He was awarded the Nobel prize in physics for this work in 1986. (AIP Emilio Segrè Visual Archives, W. F. Meggers Gallery of Nobel Laureates)

Image not available due to copyright restrictions

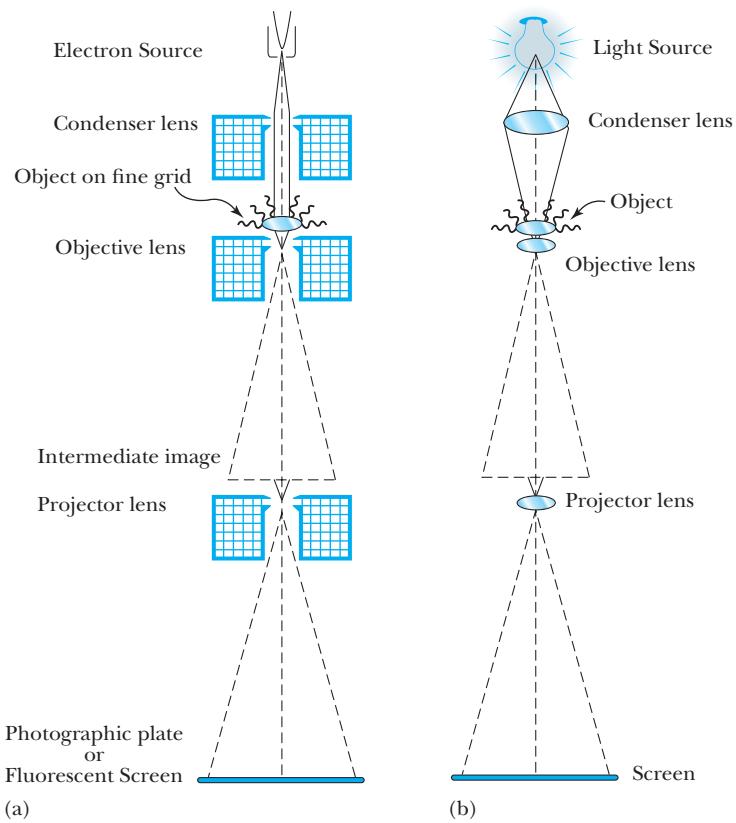


Figure 5.11 (a) Schematic drawing of a transmission electron microscope with magnetic lenses. (b) Schematic of a light-projection microscope.

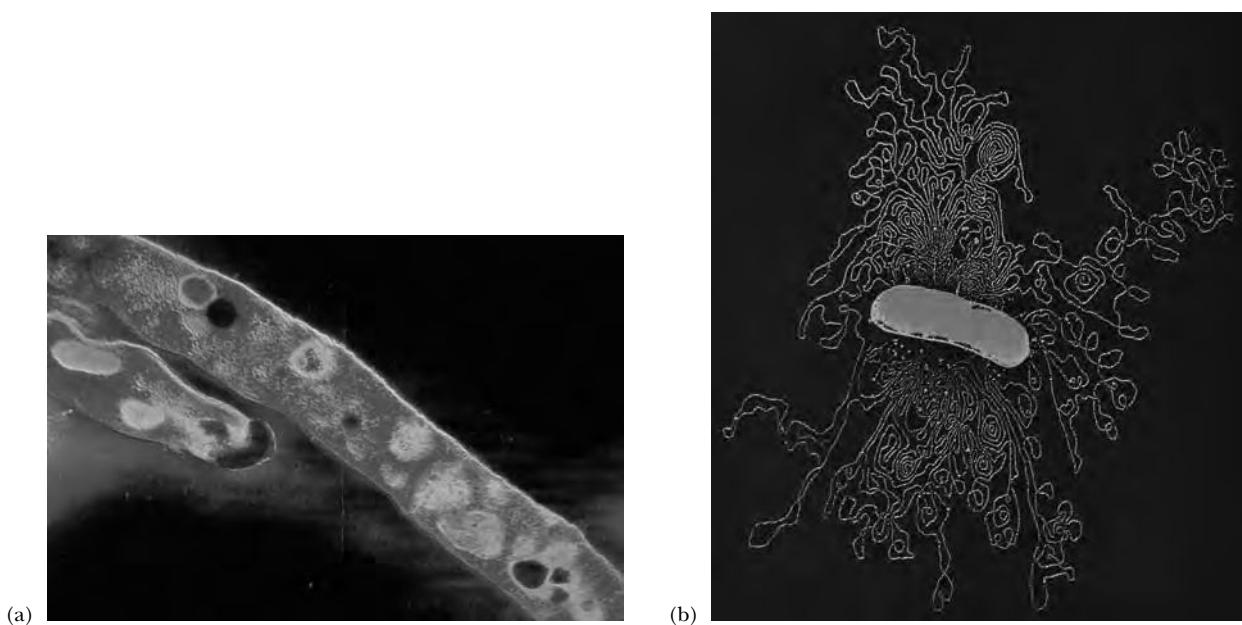


Figure 5.12 (a) A false-color TEM micrograph of tuberculosis bacteria. (b) A TEM micrograph of a microbe leaking DNA ($\times 40,000$). (CNRI/Photo Researchers, Inc., Dr. Gopal Murti/Photo Researchers, Inc.)

improve resolution—it only permits electrons to sample regions deeper inside an object. Figures 5.13a and 5.13b show, respectively, a diagram of a modern TEM and a photo of the same instrument.

A second type of electron microscope with less resolution and magnification than the TEM, but capable of producing striking three-dimensional images, is the scanning electron microscope (SEM). Figure 5.14 shows dramatic three-dimensional SEM micrographs made possible by the large range of focus (depth of field) of the SEM, which is several hundred times better than that of a light microscope. The SEM was the brainchild of the same Max Knoll who helped invent the TEM. Knoll had recently moved to the television department at Telefunken when he conceived of the idea in 1935. The SEM produces a sort of giant television image by collecting electrons scattered from an object, rather than light. The first operating scanning microscope was built by M. von Ardenne in 1937, and it was extensively developed and perfected by Vladimir Zworykin and collaborators at RCA Camden in the early 1940s.

Figure 5.15 shows how a typical SEM works. Such a device might be operated with 20-keV electrons and have a resolution of about 10 nm and a magnification ranging from 10 to 100,000. As shown in Figure 5.15, an electron beam is sharply focused on a specimen by magnetic lenses and then scanned (rastered) across a tiny region on the surface of the specimen. The high-energy primary beam scatters lower-energy secondary electrons out of the object depending on specimen composition and surface topography. These secondary electrons are detected by a plastic scintillator coupled to a photomultiplier, amplified, and used to modulate the brightness of a simultaneously

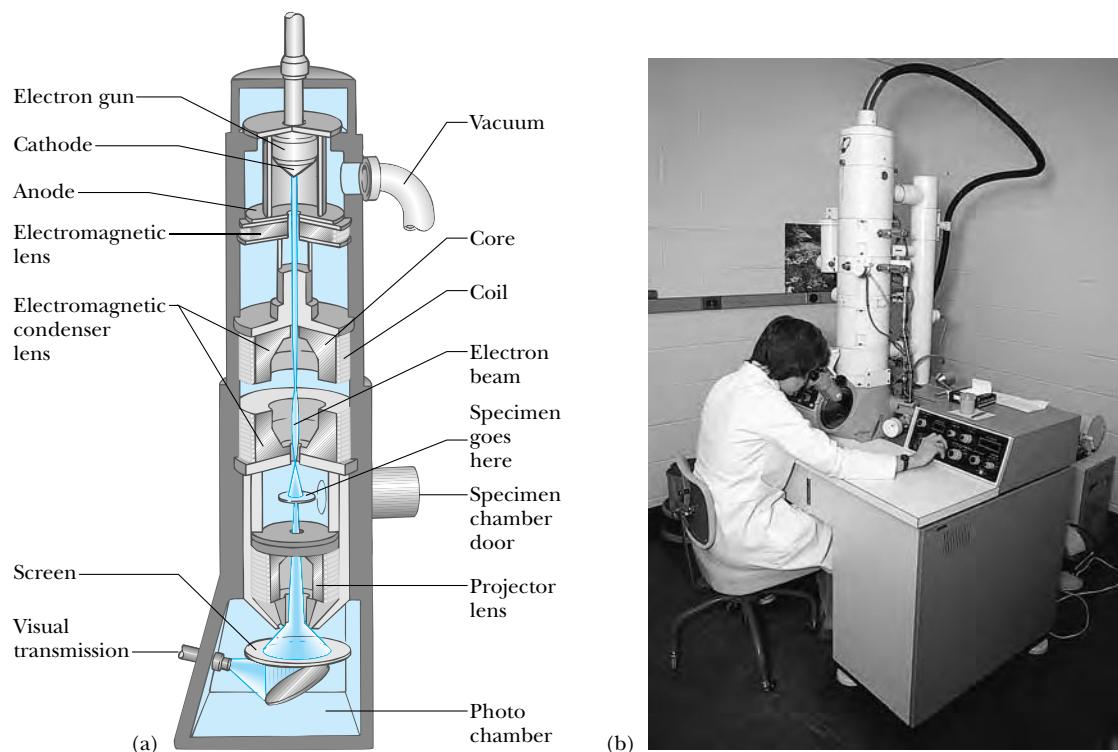


Figure 5.13 (a) Diagram of a transmission electron microscope. (b) A photo of the same TEM. (*W. Ormerod/Visuals Unlimited*)

rastered display CRT. The ratio of the display raster size to the microscope electron beam raster size determines the magnification. Modern SEM's can also collect x-rays and high-energy electrons from the specimen to detect chemical elements at certain locations on the specimen's surface, thus answering the bonus question, "Is the bitty bump on the bilayer boron or bismuth?"

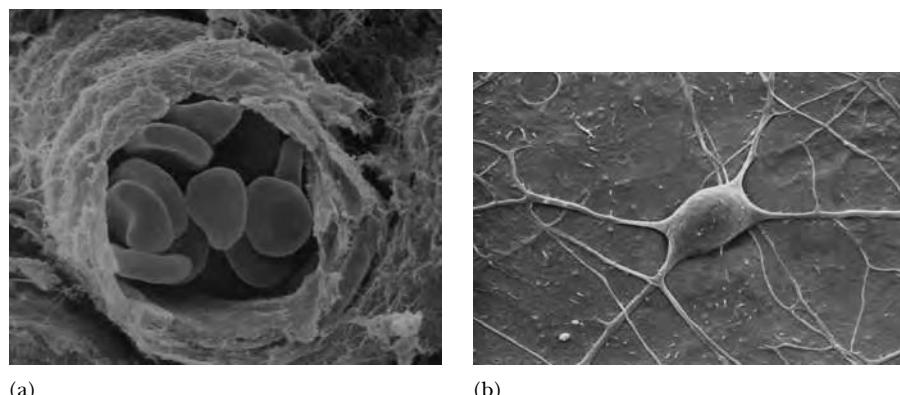


Figure 5.14 (a) A SEM micrograph showing blood cells in a tiny artery. (b) A SEM micrograph of a single neuron ($\times 4000$). (*P. Motta & S. Corrier/Photo Researchers, Inc., David McCarthy/Photo Researchers, Inc.*)

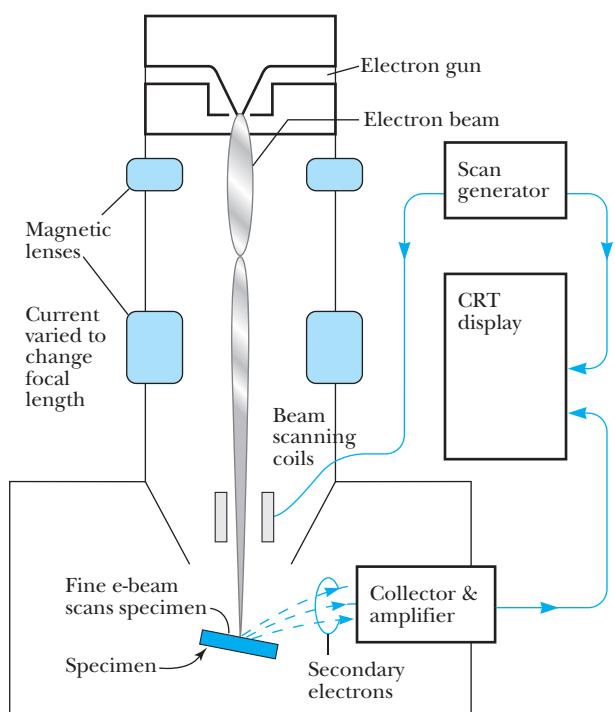


Figure 5.15 The working parts of a scanning electron microscope.

The newer, higher-resolution scanning tunneling microscope (STM) and atomic force microscope (AFM), which can image individual atoms and molecules, are discussed in Chapter 7. These instruments are exciting not only for their superb pictures of surface topography and individual atoms (see Figure 5.16 for an AFM picture) but also for their potential as microscopic machines capable of detecting and moving a few atoms at a time in proposed microchip terabit memories and mass spectrometers.

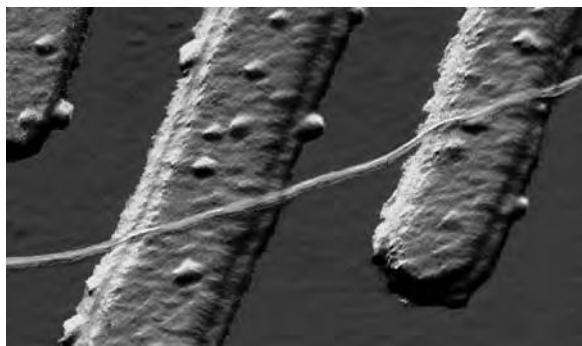


Figure 5.16 World's smallest electrical wire. An AFM image of a carbon nanotube wire on platinum electrodes. The wire is 1.5 nm wide, a mere 10 atoms. The magnification is 120,000. (*Delft University of Technology/Photo Researchers, Inc.*)

5.3 WAVE GROUPS AND DISPERSION

The matter wave representing a moving particle must reflect the fact that the particle has a large probability of being found in a small region of space only at a specific time. This means that a traveling sinusoidal matter wave of infinite extent and constant amplitude cannot properly represent a localized moving particle. What is needed is a pulse, or “wave group,” of limited spatial extent. Such a pulse can be formed by adding sinusoidal waves with different wavelengths. The resulting wave group can then be shown to move with a speed v_g (the group speed) identical to the classical particle speed. This argument is shown schematically in Figure 5.17 and will be treated in detail after the introduction of some general ideas about wave groups.

Actually, all observed waves are limited to definite regions of space and are called *pulses*, *wave groups*, or *wave packets* in the case of matter waves. The plane wave with an exact wavelength and infinite extension is an abstraction. Water waves from a stone dropped into a pond, light waves emerging from a briefly opened shutter, a wave generated on a taut rope by a single flip of one end, and a sound wave emitted by a discharging capacitor must all be modeled by wave groups. A wave group consists of a superposition of waves with *different wavelengths*, with the amplitude and phase of each component wave adjusted so that the waves interfere constructively over a small region of space. Outside of this region the combination of waves produces a net amplitude that approaches zero rapidly as a result of destructive interference. Perhaps the most familiar physical example in which wave groups arise is the phenomenon of beats. Beats occur when two sound waves of slightly different wavelength (and hence different frequency) are combined. The resultant sound wave has a frequency equal to the average of the two combining waves and an amplitude that fluctuates, or “beats,” at a rate

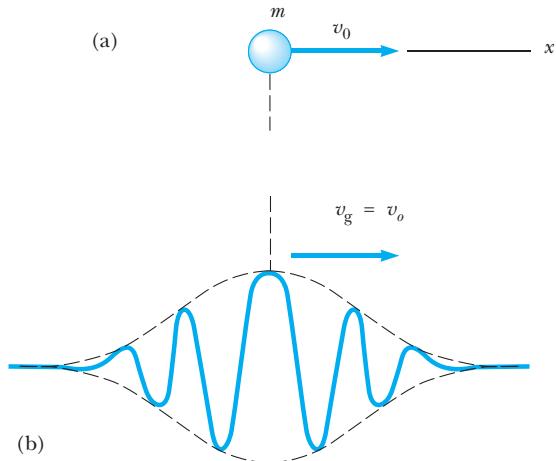


Figure 5.17 Representing a particle with matter waves: (a) particle of mass m and speed v_0 ; (b) superposition of many matter waves with a spread of wavelengths centered on $\lambda_0 = h/mv_0$ correctly represents a particle.

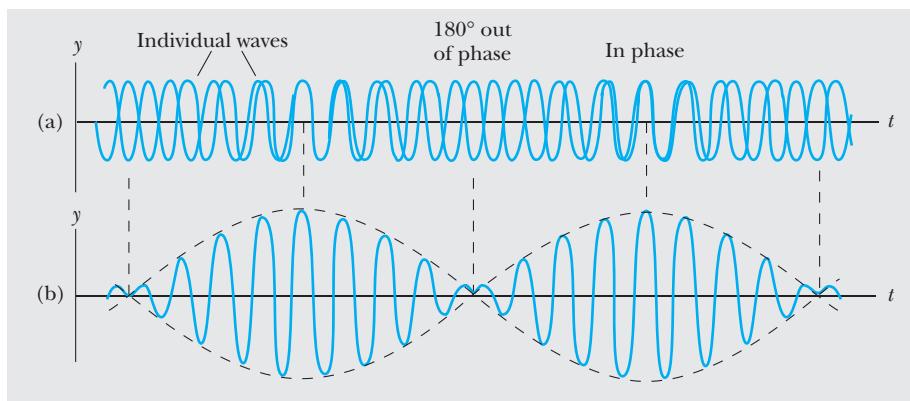


Figure 5.18 Beats are formed by the combination of two waves of slightly different frequency traveling in the same direction. (a) The individual waves. (b) The combined wave has an amplitude (broken line) that oscillates in time.

given by the difference of the two original frequencies. This case is illustrated in Figure 5.18.

Let us examine this situation mathematically. Consider a one-dimensional wave propagating in the positive x direction with a *phase speed* v_p . Note that v_p is the speed of a point of constant phase on the wave, such as a wave crest or trough. This traveling wave with wavelength λ , frequency f , and amplitude A may be described by

$$y = A \cos\left(\frac{2\pi x}{\lambda} - 2\pi ft\right) \quad (5.10)$$

where λ and f are related by

$$v_p = \lambda f \quad (5.11)$$

A more compact form for Equation 5.10 results if we take $\omega = 2\pi f$ (where ω is the angular frequency) and $k = 2\pi/\lambda$ (where k is the wavenumber). With these substitutions the infinite wave becomes

$$y = A \cos(kx - \omega t) \quad (5.12)$$

with

$$v_p = \frac{\omega}{k} \quad (5.13) \quad \text{Phase velocity}$$

Let us now form the superposition of two waves of equal amplitude both traveling in the positive x direction but with slightly different wavelengths, frequencies, and phase velocities. The resultant amplitude y is given by

$$y = y_1 + y_2 = A \cos(k_1 x - \omega_1 t) + A \cos(k_2 x - \omega_2 t)$$

Using the trigonometric identity

$$\cos a + \cos b = 2 \cos \frac{1}{2}(a - b) \cdot \cos \frac{1}{2}(a + b)$$

we find

$$y = 2A \cos \frac{1}{2} \{(k_2 - k_1)x - (\omega_2 - \omega_1)t\} \cdot \cos \frac{1}{2} \{(k_1 + k_2)x - (\omega_1 + \omega_2)t\} \quad (5.14)$$

For the case of two waves with slightly different values of k and ω , we see that $\Delta k = k_2 - k_1$ and $\Delta\omega = \omega_2 - \omega_1$ are small, but $(k_1 + k_2)$ and $(\omega_1 + \omega_2)$ are large. Thus, Equation 5.14 may be interpreted as a broad sinusoidal envelope

$$2A \cos \left(\frac{\Delta k}{2}x - \frac{\Delta\omega}{2}t \right)$$

limiting or modulating a high-frequency wave within the envelope

$$\cos \left[\frac{1}{2}(k_1 + k_2)x - \frac{1}{2}(\omega_1 + \omega_2)t \right]$$

This superposition of two waves is shown in Figure 5.19.

Although our model is primitive and does not represent a pulse limited to a small region of space, it shows several interesting features common to more complicated models. For example, the envelope and the wave within the envelope move at different speeds. The speed of either the high-frequency wave or the envelope is given by dividing the coefficient of the t term by the coefficient of the x term as was done in Equations 5.12 and 5.13. For the wave *within the envelope*,

$$v_p = \frac{(\omega_1 + \omega_2)/2}{(k_1 + k_2)/2} \approx \frac{\omega_1}{k_1} = v_1$$

Thus, the high-frequency wave moves at the phase velocity v_1 of one of the waves or at v_2 because $v_1 \approx v_2$. The envelope or group described by $2A \cos[(\Delta k/2)x - (\Delta\omega/2)t]$ moves with a different velocity however, the group velocity given by

$$v_g = \frac{(\omega_2 - \omega_1)/2}{(k_2 - k_1)/2} = \frac{\Delta\omega}{\Delta k} \quad (5.15)$$

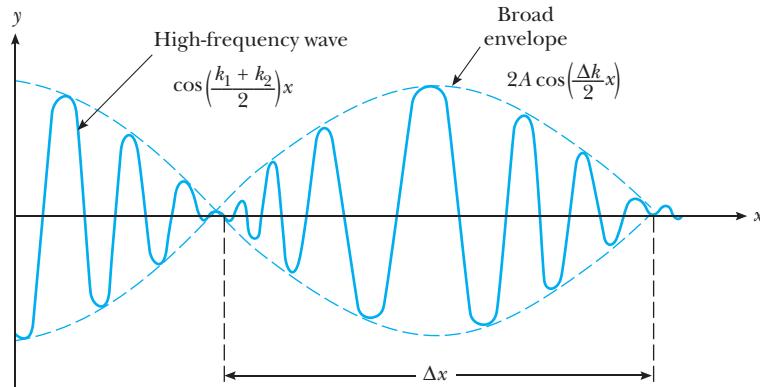


Figure 5.19 Superposition of two waves of slightly different wavelengths resulting in primitive wave groups; t has been set equal to zero in Equation 5.14.

Another general characteristic of wave groups for waves of any type is both a limited duration in time, Δt , and a limited extent in space, Δx . It is found that the smaller the spatial width of the pulse, Δx , the larger the range of wavelengths or wavenumbers, Δk , needed to form the pulse. This may be stated mathematically as

$$\Delta x \Delta k \approx 1 \quad (5.16)$$

Likewise, if the time duration, Δt , of the pulse is small, we require a wide spread of frequencies, $\Delta\omega$, to form the group. That is,

$$\Delta t \Delta\omega \approx 1 \quad (5.17)$$

In pulse electronics, this condition is known as the “response time–bandwidth formula.”³ In this situation Equation 5.17 shows that in order to amplify a voltage pulse of time width Δt without distortion, a pulse amplifier must equally amplify all frequencies in a frequency band of width $\Delta\omega$.

Equations 5.16 and 5.17 are important because they constitute “uncertainty relations,” or “reciprocity relations,” for pulses of any kind—electromagnetic, sound, or even matter waves. In particular, Equation 5.16 shows that Δx , the uncertainty in spatial extent of a pulse, is inversely proportional to Δk , the range of wavenumbers making up the pulse: **both Δx and Δk cannot become arbitrarily small, but as one decreases the other must increase.**

It is interesting that our simple two-wave model also shows the general principles given by Equations 5.16 and 5.17. If we call (rather artificially) the spatial extent of our group the distance between adjacent minima (labeled Δx in Figure 5.12), we find from the envelope term $2A \cos(\frac{1}{2}\Delta kx)$ the condition $\frac{1}{2}\Delta k \Delta x = \pi$ or

$$\Delta k \Delta x = 2\pi \quad (5.18)$$

Here, $\Delta k = k_2 - k_1$ is the range of wavenumbers present. Likewise, if x is held constant and t is allowed to vary in the envelope portion of Equation 5.14, the result is $\frac{1}{2}(\omega_2 - \omega_1) \Delta t = \pi$, or

$$\Delta\omega \Delta t = 2\pi \quad (5.19)$$

Therefore, Equations 5.18 and 5.19 agree with the general principles, respectively, of $\Delta k \Delta x \approx 1$ and $\Delta\omega \Delta t \approx 1$.

The addition of only two waves with discrete frequencies is instructive but produces an infinite wave instead of a true pulse. In the general case, many waves having a continuous distribution of wavelengths must be added to form a packet that is finite over a limited range and really zero everywhere else. In this case Equation 5.15 for the **group velocity**, v_g becomes

$$v_g = \left. \frac{d\omega}{dk} \right|_{k_0} \quad (5.20) \quad \text{Group velocity}$$

³It should be emphasized that Equations 5.16 and 5.17 are true in general and that the quantities Δx , Δk , Δt , and $\Delta\omega$ represent the spread in values present in an *arbitrary* pulse formed from the superposition of two or *more* waves.

where the derivative is to be evaluated at k_0 , the central wavenumber of the many waves present. The connection between the group velocity and the phase velocity of the composite waves is easily obtained. Because $\omega = kv_p$, we find

$$v_g = \frac{d\omega}{dk} \Big|_{k_0} = v_p \Big|_{k_0} + k \frac{dv_p}{dk} \Big|_{k_0} \quad (5.21)$$

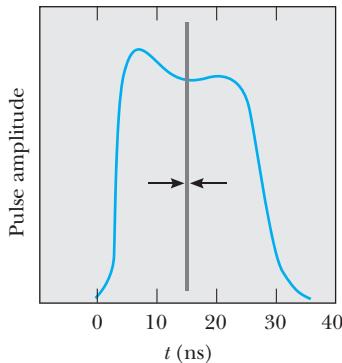


Figure 5.20 Dispersion in a 1-ns laser pulse. A pulse that starts with the width shown by the vertical lines has a time width of approximately 30 ns after traveling 1 km along an optical fiber.

where v_p is the phase velocity and is, in general, a function of k or λ . Materials in which the phase velocity varies with wavelength are said to exhibit **dispersion**. An example of a dispersive medium is glass, in which the index of refraction varies with wavelength and different colors of light travel at different speeds. Media in which the phase velocity does not vary with wavelength (such as vacuum for electromagnetic waves) are termed *nondispersive*. The term *dispersion* arises from the fact that the individual harmonic waves that form a pulse travel at different phase velocities and cause an originally sharp pulse to change shape and become spread out, or dispersed. As an example, dispersion of a laser pulse after traveling 1 km along an optical fiber is shown in Figure 5.20. In a nondispersive medium where all waves have the same velocity, the group velocity is equal to the phase velocity. In a dispersive medium the group velocity can be less than or greater than the phase velocity, depending on the sign of dv_p/dk , as shown by Equation 5.21.

EXAMPLE 5.4 Group Velocity in a Dispersive Medium

In a particular substance the phase velocity of waves doubles when the wavelength is halved. Show that wave groups in this system move at twice the *central* phase velocity.

Solution From the given information, the dependence of phase velocity on wavelength must be

$$v_p = \frac{A'}{\lambda} = Ak$$

for some constants A' and A . From Equation 5.21 we obtain

$$v_g = v_p \Big|_{k_0} + k \frac{dv_p}{dk} \Big|_{k_0} = Ak_0 + Ak_0 = 2Ak_0$$

Thus,

$$v_g = 2v_p \Big|_{k_0}$$

EXAMPLE 5.5 Group Velocity in Deep Water Waves

Newton showed that the phase velocity of deep water waves having wavelength λ is given by

$$v_p = \sqrt{\frac{g\lambda}{2\pi}}$$

where g is the acceleration of gravity and where the minor contribution of surface tension has been ignored. Show that in this case the velocity of a group of these waves is one-half of the phase velocity of the central wavelength.

Solution Because $k = 2\pi/\lambda$, we can write v_p as

$$v_p = \left(\frac{g}{k}\right)^{1/2}$$

Therefore, we find

$$\begin{aligned} v_g &= v_p \Big|_{k_0} + k \frac{dv_p}{dk} \Big|_{k_0} = \left(\frac{g}{k_0}\right)^{1/2} - \frac{1}{2}\left(\frac{g}{k_0}\right)^{1/2} \\ &= \frac{1}{2}\left(\frac{g}{k_0}\right)^{1/2} \equiv \frac{1}{2}v_p \Big|_{k_0} \end{aligned}$$

Matter Wave Packets

We are now in a position to apply our general theory of wave groups to electrons. We shall show both the dispersion of de Broglie waves and the satisfying result that the wave packet and the particle move at the same velocity. According to de Broglie, *individual* matter waves have a frequency f and a wavelength λ given by

$$f = \frac{E}{\hbar} \quad \text{and} \quad \lambda = \frac{\hbar}{p}$$

where E and p are the relativistic energy and momentum of the particle, respectively. The phase speed of these matter waves is given by

$$v_p = f\lambda = \frac{E}{p} \quad (5.22)$$

The phase speed can be expressed as a function of p or k alone by substituting $E = (\gamma^2 c^2 + m^2 c^4)^{1/2}$ into Equation 5.22:

$$v_p = c \sqrt{1 + \left(\frac{mc}{p}\right)^2} \quad (5.23)$$

The dispersion relation for de Broglie waves can be obtained as a function of k by substituting $p = \hbar/\lambda = \hbar k$ into Equation 5.23. This gives

$$v_p = c \sqrt{1 + \left(\frac{mc}{\hbar k}\right)^2} \quad (5.24)$$

Phase velocity of matter waves

Equation 5.24 shows that individual de Broglie waves representing a particle of mass m show dispersion even in *empty space* and always travel at a speed that is *greater than* or at least equal to c . Because these component waves travel at different speeds, the width of the wave packet, Δx , spreads or disperses as time progresses, as will be seen in detail in Chapter 6. To obtain the group speed, we use

$$v_g = \left[v_p + k \frac{dv_p}{dk} \right]_{k_0}$$

and Equation 5.24. After some algebra, we find

$$v_g = \frac{c}{\left[1 + \left(\frac{mc}{\hbar k_0} \right)^2 \right]^{1/2}} = \frac{c^2}{v_p} \Big|_{k_0} \quad (5.25)$$

Solving for the phase speed from Equation 5.22, we find

$$v_p = \frac{E}{p} = \frac{\gamma mc^2}{\gamma mv} = \frac{c^2}{v}$$

where v is the particle's speed. Finally, substituting $v_p = c^2/v$ into Equation 5.25 for v_g shows that the group velocity of the matter wave packet is the same as the particle speed. This agrees with our intuition that the matter wave envelope should move at the same speed as the particle.

5.4 FOURIER INTEGRALS

In this section we show in detail how to construct wave groups, or pulses, that are truly localized in space or time and also show that very general reciprocity relations of the type $\Delta k \Delta x \approx 1$ and $\Delta \omega \Delta t \approx 1$ hold for these pulses.

To form a true pulse that is zero everywhere outside of a finite spatial range Δx requires adding together an infinite number of harmonic waves with continuously varying wavelengths and amplitudes. This addition can be done with a Fourier integral, which is defined as follows:

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} a(k) e^{ikx} dk \quad (5.26)$$

Here $f(x)$ is a spatially localized wave group, $a(k)$ gives the amount or amplitude of the wave with wavenumber $k = (2\pi/\lambda)$ to be added, and $e^{ikx} = \cos kx + i \sin kx$ is Euler's compact expression for a harmonic wave. The amplitude distribution function $a(k)$ can be obtained if $f(x)$ is known by using the symmetric formula

$$a(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x) e^{-ikx} dx \quad (5.27)$$

Equations 5.26 and 5.27 apply to the case of a spatial pulse at fixed time, but it is important to note that they are mathematically identical to the case of a time pulse passing a fixed position. This case is common in electrical engineering and involves adding together a continuously varying set of frequencies:

$$V(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} g(\omega) e^{i\omega t} d\omega \quad (5.28)$$

$$g(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} V(t) e^{-i\omega t} dt \quad (5.29)$$

where $V(t)$ is the strength of a signal as a function of time, and $g(\omega)$ is the *spectral content* of the signal and gives the amount of the harmonic wave with frequency ω that is present.

Let us now consider several examples of how to use Equations 5.26 through 5.29 and how they lead to uncertainty relationships of the type $\Delta \omega \Delta t \approx 1$ and $\Delta k \Delta x \approx 1$.

EXAMPLE 5.6

This example compares the spectral contents of infinite and truncated sinusoidal waves. A truncated sinusoidal wave is a wave cut off or truncated by a shutter, as shown in Figure 5.21. (a) What is the spectral content of an infinite sinusoidal wave $e^{i\omega_0 t}$? (b) Find and sketch the spectral content of a truncated sinusoidal wave given by

$$V(t) = e^{i\omega_0 t} \quad -T < t < +T$$

$$V(t) = 0 \quad \text{otherwise}$$

(c) Show that for this truncated sinusoid $\Delta t \Delta \omega = \pi$, where Δt and $\Delta \omega$ are the half-widths of $v(t)$ and $g(\omega)$, respectively.

Solution (a) The spectral content consists of a single strong contribution at the frequency ω_0 .

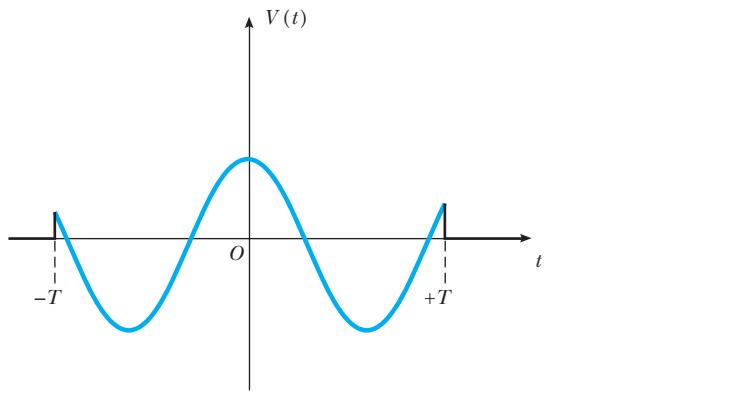


Figure 5.21 (Example 5.6) The real part of a truncated sinusoidal wave.

$$(b) \quad g(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} V(t) e^{-i\omega t} dt = \frac{1}{\sqrt{2\pi}} \int_{-T}^{+T} e^{i(\omega_0 - \omega)t} dt \\ = \frac{1}{\sqrt{2\pi}} \int_{-T}^{+T} [\cos(\omega_0 - \omega)t + i \sin(\omega_0 - \omega)t] dt$$

Because the sine term is an odd function and the cosine is even, the integral reduces to

$$g(\omega) = \frac{2}{\sqrt{2\pi}} \int_0^T \cos(\omega_0 - \omega)t dt = \sqrt{\frac{2}{\pi}} \frac{\sin(\omega_0 - \omega)T}{(\omega_0 - \omega)} = \sqrt{\frac{2}{\pi}} (T) \frac{\sin(\omega_0 - \omega)T}{(\omega_0 - \omega)T}$$

A sketch of $g(\omega)$ (Figure 5.22) shows a typical sin Z/Z profile centered on ω_0 . Note that both positive and negative amounts of different frequencies must be added to produce the truncated sinusoid. Furthermore, the strongest frequency contribution comes from the frequency region near $\omega = \omega_0$, as expected.

(c) Δt clearly equals T and $\Delta\omega$ may be taken to be half the width of the main lobe of $g(\omega)$, $\Delta\omega = \pi/T$. Thus, we get

$$\Delta\omega \Delta t = \frac{\pi}{T} \times T = \pi$$

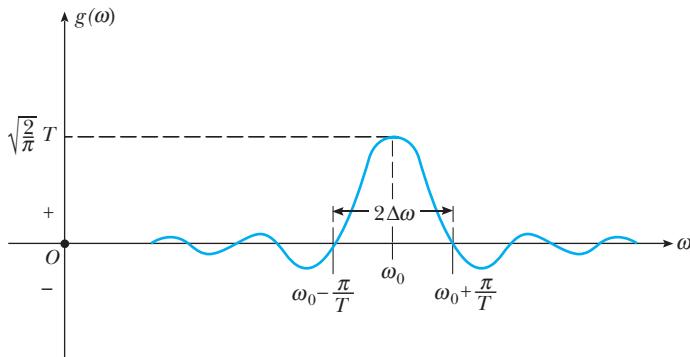


Figure 5.22 (Example 5.6) The Fourier transform of a truncated sinusoidal wave. The curve shows the amount of a given frequency that must be added to produce the truncated wave.

We see that the product of the spread in frequency, $\Delta\omega$, and the spread in time, Δt , is a constant independent of T .

EXAMPLE 5.7 A Matter Wave Packet

(a) Show that the matter wave packet whose amplitude distribution $a(k)$ is a rectangular pulse of height unity, width Δk , and centered at k_0 (Fig. 5.23) has the form

$$f(x) = \frac{\Delta k}{\sqrt{2\pi}} \frac{\sin(\Delta k \cdot x/2)}{(\Delta k \cdot x/2)} e^{ik_0 x}$$

Solution

$$\begin{aligned} f(x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} a(k) e^{ikx} dk = \frac{1}{\sqrt{2\pi}} \int_{k_0 - (\Delta k/2)}^{k_0 + (\Delta k/2)} e^{ikx} dk = \frac{1}{\sqrt{2\pi}} \frac{e^{ik_0 x}}{x} 2 \sin(\Delta k \cdot x/2) \\ &= \frac{\Delta k}{\sqrt{2\pi}} \frac{\sin(\Delta k \cdot x/2)}{(\Delta k \cdot x/2)} e^{ik_0 x} \end{aligned}$$

(b) Observe that this wave packet is a complex function. Later in this chapter we shall see how the definition of probability density results in a real function, but for the time being consider only the real part of $f(x)$ and make a sketch of its behavior, showing its envelope and the cosine function within. Determine Δx , and show that an uncertainty relation of the form $\Delta x \Delta k \approx 1$ holds.

Solution The real part of the wave packet is shown in Figure 5.24 where the full width of the main lobe is $\Delta x = 4\pi/\Delta k$. This immediately gives the uncertainty relation $\Delta x \Delta k = 4\pi$. Note that the constant on the right-hand side of the uncertainty relation depends on the shape chosen for $a(k)$ and the precise definition of Δx and Δk .

Exercise 3 Assume that a narrow triangular voltage pulse $V(t)$ arises in some type of radar system (see Fig. 5.25). (a) Find and sketch the spectral content $g(\omega)$. (b) Show that a relation of the type $\Delta\omega \Delta t \approx 1$ holds. (c) If the width of the pulse is

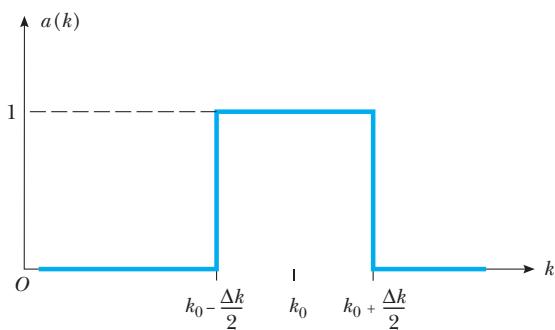


Figure 5.23 (Example 5.7) A simple amplitude distribution specifying a uniform contribution of all wavenumbers from $k_0 - \Delta k/2$ to $k_0 + \Delta k/2$. Although we have used only positive k 's here, both positive and negative k values are allowed, in general corresponding to waves traveling to the right ($k > 0$) or left ($k < 0$).

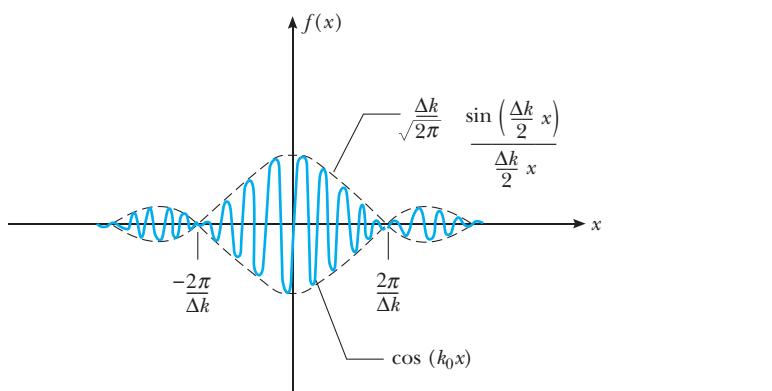


Figure 5.24 (Example 5.7) The real part of the wave packet formed by the uniform amplitude distribution shown in Figure 5.23.

$2\tau = 10^{-9}$ s, what range of frequencies must this system pass if the pulse is to be undistorted? Take $\Delta t = \tau$ and define $\Delta\omega$ similarly.

Answer (a) $g(\omega) = (\sqrt{2/\pi})(1/\omega^2\tau)(1 - \cos \omega\tau)$. (b) $\Delta\omega \Delta t = 2\pi$. (c) $2\Delta f = 4.00 \times 10^9$ Hz.

Constructing Moving Wave Packets

Figure 5.24 represents a snapshot of the wave packet at $t = 0$. To construct a moving wave packet representing a moving particle, we replace kx in Equation 5.26 with $(kx - \omega t)$. Thus, the representation of the moving wave packet becomes

$$f(x, t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} a(k) e^{i(kx - \omega t)} dk \quad (5.30)$$

It is important to realize that here $\omega = \omega(k)$, that is, ω is a function of k and therefore depends on the type of wave and the medium traversed. In general, it is difficult to solve this integral analytically. For matter waves, the QMTools software available from our companion Web site (<http://info.brookscole.com/mp3e>) produces the same result by solving numerically a certain differential equation that governs the behavior of such waves. This approach will be explored further in the next chapter.

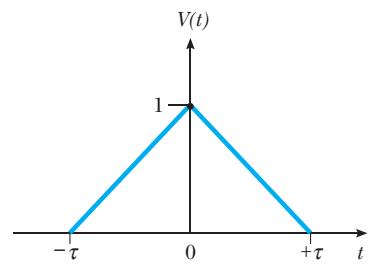


Figure 5.25 (Exercise 3).

5.5 THE HEISENBERG UNCERTAINTY PRINCIPLE

In the period 1924–25, Werner Heisenberg, the son of a professor of Greek and Latin at the University of Munich, invented a complete theory of quantum mechanics called matrix mechanics. This theory overcame some of the problems with the Bohr theory of the atom, such as the postulate of “unobservable” electron orbits. Heisenberg’s formulation was based primarily on measurable quantities such as the transition probabilities for electronic jumps between quantum states. Because transition probabilities depend on the initial and final states, Heisenberg’s mechanics used variables labeled by two subscripts. Although at first Heisenberg presented his theory in the form of non-commuting algebra, Max Born quickly realized that this theory could be more

elegantly described by matrices. Consequently, Born, Heisenberg, and Pascual Jordan soon worked out a comprehensive theory of matrix mechanics. Although the matrix formulation was quite elegant, it attracted little attention outside of a small group of gifted physicists because it was difficult to apply in specific cases, involved mathematics unfamiliar to most physicists, and was based on rather vague physical concepts.

Although we will investigate this remarkable form of quantum mechanics no further, we shall discuss another of Heisenberg's discoveries, the **uncertainty principle**, elucidated in a famous paper in 1927. In this paper Heisenberg introduced the notion that **it is impossible to determine simultaneously with unlimited precision the position and momentum of a particle**. In words we may state the uncertainty principle as follows:

If a measurement of position is made with precision Δx and a simultaneous measurement of momentum in the x direction is made with precision Δp_x , then the product of the two uncertainties can never be smaller than $\hbar/2$. That is,

$$\Delta p_x \Delta x \geq \frac{\hbar}{2} \quad (5.31)$$

In his paper of 1927, Heisenberg was careful to point out that the inescapable uncertainties Δp_x and Δx do not arise from imperfections in practical measuring instruments. Rather, they arise from the need to use a large range of wavenumbers, Δk , to represent a matter wave packet localized in a small region, Δx . The uncertainty principle represents a sharp break with the ideas of classical physics, in which it is assumed that, with enough skill and ingenuity, it is possible to simultaneously measure a particle's position and momentum to any desired degree of precision. As shown in Example 5.8, however, there is no contradiction between the uncertainty principle and classical laws for macroscopic systems because of the small value of \hbar .

One can show that $\Delta p_x \Delta x \geq \hbar/2$ comes from the uncertainty relation governing any type of wave pulse formed by the superposition of waves with different wavelengths. In Section 5.3 we found that to construct a wave group localized in a small region Δx , we had to add up a large range of wavenumbers Δk , where $\Delta k \Delta x \approx 1$ (Eq. 5.16). The precise value of the number on the right-hand side of Equation 5.16 depends on the functional form $f(x)$ of the wave group as well as on the specific definition of Δx and Δk . A different choice of $f(x)$ or a different rule for defining Δx and Δk (or both) will give a slightly different number. With Δx and Δk defined as standard deviations, it can be shown that the smallest number, $\frac{1}{2}$, is obtained for a Gaussian wavefunction.⁴ In this minimum uncertainty case we have

$$\Delta x \Delta k = \frac{1}{2}$$

⁴See Section 6.7 for a definition of the standard deviation and Problem 6.34 for a complete mathematical proof of this statement.

This photograph of Werner Heisenberg was taken around 1924. Heisenberg obtained his Ph.D. in 1923 at the University of Munich where he studied under Arnold Sommerfeld and became an enthusiastic mountain climber and skier. Later, he worked as an assistant to Max Born at Göttingen and Niels Bohr in Copenhagen. While physicists such as de Broglie and Schrödinger tried to develop visualizable models of the atom, Heisenberg, with the help of Born and Pascual Jordan, developed an abstract mathematical model called matrix mechanics to explain the wavelengths of spectral lines. The more successful wave mechanics of Schrödinger an-

Image not available due to copyright restrictions

WERNER HEISENBERG

(1901–1976)

nounced a few months later was shown to be equivalent to Heisenberg's approach. Heisenberg made many other significant contributions to physics, including his famous uncertainty principle, for which he received the Nobel prize in 1932, the prediction of two forms of molecular hydrogen, and theoretical models of the nucleus. During World War II he was director of the Max Planck Institute at Berlin where he was in charge of German research on atomic weapons. Following the war, he moved to West Germany and became director of the Max Planck Institute for Physics at Göttingen.

For any other choice of $f(x)$,

$$\Delta x \Delta k \geq \frac{1}{2} \quad (5.32)$$

and using $\Delta p_x = \hbar \Delta k$, $\Delta x \Delta k \geq \frac{1}{2}$ immediately becomes

$$\Delta p_x \Delta x \geq \frac{\hbar}{2} \quad (5.33)$$

The basic meaning of $\Delta p \Delta x \geq \hbar/2$ is that as one uncertainty increases the other decreases. In the extreme case as one uncertainty approaches ∞ , the other must approach zero. This extreme case is illustrated by a plane wave $e^{ik_0 x}$ that has a precise momentum $\hbar k_0$ and an infinite extent—that is, the wavefunction is not concentrated in any segment of the x axis.

Another important uncertainty relation involves the uncertainty in energy of a wave packet, ΔE , and the time, Δt , taken to measure that energy. Starting with $\Delta \omega \Delta t \geq \frac{1}{2}$ as the minimum form of the time–frequency uncertainty principle, and using the de Broglie relation for the connection between the matter wave energy and frequency, $E = \hbar \omega$, we immediately find the **energy–time uncertainty principle**

$$\Delta E \Delta t \geq \frac{\hbar}{2} \quad (5.34)$$

Energy–time uncertainty principle

Equation 5.34 states that the precision with which we can know the energy of some system is limited by the time available for measuring the energy. A common application of the energy–time uncertainty is in calculating the lifetimes of very short-lived subatomic particles whose lifetimes cannot be measured directly, but whose uncertainty in energy or mass can be measured. (See Problem 26.)

A Different View of the Uncertainty Principle

Although we have indicated that $\Delta p_x \Delta x \geq \hbar/2$ arises from the theory of forming pulses or wave groups, there is a more physical way to view the origin of the un-

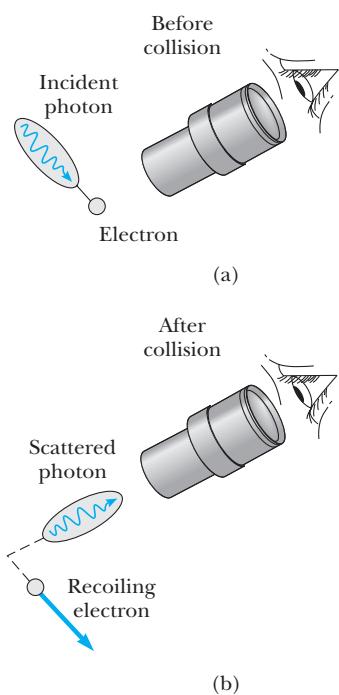


Figure 5.26 A thought experiment for viewing an electron with a powerful microscope. (a) The electron is shown before colliding with the photon. (b) The electron recoils (is disturbed) as a result of the collision with the photon.

certainty principle. We consider certain idealized experiments (called *thought experiments*) and show that it is impossible to carry out an experiment that allows the position and momentum of a particle to be simultaneously measured with an accuracy that violates the uncertainty principle. The most famous thought experiment along these lines was introduced by Heisenberg himself and involves the measurement of an electron's position by means of a microscope (Fig. 5.26), which forms an image of the electron on a screen or the retina of the eye.

Because light can scatter from and perturb the electron, let us minimize this effect by considering the scattering of only a single light quantum from an electron initially at rest (Fig. 5.27). To be collected by the lens, the photon must be scattered through an angle ranging from $-\theta$ to $+\theta$, which consequently imparts to the electron an x momentum value ranging from $+(h \sin \theta)/\lambda$ to $-(h \sin \theta)/\lambda$. Thus the uncertainty in the electron's momentum is $\Delta p_x = (2h \sin \theta)/\lambda$. After passing through the lens, the photon lands somewhere on the screen, but the image and consequently the position of the electron is "fuzzy" because the photon is diffracted on passing through the lens aperture. According to physical optics, the resolution of a microscope or the uncertainty in the image of the electron, Δx , is given by $\Delta x = \lambda/(2 \sin \theta)$. Here 2θ is the angle subtended by the objective lens, as shown in Figure 5.27.⁵ Multiplying the expressions for Δp_x and Δx , we find for the electron

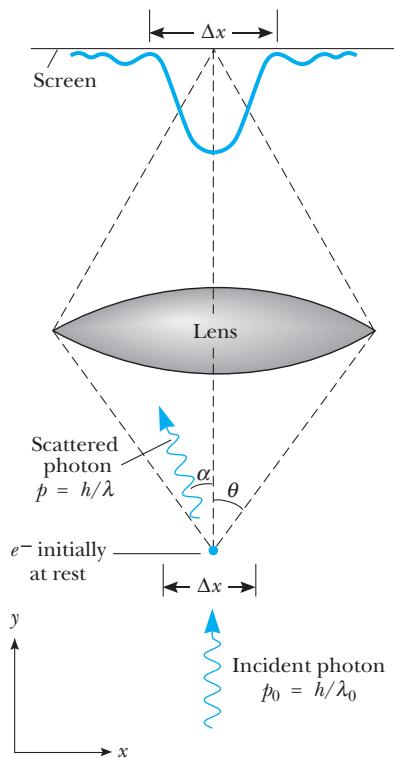


Figure 5.27 The Heisenberg microscope.

⁵The resolving power of the microscope is treated clearly in F. A. Jenkins and H. E. White, *Fundamentals of Optics*, 4th ed., New York, McGraw-Hill Book Co., 1976, pp. 332–334.

$$\Delta p_x \Delta x \approx \left(\frac{2\hbar}{\lambda} \sin \theta \right) \left(\frac{\lambda}{2 \sin \theta} \right) = \hbar$$

in agreement with the uncertainty relation. Note also that this principle is inescapable and relentless! If Δx is reduced by increasing θ or the lens size, there is an equivalent increase in the uncertainty of the electron's momentum.

Examination of this simple experiment shows several key physical properties that lead to the uncertainty principle:

- The indivisible nature of light particles or quanta (nothing less than a single photon can be used!).
- The wave property of light as shown in diffraction.
- The impossibility of predicting or measuring the precise classical path of a single scattered photon and hence of knowing the precise momentum transferred to the electron.⁶

We conclude this section with some examples of the types of calculations that can be done with the uncertainty principle. In the spirit of Fermi or Heisenberg, these “back-of-the-envelope calculations” are surprising for their simplicity and essential description of quantum systems of which the details are unknown.

EXAMPLE 5.8 The Uncertainty Principle Changes Nothing for Macroscopic Objects

(a) Show that the spread of velocities caused by the uncertainty principle does not have measurable consequences for macroscopic objects (objects that are large compared with atoms) by considering a 100-g racquetball confined to a room 15 m on a side. Assume the ball is moving at 2.0 m/s along the x axis.

Solution

$$\Delta p_x \geq \frac{\hbar}{2 \Delta x} = \frac{1.05 \times 10^{-34} \text{ J}\cdot\text{s}}{2 \times 15 \text{ m}} = 3.5 \times 10^{-36} \text{ kg}\cdot\text{m/s}$$

Thus the minimum spread in velocity is given by

$$\Delta v_x = \frac{\Delta p_x}{m} = \frac{3.5 \times 10^{-36} \text{ kg}\cdot\text{m/s}}{0.100 \text{ kg}} = 3.5 \times 10^{-35} \text{ m/s}$$

This gives a relative uncertainty of

$$\frac{\Delta v_x}{v_x} = \frac{3.5 \times 10^{-35}}{2.0} = 1.8 \times 10^{-35}$$

which is certainly not measurable.

(b) If the ball were to suddenly move along the y axis perpendicular to its well-defined classical trajectory along x , how far would it move in 1 s? Assume that the ball moves in the y direction with the top speed in the spread Δv_y produced by the uncertainty principle.

Solution It is important to realize that uncertainty relations hold in the y and z directions as well as in the x direction. This means that $\Delta p_x \Delta x \geq \hbar/2$, $\Delta p_y \Delta y \geq \hbar/2$, and $\Delta p_z \Delta z \geq \hbar/2$ and because all the position uncertainties are equal, all of the velocity spreads are equal. Consequently, we have $\Delta v_y = 3.5 \times 10^{-35} \text{ m/s}$ and the ball moves $3.5 \times 10^{-35} \text{ m}$ in the y direction in 1 s. This distance is again an immeasurably small quantity, being 10^{-20} times the size of a nucleus!

Exercise 4 How long would it take the ball to move 50 cm in the y direction? (The age of the universe is thought to be 15 billion years, give or take a few billion).

⁶Attempts to measure the photon's position by scattering electrons from it in a Compton process only serve to make its path to the lens more uncertain.

EXAMPLE 5.9 Do Electrons Exist Within the Nucleus?

Estimate the kinetic energy of an electron confined within a nucleus of size 1.0×10^{-14} m by using the uncertainty principle.

Solution Taking Δx to be the half-width of the confinement length in the equation $\Delta p_x \geq \frac{\hbar}{2 \Delta x}$, we have

$$\Delta p_x \geq \frac{6.58 \times 10^{-16} \text{ eV}\cdot\text{s}}{1.0 \times 10^{-14} \text{ m}} \times \frac{3.00 \times 10^8 \text{ m/s}}{c}$$

or

$$\Delta p_x \geq 2.0 \times 10^7 \frac{\text{eV}}{c}$$

This means that measurements of the component of momentum of electrons trapped inside a nucleus would range from less than $-20 \text{ MeV}/c$ to greater than $+20 \text{ MeV}/c$ and that some electrons would have momentum at least as large as $20 \text{ MeV}/c$. Because this appears to be a large momentum, to be safe we calculate the electron's energy relativistically.

$$\begin{aligned} E^2 &= p^2 c^2 + (m_e c^2)^2 \\ &= (20 \text{ MeV}/c)^2 c^2 + (0.511 \text{ MeV})^2 \\ &= 400(\text{MeV})^2 \end{aligned}$$

or

$$E \geq 20 \text{ MeV}$$

Finally, the kinetic energy of an intranuclear electron is

$$K = E - m_e c^2 \geq 19.5 \text{ MeV}$$

Since electrons emitted in radioactive decay of the nucleus (beta decay) have energies much less than 19.5 MeV (about 1 MeV or less) and it is known that no other mechanism could carry off an intranuclear electron's energy during the decay process, we conclude that electrons observed in beta decay do not come from within the nucleus but are actually created at the instant of decay.

EXAMPLE 5.10 The Width of Spectral Lines

Although an excited atom can radiate at any time from $t = 0$ to $t = \infty$, the average time after excitation at which a group of atoms radiates is called the **lifetime**, τ , of

a particular excited state. (a) If $\tau = 1.0 \times 10^{-8}$ s (a typical value), use the uncertainty principle to compute the line width Δf of light emitted by the decay of this excited state.

Solution We use $\Delta E \Delta t \approx \hbar/2$, where ΔE is the uncertainty in energy of the excited state, and $\Delta t = 1.0 \times 10^{-8}$ s is the average time available to measure the excited state. Thus,

$$\Delta E \approx \hbar/2 \Delta t = \hbar/(2.0 \times 10^{-8} \text{ s})$$

Since ΔE is also the uncertainty in energy of a photon emitted when the excited state decays, and $\Delta E = \hbar \Delta f$ for a photon,

$$\hbar \Delta f = \hbar/(2.0 \times 10^{-8} \text{ s})$$

or

$$\Delta f = \frac{1}{4\pi \times 10^{-8} \text{ s}} = 8.0 \times 10^6 \text{ Hz}$$

(b) If the wavelength of the spectral line involved in this process is 500 nm, find the fractional broadening $\Delta f/f$.

Solution First, we find the center frequency of this line as follows:

$$f_0 = \frac{c}{\lambda} = \frac{3.0 \times 10^8 \text{ m/s}}{500 \times 10^{-9} \text{ m}} = 6.0 \times 10^{14} \text{ Hz}$$

Hence,

$$\frac{\Delta f}{f_0} = \frac{8.0 \times 10^6 \text{ Hz}}{6.0 \times 10^{14} \text{ Hz}} = 1.3 \times 10^{-8}$$

This narrow natural line width can be seen with a sensitive interferometer. Usually, however, temperature and pressure effects overshadow the natural line width and broaden the line through mechanisms associated with the Doppler effect and atomic collisions.

Exercise 5 Using the nonrelativistic Doppler formula, calculate the Doppler broadening of a 500-nm line emitted by a hydrogen atom at 1000 K. Do this by considering the atom to be moving either directly toward or away from an observer with an energy of $\frac{3}{2}k_B T$.

Answer 0.0083 nm, or 0.083 Å.

5.6 IF ELECTRONS ARE WAVES, WHAT'S WAVING?

Although we have discussed in some detail the notion of de Broglie matter waves, we have not discussed the precise nature of the field $\Psi(x, y, z, t)$ or **wavefunction** that represents the matter waves. We have delayed this discussion because Ψ

(Greek letter psi) is rather abstract. Ψ is definitely *not* a measurable disturbance requiring a medium for propagation like a water wave or a sound wave. Instead, the stuff that is waving requires no medium. Furthermore, Ψ is in general represented by a complex number and is used to calculate the probability of finding the particle at a given time in a small volume of space. If any of this seems confusing, you should not lose heart, as the nature of the wavefunction has been confusing people since its invention. It even confused its inventor, Erwin Schrödinger, who incorrectly interpreted $\Psi^*\Psi$ as the electric charge density.⁷ The great philosopher of the quantum theory, Bohr, immediately objected to this interpretation. Subsequently, Max Born offered the currently accepted statistical view of $\Psi^*\Psi$ in late 1926. The confused state of affairs surrounding Ψ at that time was nicely described in a poem by Walter Huckel:

Erwin with his psi can do
Calculations quite a few.
But one thing has not been seen
Just what does psi really mean?
(English translation by Felix Bloch)

The currently held view is that a particle is described by a function $\Psi(x, y, z, t)$ called the **wavefunction**. The quantity $\Psi^*\Psi = |\Psi|^2$ represents the probability per unit volume of finding the particle at a time t in a small volume of space centered on (x, y, z) . We will treat methods of finding Ψ in much more detail in Chapter 6, but for now all we require is the idea that **the probability of finding a particle is directly proportional to $|\Psi|^2$** .

5.7 THE WAVE–PARTICLE DUALITY

The Description of Electron Diffraction in Terms of Ψ

In this chapter and previous chapters we have seen evidence for both the wave properties and the particle properties of electrons. Historically, the particle properties were first known and connected with a definite mass, a discrete charge, and detection or localization of the electron in a small region of space. Following these discoveries came the confirmation of the wave nature of electrons in scattering at low energy from metal crystals. In view of these results and because of the everyday experience of seeing the world in terms of *either* grains of sand *or* diffuse water waves, it is no wonder that we are tempted to simplify the issue and ask, “Well, is the electron a wave or a particle?” The answer is that **electrons are very delicate and rather plastic—they behave like either particles or waves, depending on the kind of experiment performed on them. In any case, it is impossible to measure both the wave and particle properties simultaneously.**⁸ The view of Bohr was expressed in an idea known as **complementarity**. As different as they are, both wave and particle views are needed and they complement each other to fully describe the electron. The view of

Complementarity

⁷ Ψ^* represents the complex conjugate of Ψ . Thus, if $\Psi = a + ib$, then $\Psi^* = a - ib$. In exponential form, if $\Psi = Ae^{i\theta}$, then $\Psi^* = Ae^{-i\theta}$. Note that $\Psi^*\Psi = |\Psi|^2$; a , b , A , and θ are all real quantities.

⁸Many feel that the elder Bragg's remark, originally made about light, is a more satisfying answer: Electrons behave like waves on Mondays, Wednesdays, and Fridays, like particles on Tuesdays, Thursdays, and Saturdays, and like nothing at all on Sundays.

Feynman⁹ was that both electrons and photons behave in their own inimitable way. This is like nothing we have seen before, because we do not live at the very tiny scale of atoms, electrons, and photons.

Perhaps the best way to crystallize our ideas about the wave–particle duality is to consider a “simple” double-slit electron diffraction experiment. This experiment highlights much of the mystery of the wave–particle paradox, shows the impossibility of measuring *simultaneously* both wave and particle properties, and illustrates the use of the wavefunction, Ψ , in determining interference effects. A schematic of the experiment with monoenergetic (single-wavelength) electrons is shown in Figure 5.28. A parallel beam of electrons falls on a double slit, which has individual openings much smaller than D so that single-slit diffraction effects are negligible. At a distance from the slits much greater than D is an electron detector capable of detecting individual electrons. It is important to note that the detector always registers discrete particles localized in space and time. In a real experiment this can be achieved if the electron source is weak enough (see Fig. 5.29): **In all cases if the detector collects electrons at different positions for a long enough time, a typical wave interference pattern for the counts per minute or probability of arrival of electrons is found** (see Fig. 5.28). If one imagines a single electron to produce in-phase “wavelets” at the slits, standard wave theory can be used to find the angular separation, θ , of the

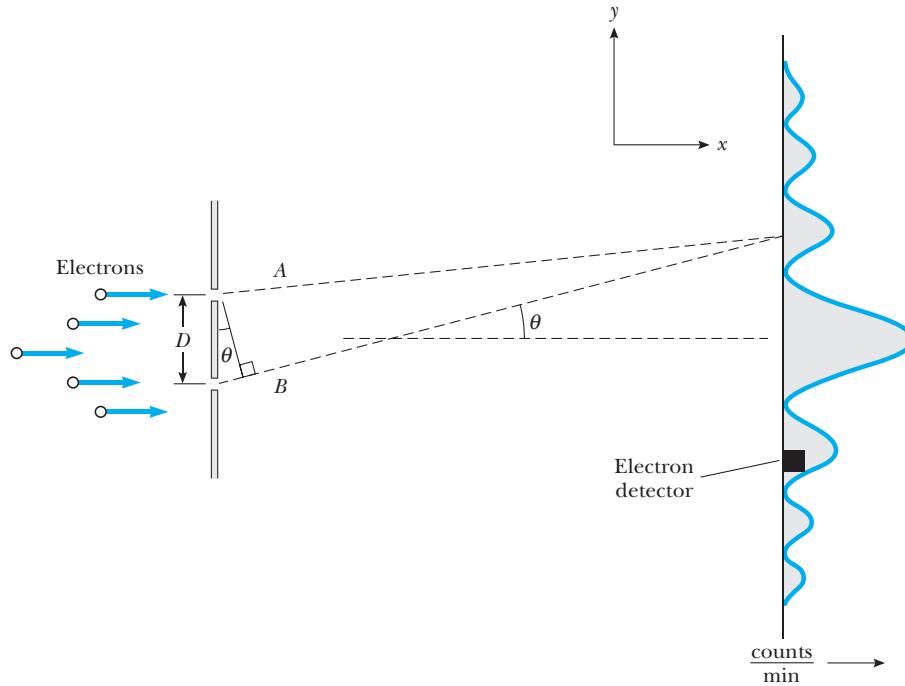


Figure 5.28 Electron diffraction. D is much greater than the individual slit widths and much less than the distance between the slits and the detector.

⁹R. Feynman, *The Character of Physical Law*, Cambridge, MA, MIT Press, 1982.

Images not available due to copyright restrictions

central probability maximum from its neighboring minimum. The minimum occurs when the path length difference between A and B in Figure 5.28 is half a wavelength, or

$$D \sin \theta = \lambda/2$$

As the electron's wavelength is given by $\lambda = h/p_x$, we see that

$$\sin \theta \approx \theta = \frac{h}{2p_x D} \quad (5.35)$$

for small θ . Thus we can see that the dual nature of the electron is clearly shown in this experiment: **although the electrons are detected as particles at a localized spot at some instant of time, the probability of arrival at that spot is determined by finding the intensity of two interfering matter waves.**

But there is more. What happens if one slit is covered during the experiment? In this case one obtains a symmetric curve peaked around the center of the open slit, much like the pattern formed by bullets shot through a hole in armor plate. Plots of the counts per minute or probability of arrival of electrons with the lower or upper slit *closed* are shown in Figure 5.30. These are expressed as the appropriate square of the absolute value of some wavefunction, $|\Psi_1|^2 = \Psi_1^* \Psi_1$ or $|\Psi_2|^2 = \Psi_2^* \Psi_2$, where Ψ_1 and Ψ_2 represent the cases of the electron passing through slit 1 and slit 2, respectively. If an experiment is now performed with slit 1 open and slit 2 blocked for time T and then slit 1 blocked and slit 2 open for time T , the accumulated pattern of counts per minute is completely different from the case with

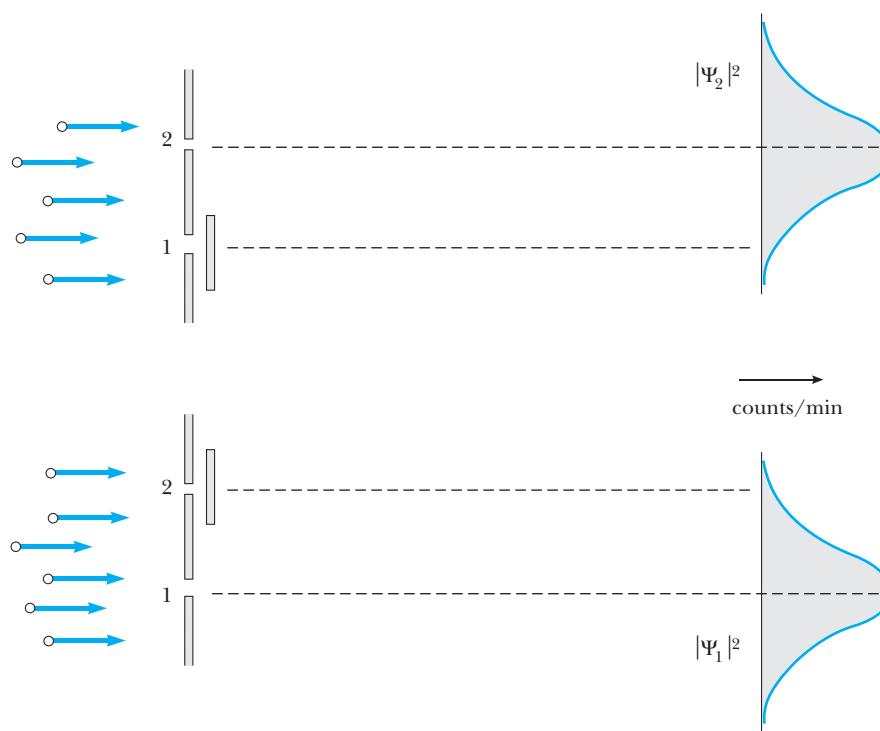


Figure 5.30 The probability of finding electrons at the screen with either the lower or upper slit closed.

both slits open. Note in Figure 5.31 that there is no longer a maximum probability of arrival of an electron at $\theta = 0$. In fact, **the interference pattern has been lost and the accumulated result is simply the sum of the individual results**. The results shown by the black curves in Figure 5.31 are easier to understand and more reasonable than the interference effects seen with both slits open (blue curve). When only one slit is open at a time, we know the electron has the same localizability and indivisibility at the slits as we measure at the detector, because the electron clearly goes through slit 1 or slit 2. Thus, the total must be analyzed as the sum of those electrons that come through slit 1, $|\Psi_1|^2$, and those that come through slit 2, $|\Psi_2|^2$. When both slits are open, it is tempting to assume that the electron goes through either slit 1 or slit 2 and that the counts per minute are again given by $|\Psi_1|^2 + |\Psi_2|^2$. We know, however, that the experimental results contradict this. Thus, our assumption that the electron is localized and goes through only one slit when both slits are open must be wrong (a painful conclusion!). Somehow the electron must be simultaneously present at both slits in order to exhibit interference.

To find the probability of detecting the electron at a particular point on the screen with both slits open, we may say that the electron is in a *superposition state* given by

$$\Psi = \Psi_1 + \Psi_2$$

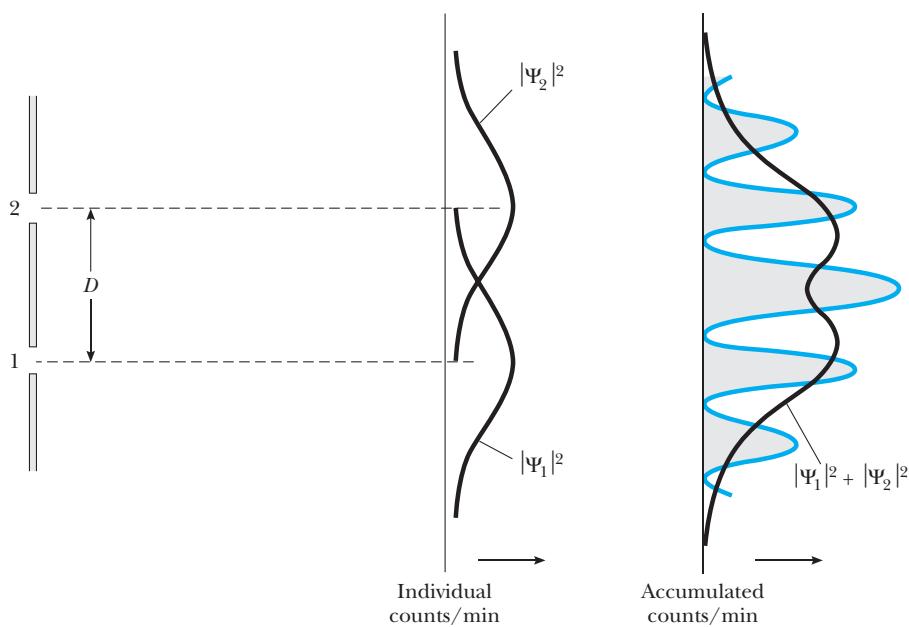


Figure 5.31 Accumulated results from the two-slit electron diffraction experiment with each slit closed half the time. For comparison, the results with both slits open are shown in color.

Thus, the probability of detecting the electron at the screen is equal to the quantity $|\Psi_1 + \Psi_2|^2$ and not $|\Psi_1|^2 + |\Psi_2|^2$. Because matter waves that start out in phase at the slits in general travel different distances to the screen (see Fig. 5.28), Ψ_1 and Ψ_2 will possess a relative phase difference ϕ at the screen. Using a phasor diagram (Fig. 5.32) to find $|\Psi_1 + \Psi_2|^2$ immediately yields

$$|\Psi|^2 = |\Psi_1 + \Psi_2|^2 = |\Psi_1|^2 + |\Psi_2|^2 + 2|\Psi_1||\Psi_2| \cos\phi$$

Note that the term $2|\Psi_1||\Psi_2| \cos\phi$ is an interference term that predicts the interference pattern actually observed in this case. For ease of comparison, a summary of the results found in both cases is given in Table 5.1.

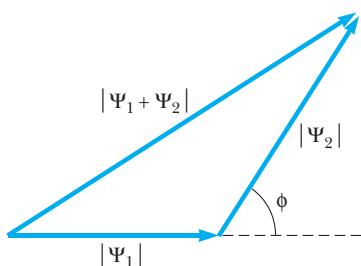


Figure 5.32 Phasor diagram to represent the addition of two complex wavefunctions, Ψ_1 and Ψ_2 , differing in phase by ϕ .

Table 5.1

Case	Wavefunction	Counts/Minute at Screen
Electron is measured to pass through slit 1 or slit 2	Ψ_1 or Ψ_2	$ \Psi_1 ^2 + \Psi_2 ^2$
No measurements made on electron at slits	$\Psi_1 + \Psi_2$	$ \Psi_1 ^2 + \Psi_2 ^2 + 2 \Psi_1 \Psi_2 \cos \phi$

A Thought Experiment: Measuring Through Which Slit the Electron Passes

Another way to view the electron double-slit experiment is to say that the electron passes through the upper or lower slit only when one *measures* the electron to do so. Once one measures unambiguously which slit the electron passes through (yes, you guessed it . . . here comes the uncertainty principle again . . .), the act of measurement disturbs the electron's path enough to destroy the delicate interference pattern.

Let us look again at our two-slit experiment to see in detail how the interference pattern is destroyed.¹⁰ To determine which slit the electron goes through, imagine that a group of particles is placed right behind the slits, as shown in Figure 5.33. If we use the recoil of a small particle to determine

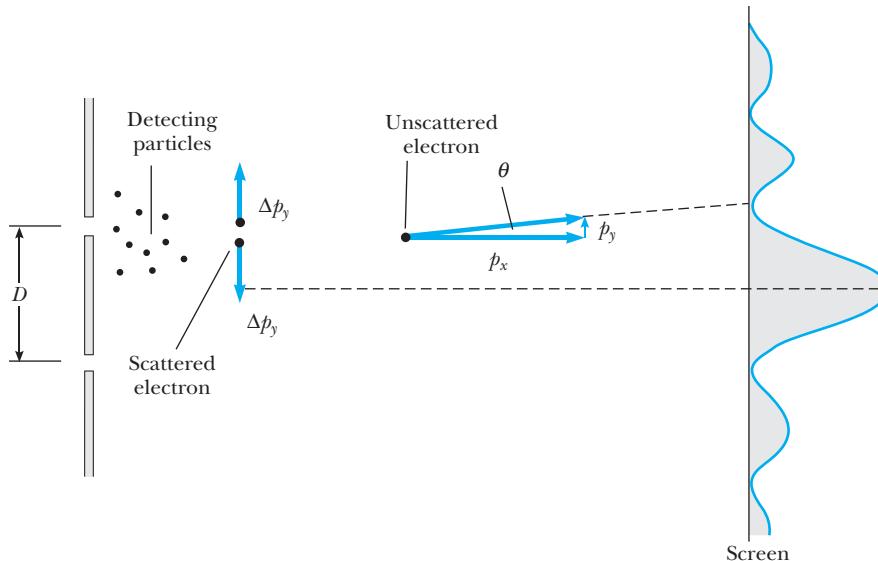


Figure 5.33 A thought experiment to determine through which slit the electron passes.

¹⁰Although we shall use the uncertainty principle in its standard form, it is worth noting that an alternative statement of the uncertainty principle involves this pivotal double-slit experiment: *It is impossible to design any device to determine through which slit the electron passes that will not at the same time disturb the electron and destroy the interference pattern.*

which slit the electron goes through, we must have the uncertainty in the detecting particle's position, $\Delta y \ll D$. Also, during the collision the detecting particle suffers a change in momentum, Δp_y , equal and opposite to the change in momentum experienced by the electron, as shown in Figure 5.33. An undeviated electron landing at the first minimum *and producing an interference pattern* has

$$\tan \theta \approx \theta = \frac{p_y}{p_x} = \frac{h}{2p_x D}$$

from Equation 5.35. Thus, we require that an electron scattered by a detecting particle have

$$\frac{\Delta p_y}{p_x} \ll \theta = \frac{h}{2p_x D}$$

or

$$\Delta p_y \ll \frac{h}{2D}$$

if the interference pattern is not to be distorted. Because the change in momentum of the scattered electron is equal to the change in momentum of the detecting particle, $\Delta p_y \ll h/2D$ also applies to the detecting particle. Thus, we have for the detecting particle

$$\Delta p_y \Delta y \ll \frac{h}{2D} \cdot D$$

or

$$\Delta p_y \Delta y \ll \frac{h}{2}$$

This is a clear violation of the uncertainty principle. Hence we see that **the small uncertainties needed, both to observe interference and to know which slit the electron goes through, are impossible, because they violate the uncertainty principle.** If Δy is small enough to determine which slit the electron goes through, Δp_y is so large that electrons heading for the first minimum are scattered into adjacent maxima and the interference pattern is destroyed.

Exercise 6 In a real experiment it is likely that some electrons would miss the detecting particles. Thus, we would really have two categories of electrons arriving at the detector: those measured to pass through a definite slit and those not observed, or just missed, at the slits. In this case what kind of pattern of counts per minute would be accumulated by the detector?

Answer A mixture of an interference pattern $|\Psi_1 + \Psi_2|^2$ (those not measured) and $|\Psi_1|^2 + |\Psi_2|^2$ (those measured) would result.

5.8 A FINAL NOTE

Scientists once viewed the world as being made up of distinct and unchanging parts that interact according to strictly deterministic laws of cause and effect. In the classical limit this is fundamentally correct because classical processes involve so many quanta that deviations from the average are imperceptible. At the atomic level, however, we have seen that a given piece of matter (an electron, say) is not a distinct and unchanging part of the universe obeying completely deterministic laws. Such a particle exhibits wave properties when it interacts with a metal crystal and particle properties a short while later when it registers on a Geiger counter. Thus, rather than viewing the electron as a distinct and separate part of the universe with an intrinsic particle nature, we are led to the view that the electron and indeed all particles are amorphous entities possessing the potential to cycle endlessly between wave and particle behavior. We also find that it is much more difficult to separate the object measured from the measuring instrument at the atomic level, because the type of measuring instrument determines whether wave properties or particle properties are observed.

SUMMARY

Every lump of matter of mass m and momentum p has wavelike properties with wavelength given by the de Broglie relation

$$\lambda = \frac{h}{p} \quad (5.1)$$

By applying this wave theory of matter to electrons in atoms, de Broglie was able to explain the appearance of integers in certain Bohr orbits as a natural consequence of electron wave interference. In 1927, Davisson and Germer demonstrated directly the wave nature of electrons by showing that low-energy electrons were diffracted by single crystals of nickel. In addition, they confirmed Equation 5.1.

Although the wavelength of matter waves can be experimentally determined, it is important to understand that they are not just like other waves because their frequency and phase velocity cannot be directly measured. In particular, the phase velocity of an individual matter wave is greater than the velocity of light and varies with wavelength or wavenumber as

$$v_p = f\lambda = \left(\frac{E}{h}\right)\left(\frac{h}{p}\right) = c \left[1 + \left(\frac{mc}{\hbar k}\right)^2\right]^{1/2} \quad (5.24)$$

To represent a particle properly, a superposition of matter waves with different wavelengths, amplitudes, and phases must be chosen to interfere constructively over a limited region of space. The resulting wave packet or group can then be shown to travel with the same speed as the classical particle. In addition, a wave packet localized in a region Δx contains a range of wavenumbers Δk , where $\Delta x \Delta k \geq \frac{1}{2}$. Because $p_x = \hbar k$, this implies that there is an uncertainty principle for position and momentum:

$$\Delta p_x \Delta x \geq \frac{\hbar}{2} \quad (5.31)$$

In a similar fashion one can show that an energy–time uncertainty relation exists, given by

$$\Delta E \Delta t \geq \frac{\hbar}{2} \quad (5.34)$$

In quantum mechanics matter waves are represented by a wavefunction $\Psi(x, y, z, t)$. The probability of finding a particle represented by Ψ in a small volume centered at (x, y, z) at time t is proportional to $|\Psi|^2$. The wave–particle duality of electrons may be seen by considering the passage of electrons through two narrow slits and their arrival at a viewing screen. We find that although the electrons are detected as particles at a localized spot on the screen, the probability of arrival at that spot is determined by finding the intensity of two interfering matter waves.

Although we have seen the importance of matter waves or wavefunctions in this chapter, we have provided no method of finding Ψ for a given physical system. In the next chapter we introduce the Schrödinger wave equation. The solutions to this important differential equation will provide us with the wavefunctions for a given system.

SUGGESTIONS FOR FURTHER READING

1. D. Bohm, *Quantum Theory*, Englewood Cliffs, NJ, Prentice-Hall, 1951. Chapters 3 and 6 in this book give an excellent account of wave packets and the wave–particle duality of matter at a more advanced level.
2. R. Feynman, *The Character of Physical Law*, Cambridge, MA, The MIT Press, 1982, Chapter 6. This monograph is an incredibly lively and readable treatment of the double-slit experiment presented in Feynman's inimitable fashion.
3. B. Hoffman, *The Strange Story of the Quantum*, New York, Dover Publications, 1959. This short book presents a beautifully written nonmathematical discussion of the history of quantum mechanics.

QUESTIONS

1. Is light a wave or a particle? Support your answer by citing specific experimental evidence.
2. Is an electron a particle or a wave? Support your answer by citing some experimental results.
3. An electron and a proton are accelerated from rest through the same potential difference. Which particle has the longer wavelength?
4. If matter has a wave nature, why is this wavelike character not observable in our daily experiences?
5. In what ways does Bohr's model of the hydrogen atom violate the uncertainty principle?
6. Why is it impossible to measure the position and speed of a particle simultaneously with infinite precision?
7. Suppose that a beam of electrons is incident on three or more slits. How would this influence the interference pattern? Would the state of an electron depend on the number of slits? Explain.
8. In describing the passage of electrons through a slit and arriving at a screen, Feynman said that "electrons arrive in lumps, like particles, but the probability of arrival of these lumps is determined as the intensity of the waves would be. It is in this sense that the electron behaves sometimes like a particle and sometimes like a wave." Elaborate on this point in your own words. (For a further discussion of this point, see R. Feynman, *The Character of Physical Law*, Cambridge, MA, MIT Press, 1982, Chapter 6.)
9. Do you think that most major experimental discoveries are made by careful planning or by accident? Cite examples.
10. In the case of accidental discoveries, what traits must the experimenter possess to capitalize on the discovery?
11. Are particles even things? An extreme view of the plasticity of electrons and other particles is expressed in this fa-

mous quote of Heisenberg: “The invisible elementary particle of modern physics does not have the property of occupying space any more than it has properties like color or solidity. Fundamentally, it is not a material structure in space and time but only a symbol that allows the laws of nature to be expressed in especially simple form.”

Are you satisfied with viewing science as a set of predictive rules or do you prefer to see science as a description of an objective world of things—in the case of particle physics, tiny, scaled-down things? What problems are associated with each point of view?

PROBLEMS

5.1 The Pilot Waves of de Broglie

- Calculate the de Broglie wavelength for a proton moving with a speed of 10^6 m/s .
- Calculate the de Broglie wavelength for an electron with kinetic energy (a) 50 eV and (b) 50 keV.
- Calculate the de Broglie wavelength of a 74-kg person who is running at a speed of 5.0 m/s.
- The “seeing” ability, or resolution, of radiation is determined by its wavelength. If the size of an atom is of the order of 0.1 nm, how fast must an electron travel to have a wavelength small enough to “see” an atom?
- To “observe” small objects, one measures the diffraction of particles whose de Broglie wavelength is approximately equal to the object’s size. Find the kinetic energy (in electron volts) required for electrons to resolve (a) a large organic molecule of size 10 nm, (b) atomic features of size 0.10 nm, and (c) a nucleus of size 10 fm. Repeat these calculations using alpha particles in place of electrons.
- An electron and a photon each have kinetic energy equal to 50 keV. What are their de Broglie wavelengths?
- Calculate the de Broglie wavelength of a proton that is accelerated through a potential difference of 10 MV.
- Show that the de Broglie wavelength of an electron accelerated from rest through a small potential difference V is given by $\lambda = 1.226/\sqrt{V}$, where λ is in nanometers and V is in volts.
- Find the de Broglie wavelength of a ball of mass 0.20 kg just before it strikes the Earth after being dropped from a building 50 m tall.
- An electron has a de Broglie wavelength equal to the diameter of the hydrogen atom. What is the kinetic energy of the electron? How does this energy compare with the ground-state energy of the hydrogen atom?
- For an electron to be confined to a nucleus, its de Broglie wavelength would have to be less than 10^{-14} m . (a) What would be the kinetic energy of an electron confined to this region? (b) On the basis of this result, would you expect to find an electron in a nucleus? Explain.
- Through what potential difference would an electron have to be accelerated to give it a de Broglie wavelength of $1.00 \times 10^{-10} \text{ m}$?

5.2 The Davisson–Germer Experiment

- Figure P5.13 shows the top three planes of a crystal with planar spacing d . If $2d \sin \theta = 1.01\lambda$ for the two waves shown, and high-energy electrons of wavelength λ penetrate many planes deep into the crystal, which atomic plane produces a wave that cancels the surface reflection? This is an example of how extremely narrow maxima in high-energy electron diffraction are formed—that is, there are no diffracted beams unless $2d \sin \theta$ is equal to an integral number of wavelengths.

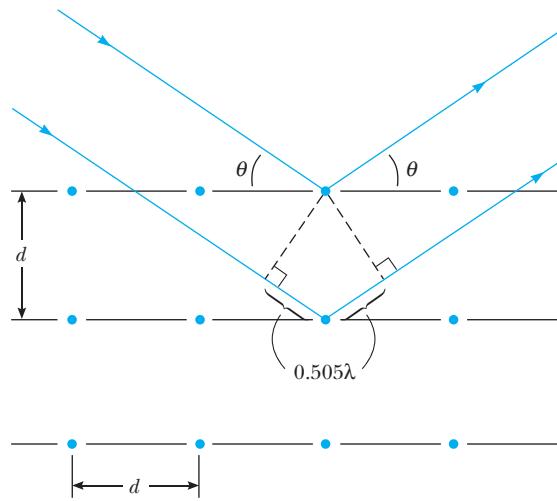


Figure P5.13

- (a) Show that the formula for low-energy electron diffraction (LEED), when electrons are incident perpendicular to a crystal surface, may be written as

$$\sin \phi = \frac{nhc}{d(2m_e c^2 K)^{1/2}}$$

where n is the order of the maximum, d is the atomic spacing, m_e is the electron mass, K is the electron’s kinetic energy, and ϕ is the angle between the incident and diffracted beams. (b) Calculate the atomic spacing

in a crystal that has consecutive diffraction maxima at $\phi = 24.1^\circ$ and $\phi = 54.9^\circ$ for 100-eV electrons.

5.3 Wave Groups and Dispersion

15. Show that the group velocity for a nonrelativistic free electron is also given by $v_g = p/m_e = v_0$, where v_0 is the electron's velocity.
16. When a pebble is tossed into a pond, a circular wave pulse propagates outward from the disturbance. If you are alert (and it's not a sleepy afternoon in late August), you will see a fine structure in the pulse consisting of surface ripples moving inward through the circular disturbance. Explain this effect in terms of group and phase velocity if the phase velocity of ripples is given by $v_p = \sqrt{2\pi S/\lambda\rho}$, where S is the surface tension and ρ is the density of the liquid.
17. The dispersion relation for free relativistic electron waves is

$$\omega(k) = \sqrt{c^2 k^2 + (m_e c^2 / \hbar)^2}$$

Obtain expressions for the phase velocity v_p and group velocity v_g of these waves and show that their product is a constant, independent of k . From your result, what can you conclude about v_g if $v_p > c$?

5.5 The Heisenberg Uncertainty Principle

18. A ball of mass 50 g moves with a speed of 30 m/s. If its speed is measured to an accuracy of 0.1%, what is the minimum uncertainty in its position?
19. A proton has a kinetic energy of 1.0 MeV. If its momentum is measured with an uncertainty of 5.0%, what is the minimum uncertainty in its position?
20. We wish to measure simultaneously the wavelength and position of a photon. Assume that the wavelength measurement gives $\lambda = 6000 \text{ \AA}$ with an accuracy of one part in a million, that is, $\Delta\lambda/\lambda = 10^{-6}$. What is the minimum uncertainty in the position of the photon?
21. A woman on a ladder drops small pellets toward a spot on the floor. (a) Show that, according to the uncertainty principle, the miss distance must be at least

$$\Delta x = \left(\frac{\hbar}{2m} \right)^{1/2} \left(\frac{H}{2g} \right)^{1/4}$$

where H is the initial height of each pellet above the floor and m is the mass of each pellet. (b) If $H = 2.0 \text{ m}$ and $m = 0.50 \text{ g}$, what is Δx ?

22. A beam of electrons is incident on a slit of variable width. If it is possible to resolve a 1% difference in momentum, what slit width would be necessary to resolve the interference pattern of the electrons if their kinetic energy is (a) 0.010 MeV, (b) 1.0 MeV, and (c) 100 MeV?
23. Suppose Fuzzy, a quantum-mechanical duck, lives in a world in which $\hbar = 2\pi \text{ J}\cdot\text{s}$. Fuzzy has a mass of 2.0 kg and is initially known to be within a region 1.0 m wide.

- (a) What is the minimum uncertainty in his speed?
- (b) Assuming this uncertainty in speed to prevail for 5.0 s, determine the uncertainty in position after this time.

24. An electron of momentum p is at a distance r from a stationary proton. The system has a kinetic energy $K = p^2/2m_e$ and potential energy $U = -ke^2/r$. Its total energy is $E = K + U$. If the electron is bound to the proton to form a hydrogen atom, its average position is at the proton but the uncertainty in its position is approximately equal to the radius, r , of its orbit. The electron's average momentum will be zero, but the uncertainty in its momentum will be given by the uncertainty principle. Treat the atom as a one-dimensional system in the following: (a) Estimate the uncertainty in the electron's momentum in terms of r . (b) Estimate the electron's kinetic, potential, and total energies in terms of r . (c) The actual value of r is the one that minimizes the total energy, resulting in a stable atom. Find that value of r and the resulting total energy. Compare your answer with the predictions of the Bohr theory.
25. An excited nucleus with a lifetime of 0.100 ns emits a γ ray of energy 2.00 MeV. Can the energy width (uncertainty in energy, ΔE) of this 2.00-MeV γ emission line be directly measured if the best gamma detectors can measure energies to $\pm 5 \text{ eV}$?
26. Typical measurements of the mass of a subatomic delta particle ($m \approx 1230 \text{ MeV}/c^2$) are shown in Figure P5.26. Although the lifetime of the delta is much too short to measure directly, it can be calculated from the energy-time uncertainty principle. Estimate the lifetime from the full width at half-maximum of the mass measurement distribution shown.

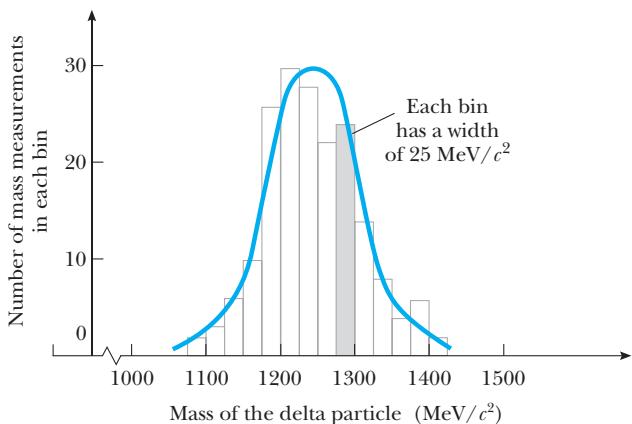


Figure P5.26 Histogram of mass measurements of the delta particle.

5.7 The Wave-Particle Duality

27. A monoenergetic beam of electrons is incident on a single slit of width 0.50 nm. A diffraction pattern is

- formed on a screen 20 cm from the slit. If the distance between successive minima of the diffraction pattern is 2.1 cm, what is the energy of the incident electrons?
28. A neutron beam with a selected speed of 0.40 m/s is directed through a double slit with a 1.0-mm separation. An array of detectors is placed 10 m from the slit. (a) What is the de Broglie wavelength of the neutrons? (b) How far off axis is the first zero-intensity point on the detector array? (c) Can we say which slit any particular neutron passed through? Explain.
29. A two-slit electron diffraction experiment is done with slits of *unequal* widths. When only slit 1 is open, the number of electrons reaching the screen per second is 25 times the number of electrons reaching the screen per second when only slit 2 is open. When both slits are open, an interference pattern results in which the destructive interference is not complete. Find the ratio of the probability of an electron arriving at an interference maximum to the probability of an electron arriving at an adjacent interference minimum. (*Hint:* Use the superposition principle).
- Additional Problems**
30. Robert Hofstadter won the 1961 Nobel prize in physics for his pioneering work in scattering 20-GeV electrons from nuclei. (a) What is the γ factor for a 20-GeV electron, where $\gamma = (1 - v^2/c^2)^{-1/2}$? What is the momentum of the electron in kg·m/s? (b) What is the wavelength of a 20-GeV electron and how does it compare with the size of a nucleus?
31. An air rifle is used to shoot 1.0-g particles at 100 m/s through a hole of diameter 2.0 mm. How far from the rifle must an observer be to see the beam spread by 1.0 cm because of the uncertainty principle? Compare this answer with the diameter of the Universe (2×10^{26} m).
32. An atom in an excited state 1.8 eV above the ground state remains in that excited state 2.0 μs before moving to the ground state. Find (a) the frequency of the emitted photon, (b) its wavelength, and (c) its approximate uncertainty in energy.
33. A π^0 meson is an unstable particle produced in high-energy particle collisions. It has a mass–energy equivalent of about 135 MeV, and it exists for an average lifetime of only 8.7×10^{-17} s before decaying into two γ rays. Using the uncertainty principle, estimate the fractional uncertainty $\Delta m/m$ in its mass determination.
34. (a) Find and sketch the spectral content of the rectangular pulse of width 2τ shown in Figure P5.34. (b) Show that a reciprocity relation $\Delta\omega \Delta t \approx \pi$ holds

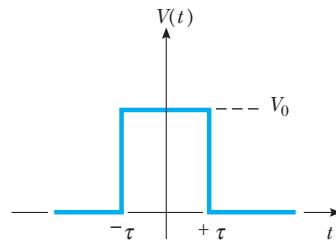


Figure P5.34

in this case. Take $\Delta t = \tau$ and define $\Delta\omega$ similarly.

(c) What range of frequencies is required to compose a pulse of width $2\tau = 1 \mu\text{s}$? A pulse of width $2\tau = 1 \text{ ns}$?

35. A *matter wave packet*. (a) Find and sketch the real part of the matter wave pulse shape $f(x)$ for a Gaussian amplitude distribution $a(k)$, where

$$a(k) = Ae^{-\alpha^2(k-k_0)^2}$$

Note that $a(k)$ is peaked at k_0 and has a width that decreases with increasing α . (*Hint:* In order to put $f(x) = (2\pi)^{-1/2} \int_{-\infty}^{+\infty} a(k)e^{ikx} dk$ into the standard form $\int_{-\infty}^{+\infty} e^{-az^2} dz$, complete the square in k .) (b) By comparing the result for the real part of $f(x)$ to the standard form of a Gaussian function with width Δx , $f(x) \propto Ae^{-(x/2\Delta x)^2}$, show that the width of the matter wave pulse is $\Delta x = \alpha$. (c) Find the width Δk of $a(k)$ by writing $a(k)$ in standard Gaussian form and show that $\Delta x \Delta k = \frac{1}{2}$, independent of α .

36. Consider a freely moving quantum particle with mass m and speed v . Its energy is $E = K + U = \frac{1}{2}mv^2 + 0$. Determine the phase speed of the quantum wave representing the particle and show that it is different from the speed at which the particle transports mass and energy.
37. In a vacuum tube, electrons are boiled out of a hot cathode at a slow, steady rate and accelerated from rest through a potential difference of 45.0 V. Then they travel altogether 28.0 cm as they go through an array of slits and fall on a screen to produce an interference pattern. Only one electron at a time will be in flight in the tube, provided the beam current is below what value? In this situation the interference pattern still appears, showing that each individual electron can interfere with itself.

DeBroglie Hypothesis

$$\lambda_{\text{DeBroglie}} = h/mv = h/p$$

In this case, we are considering the

electron to be a WAVE, and

the **electron wave will “fit” around the orbit if the momentum (and energy) is just right**

(as in the above relation). But this will happen only for specific cases - and those are the specific allowed orbits (r_n) and energies (E_n) that are allowed in the Bohr Theory!

What we now have is a **wave/particle duality for light** (E&M vs photon), **AND** a **wave/particle duality for electrons!**

DeBroglie Hypothesis

If the electron behaves as a wave, with $\lambda = h/mv$, then we should be able to test this wave behavior via interference and diffraction.

In fact, experiments show that electrons **DO EXHIBIT INTERFERENCE** when they go through multiple slits, just as the DeBroglie Hypothesis indicates.

Properties of matter waves

- The lighter the particle, greater the wavelength associated with it
- Smaller the velocity of the particle greater the wavelength associated with it
- When $v=0$, $\lambda=\infty$, this shows that matter waves are generated by the motion of particles. These waves are produced whether the particles are charged or uncharged. Wavelength is independent of charge. This fact reveals that these waves are not electromagnetic and they are new kind of waves
- The velocity of matter waves is not constant while velocity of electromagnetic wave is constant
- Velocity of matter wave is greater than velocity of light

Proof:

A particle in motion with an associated wave has two different velocities

- Due to the mechanical motion of the particle v and
the other related to the propagation of the wave denoted by ω

We know that $E=hc$

$$E=mc^2, \quad v=\frac{mc^2}{h}$$

The velocity of the wave is given by

$$\omega=v\lambda, \quad \lambda=[\frac{h}{mv}]$$

$$\omega = \frac{c^2}{v}$$

since the velocity v cannot exceed velocity of light c , Hence $\omega>c$
(velocity of light) This is an unexpected result and this can be
understood by assuming by the wave velocity equal to phase velocity or
group velocity.

- The wave and particle aspects of moving bodies can never appear together in the same experiment
- The wave nature of matter introduces an uncertainty in the location of position of the particle because a wave cannot be exactly taken at a particular point, If the wave is very large, the particle can be identified at a given point whereas if the wave is small, the particle cannot be located easily. Heisenberg's uncertainty principle is based on this concept.

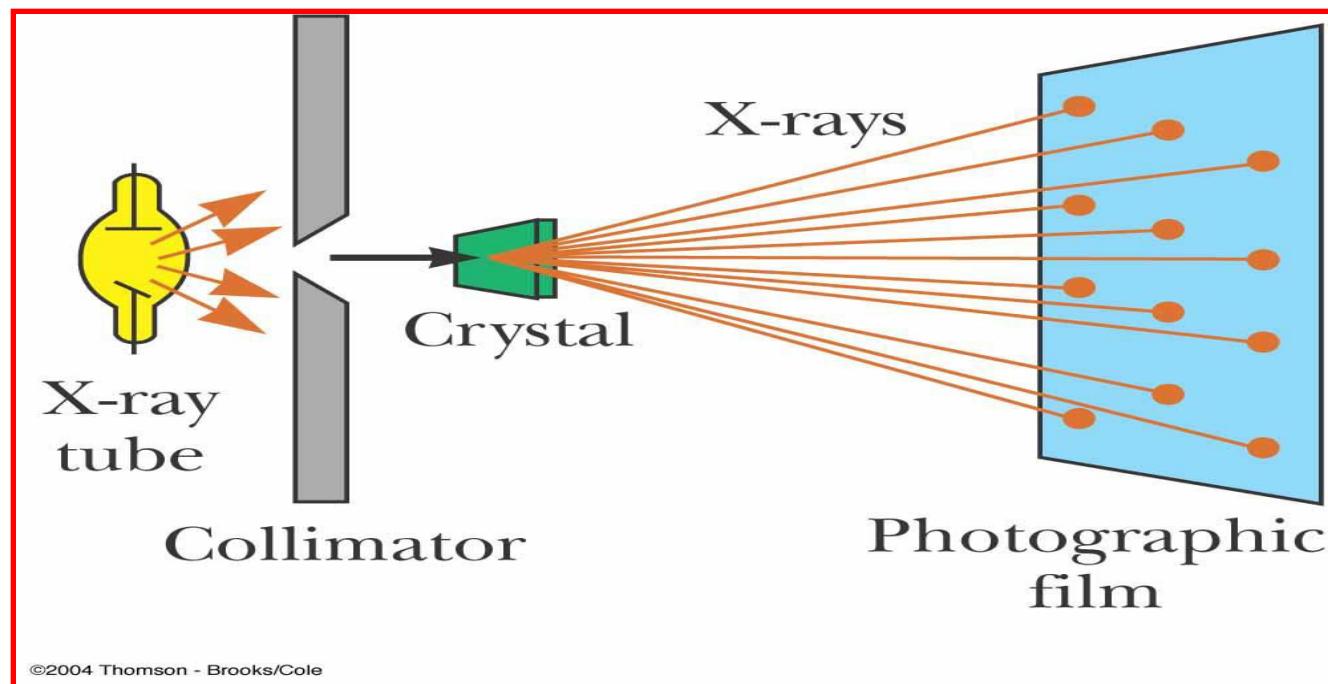
Davission-Germer Experiment



- If particles have a wave nature, then under appropriate conditions, they should exhibit diffraction
- Davisson and Germer measured the wavelength of electrons
- This provided experimental confirmation of the matter waves proposed by de Broglie

Diffraction of X-Rays by Crystals

- X-rays are electromagnetic waves of relatively short wavelength ($\lambda = 10^{-8}$ to 10^{-12} m = 100 – 0.01 Å)
- Max von Laue suggested that the regular array of atoms in a crystal (spacing in order of several Angstroms) could act as a three-dimensional diffraction grating for x-rays

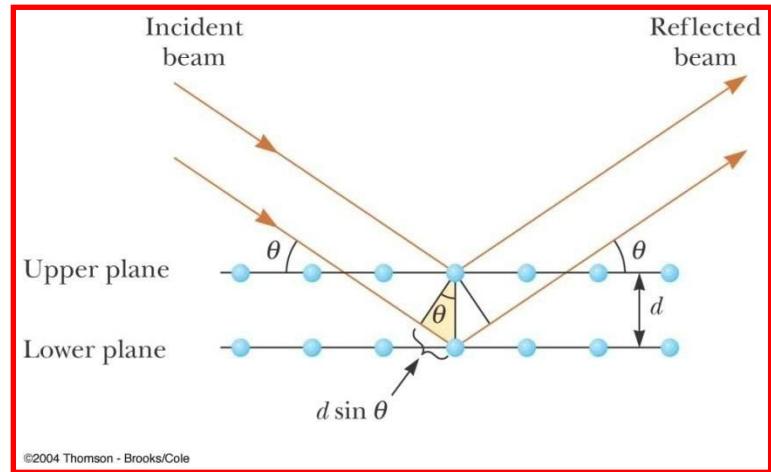


X-Ray Diffraction

- This is a two-dimensional description of the reflection (diffraction) of the x-ray beams
- The condition for *constructive interference* is

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

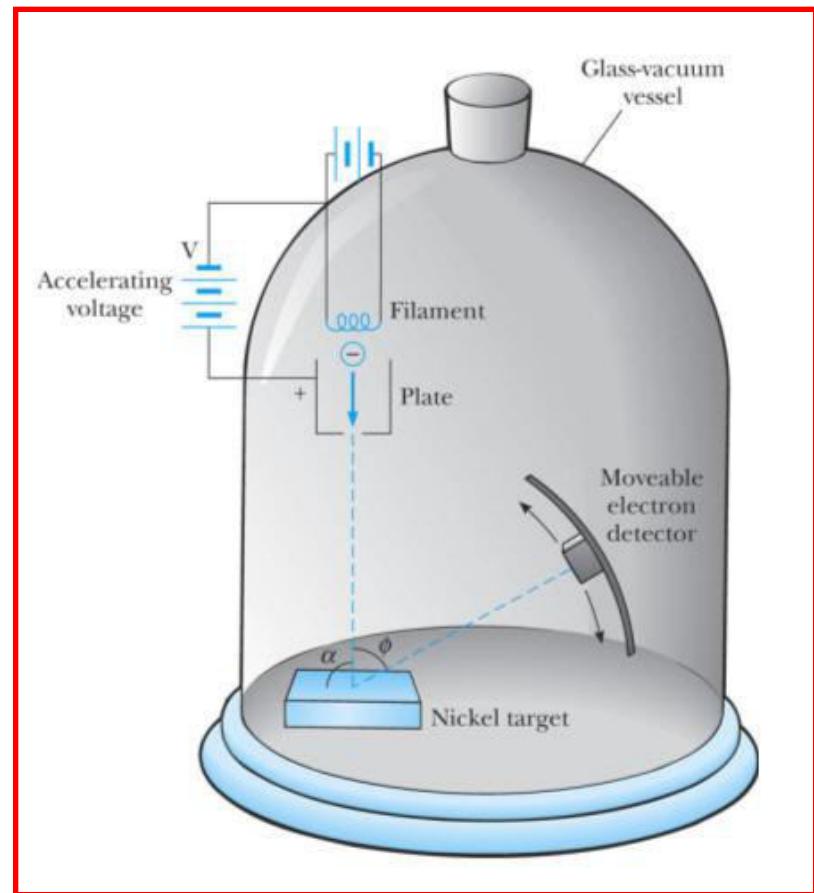
where $n = 1, 2, 3$



- This condition is known as **Bragg's law**
- This can also be used to calculate the spacing between atomic planes

Davisson and Germer Experiment

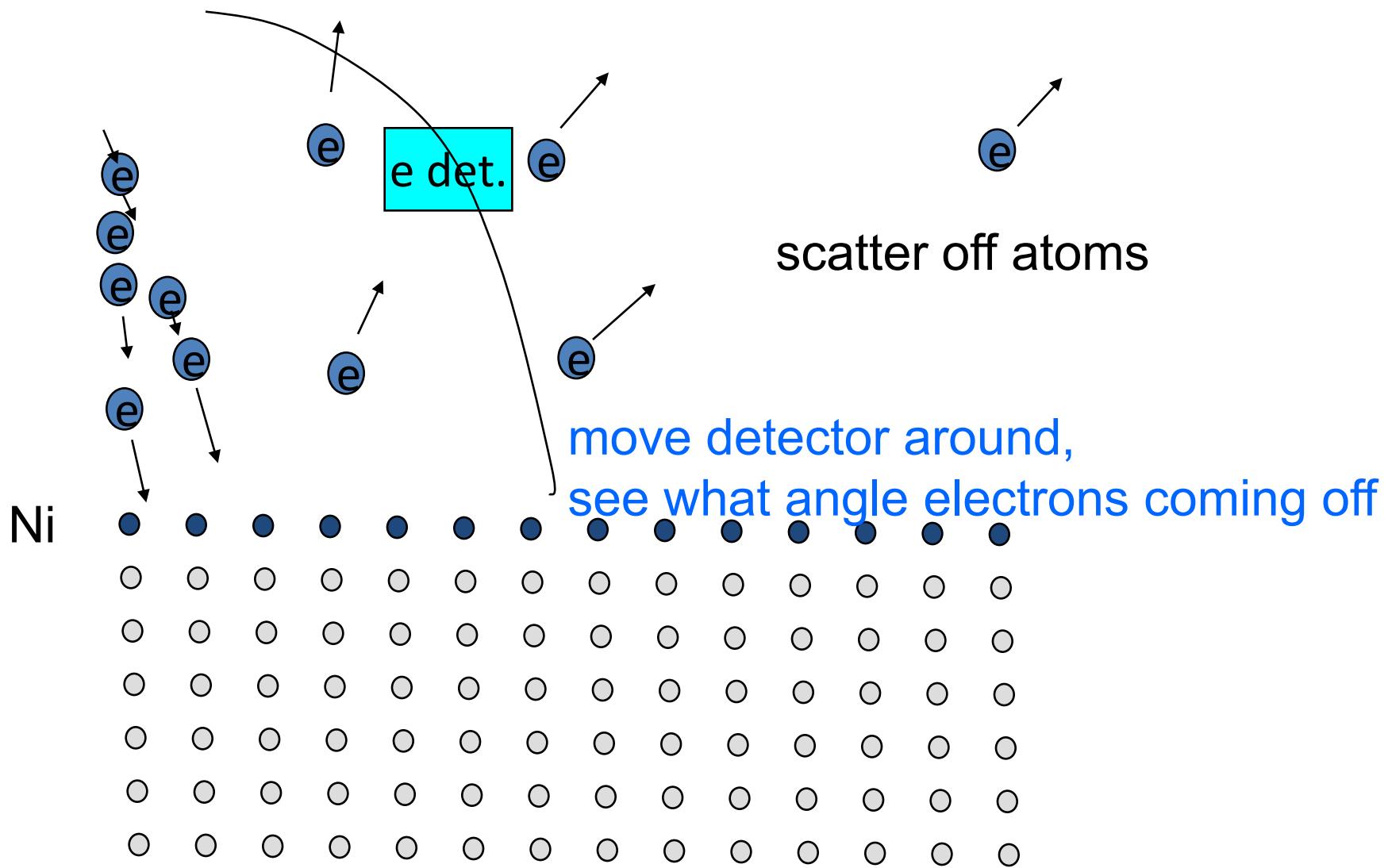
- Electrons were directed onto nickel crystals
- Accelerating voltage is used to control electron energy: $E = |e|V$
- The scattering angle and intensity (electron current) are detected
 - φ is the scattering angle



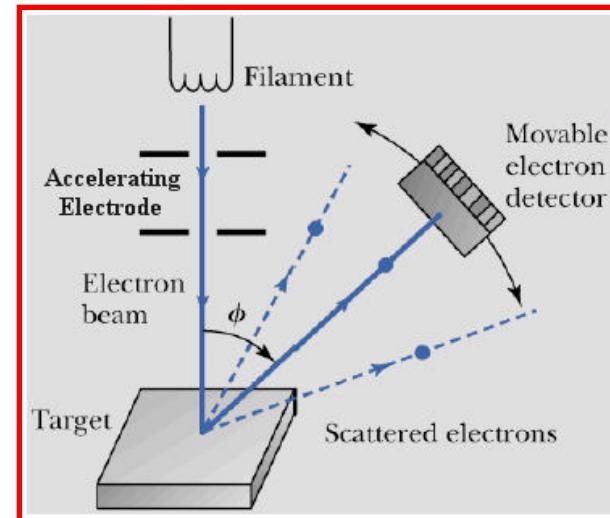
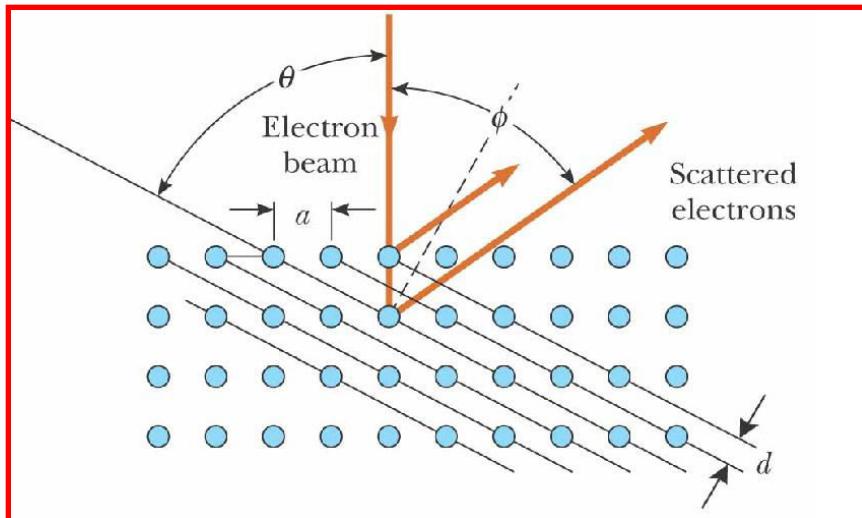
Davisson and Germer -- VERY clean nickel crystal.

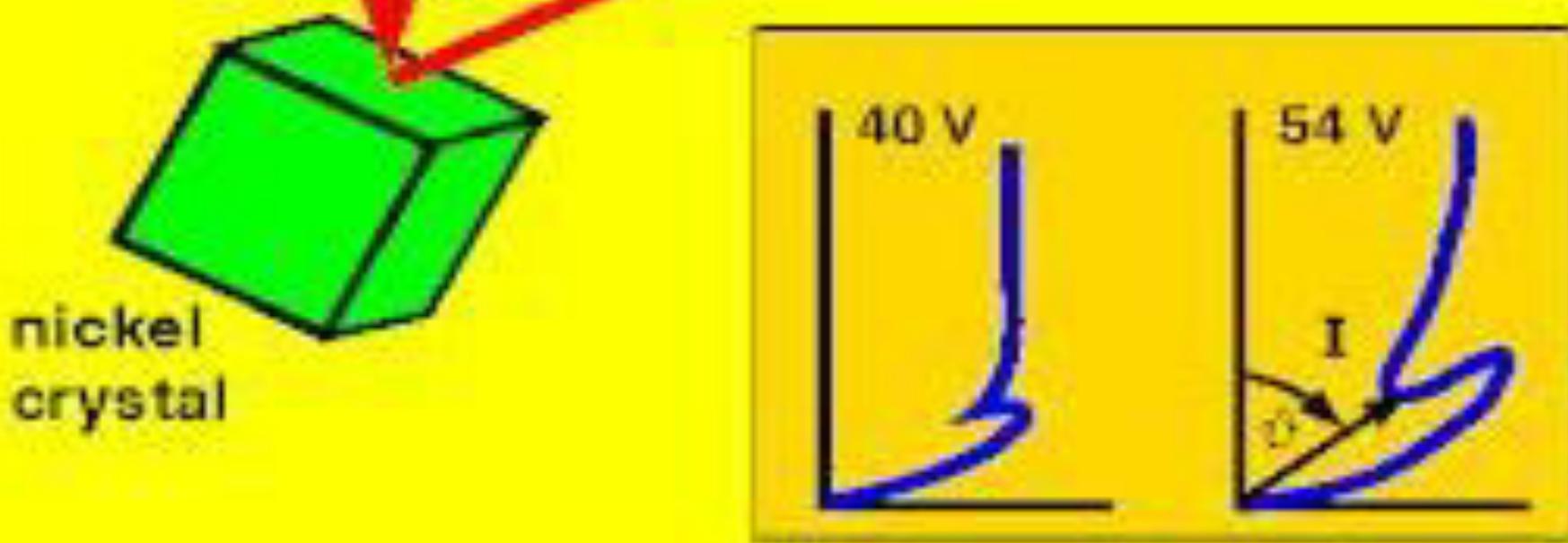
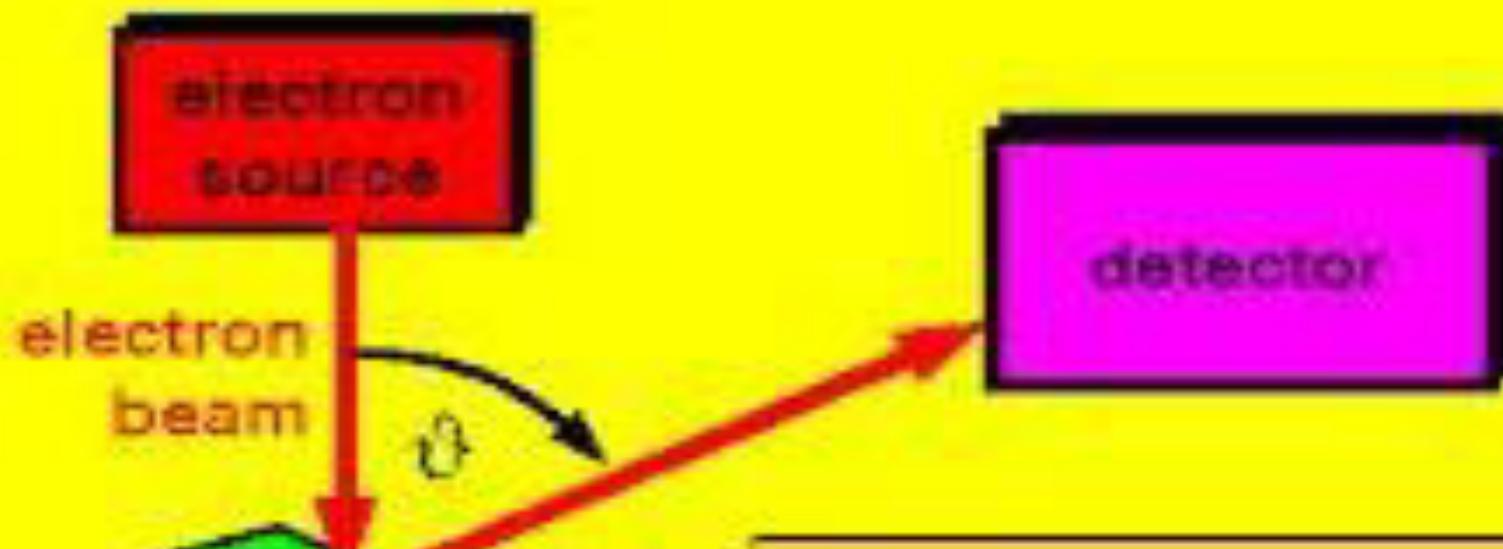
05.de_Broglie_hypothesis_08-Jan-2021_Reference_Material_I

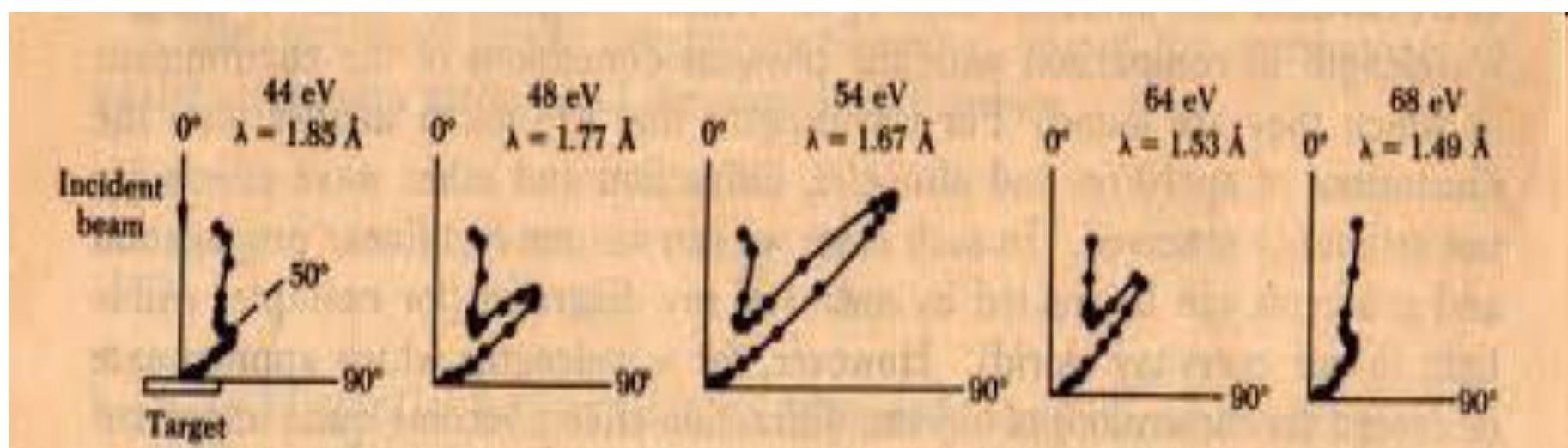
Interference is electron scattering off Ni atoms.



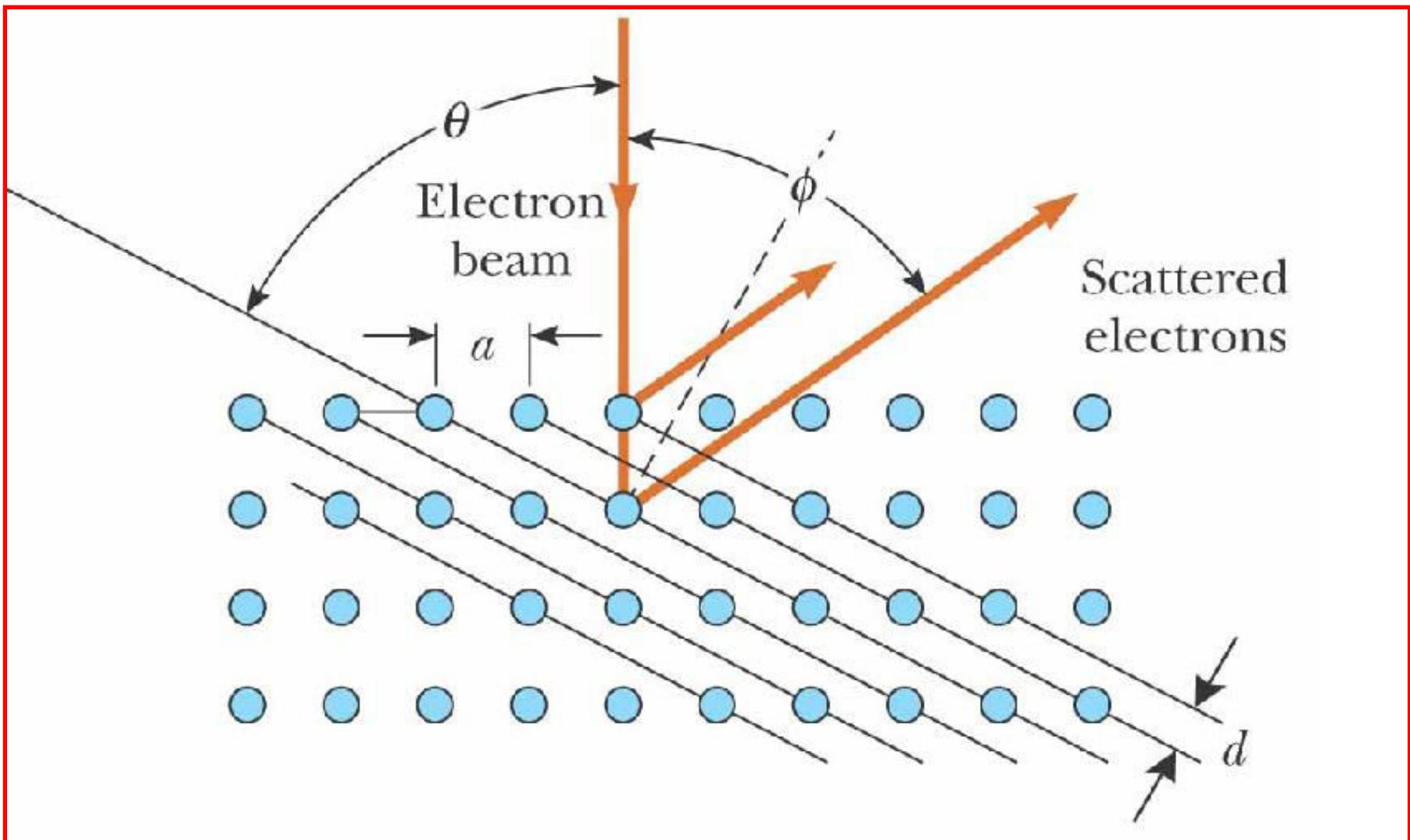
- If electrons are “just” particles, we expect a smooth monotonic dependence of scattered intensity on angle and voltage because only elastic collisions are involved
- Diffraction pattern similar to X-rays would be observed if electrons behave as waves







Davission and Germer Experiment



Davisson and Germer Experiment

- **Observations:**

- Intensity was stronger for certain angles for *specific* accelerating voltages (i.e. for specific electron energies)
- Electrons were reflected in almost the same way that X-rays of comparable wavelength

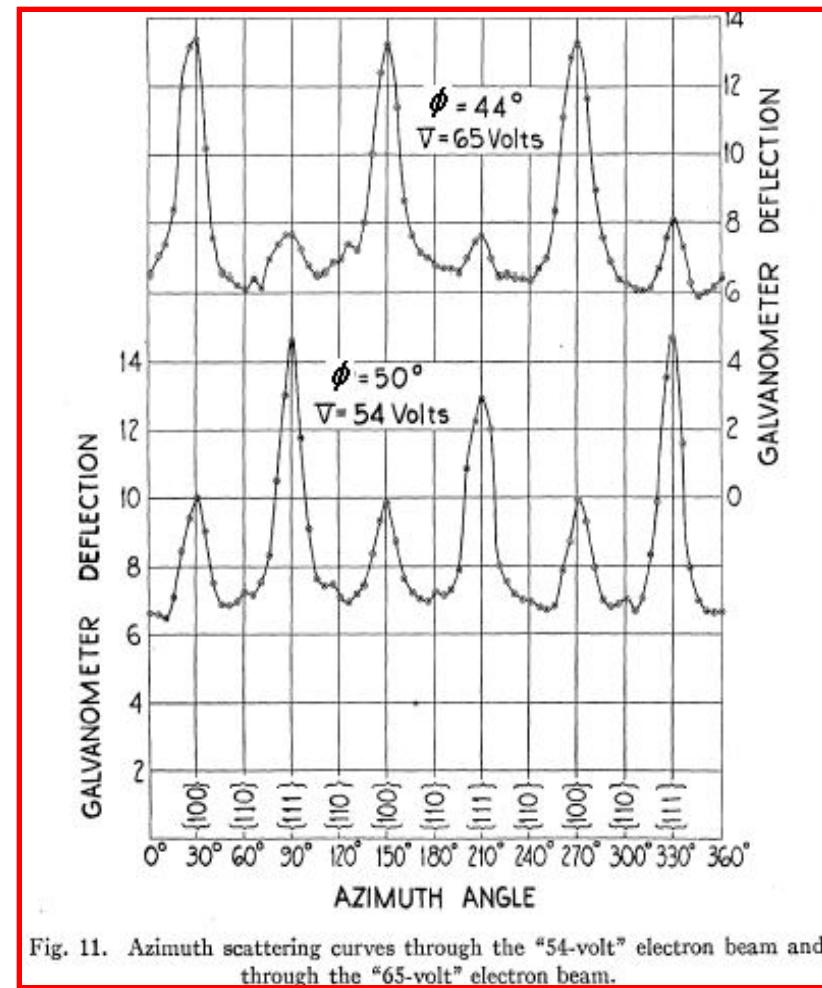
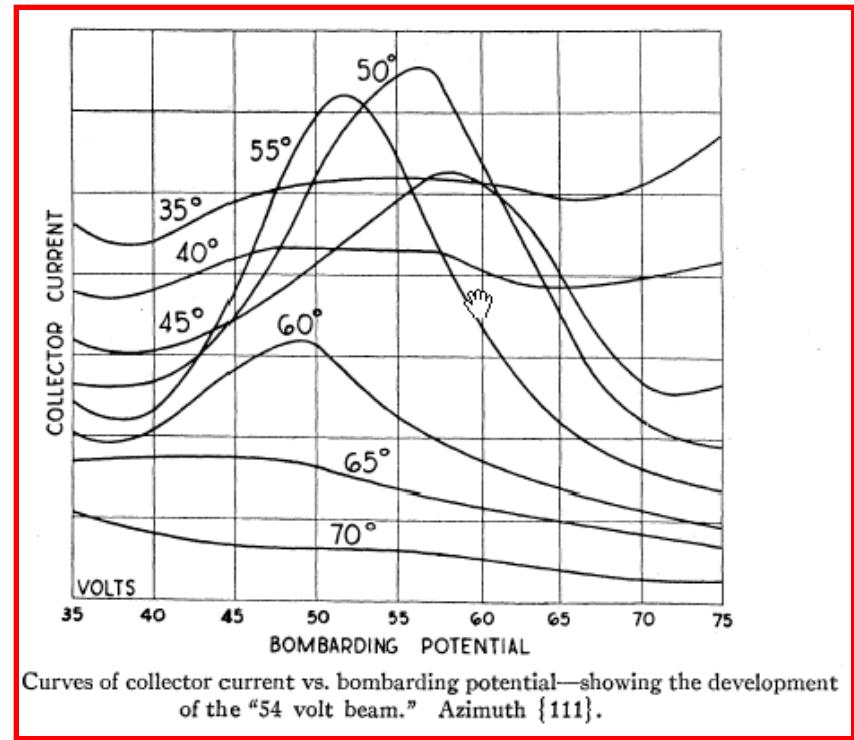


Fig. 11. Azimuth scattering curves through the "54-volt" electron beam and through the "65-volt" electron beam.

Davission and Germer Experiment

- **Observations:**

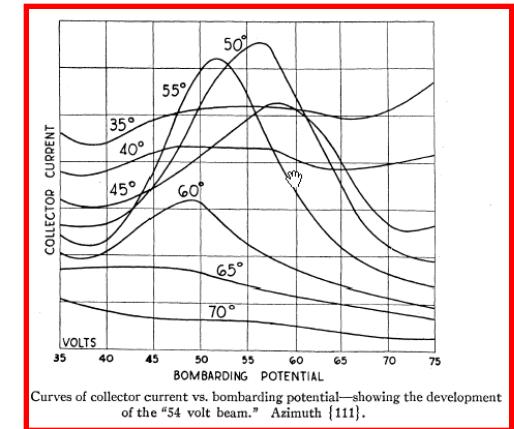
- Current vs accelerating voltage has a maximum, i.e. the highest number of electrons is scattered in a specific direction
- This can't be explained by particle-like nature of electrons \Rightarrow electrons scattered on crystals behave as waves



For $\phi \sim 50^\circ$ the maximum is at $\sim 54V$

- Assuming the wave nature of electrons we can use de Broglie's approach to calculate wavelengths of a matter wave corresponding to electrons in this experiment

- $V = 54 \text{ V} \Rightarrow E = 54 \text{ eV} = 8.64 \times 10^{-18} \text{ J}$



$$E = \frac{p^2}{2m}, \quad p = \sqrt{2mE}, \quad \lambda_B = \frac{h}{\sqrt{2mE}}$$

$$\lambda_B = \frac{6.63 \times 10^{-34} \text{ J} \cdot \text{sec}}{\sqrt{2 \times 9.1 \times 10^{-31} \text{ kg} \times 8.6 \times 10^{-18} \text{ J}}} = 1.67 \text{ \AA}^{\circ}$$

This is in excellent agreement with wavelengths of X-rays diffracted from Nickel!

Davisson and Germer Experiment

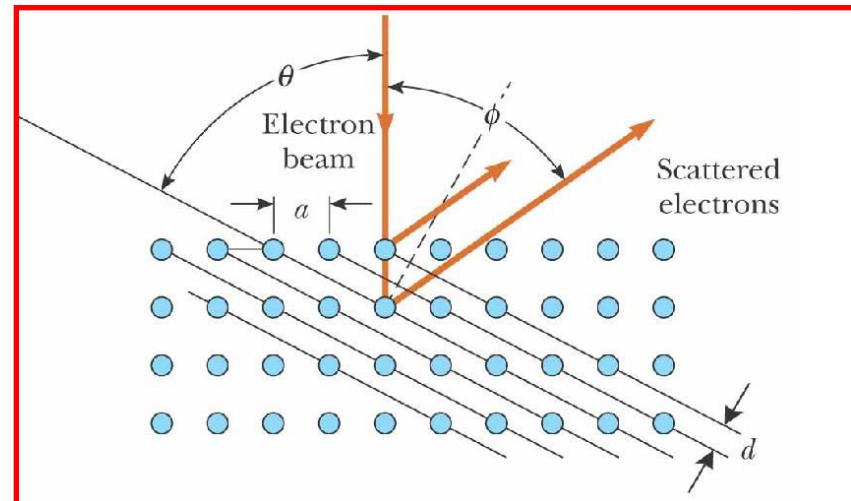
- For X-ray Diffraction on Nickel

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

$$d_{<111>} = 0.91 \text{ \AA} ; \lambda_{\text{X-ray}} = 1.65 \text{ \AA}$$

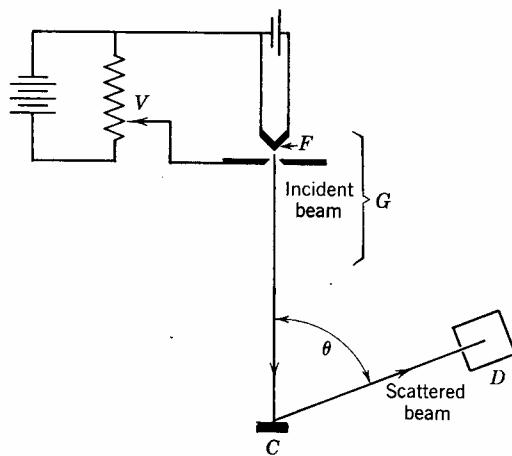
↓

$$\theta = 65^\circ \Rightarrow \phi = 50^\circ$$

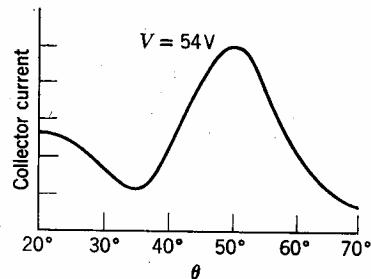
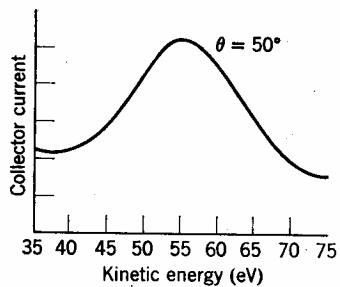


Davisson-Germer Experiment

In order to test de Broglie's hypothesis that matter behaved like waves, Davisson and Germer set up an experiment very similar to what might be used to look at the interference pattern from x-rays scattering from a crystal surface. The basic idea is that the planar nature of crystal structure provides scattering surfaces at regular intervals, thus waves that scatter from one surface can constructively or destructively interfere from waves that scatter from the next crystal plane deeper into the crystal. Their experimental apparatus is shown below.



This simple apparatus send an electron beam with an adjustable energy to a crystal surface, and then measures the current of electrons detected at a particular scattering angle theta. The results of an energy scan at a particular angle and an angle scan at a fixed energy are shown below. Both show a characteristic shape indicative of an interference pattern and consistent with the planar separation in the crystal. This was dramatic proof of the wave nature of matter.



SCHRODINGER WAVE EQUATION

- Time dependent wave equation
- Time independent equation

- (i) de Broglie wavelength can be applied for the matter waves for any field of force. Based on this, the total energy of a particle can be written as,

$$T.E = P.E + K.E \quad \text{or} \quad E = V + \left(\frac{1}{2}\right)mv^2$$

$$E = V + \frac{p^2}{2m}$$

Since $p = mv$ $p = [2m(E - V)]^{1/2}$ (1)

But from de Broglie wavelength,

$$\lambda = \frac{h}{p} \quad (2)$$

$$\lambda = \frac{h}{[2m(E - V)]^{1/2}}$$

(ii) The wave function associated with the material particles, with function of time 't' can be written as,

$$\psi = \psi_0 e^{-i\omega t} \quad (3)$$

Where ψ_0 is the amplitude of the wave at the point (x, y, z) and $\omega = 2\pi\nu$ where ν is the frequency of radiation.

(i) Schrodinger time independent equation

Let us consider a system of stationary wave associated with a moving particle. Let ψ be the wave function of the particle along X, Y and Z coordinate axes at any time t.

The differential wave equation of a progressive wave with wave velocity u can be written as in terms of Cartesian coordinate.

$$\frac{d^2\psi}{dx^2} + \frac{d^2\psi}{dy^2} + \frac{d^2\psi}{dz^2} = \frac{1}{u^2} \times \frac{d^2\psi}{dt^2} \quad (4)$$

The solution for the equation (4) is equation (3)

Differentiating equation (3) w.r.t. t twice, we get,

$$\begin{aligned} \frac{d\psi}{dt} &= -i\omega\psi_0 e^{-i\omega t} = -i\omega\psi \\ \frac{d^2\psi}{dt^2} &= -\omega^2\psi \end{aligned} \quad (5)$$

Substituting $\frac{d^2\psi}{dt^2}$ value in equation (4), we get,

$$\frac{d^2\psi}{dx^2} + \frac{d^2\psi}{dy^2} + \frac{d^2\psi}{dz^2} = -\frac{\omega^2}{u^2}\psi \quad (6)$$

$$= -\frac{(2\pi\nu)^2}{u^2}\psi$$

Where $\omega = 2\pi\nu$

$$\frac{d^2\psi}{dx^2} + \frac{d^2\psi}{dy^2} + \frac{d^2\psi}{dz^2} = -\frac{4\pi^2\nu^2}{u^2}\psi$$

Substituting the wave velocity $u (= \nu\lambda)$, we get,

$$\frac{d^2\psi}{dx^2} + \frac{d^2\psi}{dy^2} + \frac{d^2\psi}{dz^2} = -\frac{4\pi^2\nu^2}{\lambda^2\nu^2}\psi \quad (7)$$

Substituting wavelength value from equation (2) in equation (7)

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} = -\frac{4\pi^2}{h^2} \times 2m(E - V)\psi \quad (8)$$

$$\text{But } \nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

Where ∇^2 is the Laplacian operator.

Equation (8) can be written as,

$$\begin{aligned} \nabla^2 \psi &= -\frac{8\pi^2}{h^2} m(E - V)\psi \\ &= -\frac{2m}{h^2} (E - V)\psi \end{aligned} \quad (9)$$

$$\text{Where } h = \frac{h}{2\pi}$$

$$\text{Therefore, } \nabla^2 \psi + \frac{2m}{h^2} (E - V)\psi = 0 \quad (10)$$

Equation (9) or (10) is known as ***Schrodinger time independent equation.***

(ii) Schrodinger time dependent equation

From the Schrodinger second assumption, differentiating Equation (3) w.r.t. t, we get,

$$\frac{d\psi}{dt} = -i\omega\psi = -i2\pi\nu\psi$$

Substituting $E = \hbar\nu$ we get

$$= \frac{-i2\pi E}{\hbar}\psi = -i\frac{E}{\hbar}\psi$$

$$E\psi = -\frac{\hbar}{i} \frac{d\psi}{dt}$$

$$E\psi = i\hbar \frac{d\psi}{dt} \quad (12)$$

From Schrodinger time independent equation (10)

$$\nabla^2\psi + \frac{2m}{\hbar^2} E\psi - \frac{2m}{\hbar^2} V\psi = 0$$

Substituting $E\psi$ value, we get,

$$\nabla^2\psi + \frac{2m}{\hbar^2} i\hbar \times \frac{d\psi}{dt} - \frac{2m}{\hbar^2} V\psi = 0 \quad (13)$$

Multiplying equation (13) by $\frac{\hbar^2}{2m}$, we get,

$$\frac{\hbar^2}{2m} \nabla^2\psi + i\hbar \frac{d\psi}{dt} - V\psi = 0$$

Or $-\frac{\hbar^2}{2m} \nabla^2\psi + V\psi = i\hbar \frac{d\psi}{dt} \quad (14)$

$$-\frac{\hbar^2}{2m} \nabla^2 \psi + V = H \quad (15)$$

Where H is an operator known as *Hamiltonian operator*.
Therefore,

$$H\psi = E\psi \quad (16)$$

Where $E\psi = i\hbar \times \frac{d\psi}{dt}$ is an energy operator

Equations (14) or (16) is called as *Schrodinger time dependent wave equation*.

Physical significance of ψ

- It relates the particles and wave nature of matter statistically.
- It is a complex quantity and hence we cannot measure it.
- Its square is a measure of the probability of finding the particle at a particular position. It cannot predict the exact location of the particle
- The wave function is a complex quantity, whereas the probability is a real and positive quantity. Therefore, a term called position probability density $P(r, t)$ is introduced. It is defined as the product of the wave function and its complex conjugate as,

$$P(r, t) = |\psi(r, t)|^2$$

➤ The probability of finding the particle within a volume $d\tau$ is

$$P = \int |\psi|^2 d\tau$$

Where $d\tau = dx dy dz$

➤ If a particle is definitely present then its probability value is one.

i.e. $P = \int_{-\infty}^{+\infty} |\psi|^2 d\tau = 1$

➤ In optics, the amount of light is expressed in terms of its intensity rather than its amplitude, since intensity is a measurable physical quantity. Similarly, the wave function ψ has no physical meaning, whereas the probability density has physical meaning.

Chapter 6

The Schrödinger Wave Equation

So far, we have made a lot of progress concerning the properties of, and interpretation of the wave function, but as yet we have had very little to say about how the wave function may be derived in a general situation, that is to say, we do not have on hand a ‘wave equation’ for the wave function. There is no true derivation of this equation, but its form can be motivated by physical and mathematical arguments at a wide variety of levels of sophistication. Here, we will offer a simple derivation based on what we have learned so far about the wave function.

The Schrödinger equation has two ‘forms’, one in which time explicitly appears, and so describes how the wave function of a particle will evolve in time. In general, the wave function behaves like a wave, and so the equation is often referred to as the time dependent Schrödinger wave equation. The other is the equation in which the time dependence has been ‘removed’ and hence is known as the time independent Schrödinger equation and is found to describe, amongst other things, what the allowed energies are of the particle. These are not two separate, independent equations – the time independent equation can be derived readily from the time dependent equation (except if the potential is time dependent, a development we will not be discussing here). In the following we will describe how the first, time dependent equation can be ‘derived’, and in then how the second follows from the first.

6.1 Derivation of the Schrödinger Wave Equation

6.1.1 The Time Dependent Schrödinger Wave Equation

In the discussion of the particle in an infinite potential well, it was observed that the wave function of a particle of fixed energy E could most naturally be written as a linear combination of wave functions of the form

$$\Psi(x, t) = A e^{i(kx - \omega t)} \quad (6.1)$$

representing a wave travelling in the positive x direction, and a corresponding wave travelling in the opposite direction, so giving rise to a standing wave, this being necessary in order to satisfy the boundary conditions. This corresponds intuitively to our classical notion of a particle bouncing back and forth between the walls of the potential well, which suggests that we adopt the wave function above as being the appropriate wave function

for a *free* particle of momentum $p = \hbar k$ and energy $E = \hbar\omega$. With this in mind, we can then note that

$$\frac{\partial^2\Psi}{\partial x^2} = -k^2\Psi \quad (6.2)$$

which can be written, using $E = p^2/2m = \hbar^2 k^2/2m$:

$$-\frac{\hbar^2}{2m}\frac{\partial^2\Psi}{\partial x^2} = \frac{p^2}{2m}\Psi. \quad (6.3)$$

Similarly

$$\frac{\partial\Psi}{\partial t} = -i\omega\Psi \quad (6.4)$$

which can be written, using $E = \hbar\omega$:

$$i\hbar\frac{\partial\Psi}{\partial t} = \hbar\omega\Psi = E\Psi. \quad (6.5)$$

We now generalize this to the situation in which there is both a kinetic energy and a potential energy present, then $E = p^2/2m + V(x)$ so that

$$E\Psi = \frac{p^2}{2m}\Psi + V(x)\Psi \quad (6.6)$$

where Ψ is now the wave function of a particle moving in the presence of a potential $V(x)$. But if we assume that the results Eq. (6.3) and Eq. (6.5) still apply in this case then we have

$$-\frac{\hbar^2}{2m}\frac{\partial^2\Psi}{\partial x^2} + V(x)\Psi = i\hbar\frac{\partial\Psi}{\partial t} \quad (6.7)$$

which is the famous time dependent Schrödinger wave equation. It is setting up and solving this equation, then analyzing the physical contents of its solutions that form the basis of that branch of quantum mechanics known as wave mechanics.

Even though this equation does not look like the familiar wave equation that describes, for instance, waves on a stretched string, it is nevertheless referred to as a ‘wave equation’ as it can have solutions that represent waves propagating through space. We have seen an example of this: the harmonic wave function for a free particle of energy E and momentum p , i.e.

$$\Psi(x, t) = Ae^{-i(px-Et)/\hbar} \quad (6.8)$$

is a solution of this equation with, as appropriate for a free particle, $V(x) = 0$. But this equation can have distinctly non-wave like solutions whose form depends, amongst other things, on the nature of the potential $V(x)$ experienced by the particle.

In general, the solutions to the time dependent Schrödinger equation will describe the *dynamical* behaviour of the particle, in some sense similar to the way that Newton’s equation $F = ma$ describes the dynamics of a particle in classical physics. However, there is an important difference. By solving Newton’s equation we can determine the position of a particle as a function of time, whereas by solving Schrödinger’s equation, what we get is a wave function $\Psi(x, t)$ which tells us (after we square the wave function) how the *probability* of finding the particle in some region in space varies as a function of time.

It is possible to proceed from here look at ways and means of solving the full, time dependent Schrödinger equation in all its glory, and look for the physical meaning of the solutions that are found. However this route, in a sense, bypasses much important physics contained in the Schrödinger equation which we can get at by asking much simpler questions. Perhaps the most important ‘simpler question’ to ask is this: what is the wave

function for a particle of a given energy E ? Curiously enough, to answer this question requires ‘extracting’ the time dependence from the time dependent Schrödinger equation. To see how this is done, and its consequences, we will turn our attention to the closely related time independent version of this equation.

6.1.2 The Time Independent Schrödinger Equation

We have seen what the wave function looks like for a free particle of energy E – one or the other of the harmonic wave functions – and we have seen what it looks like for the particle in an infinitely deep potential well – see Section 5.3 – though we did not obtain that result by solving the Schrödinger equation. But in both cases, the time dependence entered into the wave function via a complex exponential factor $\exp[-iEt/\hbar]$. This suggests that to ‘extract’ this time dependence we guess a solution to the Schrödinger wave equation of the form

$$\Psi(x, t) = \psi(x)e^{-iEt/\hbar} \quad (6.9)$$

i.e. where the space and the time dependence of the complete wave function are contained in separate factors¹. The idea now is to see if this guess enables us to derive an equation for $\psi(x)$, the spatial part of the wave function.

If we substitute this trial solution into the Schrödinger wave equation, and make use of the meaning of partial derivatives, we get:

$$-\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} e^{-iEt/\hbar} + V(x)\psi(x)e^{-iEt/\hbar} = i\hbar \cdot -iE/\hbar e^{-iEt/\hbar} \psi(x) = E\psi(x)e^{-iEt/\hbar}. \quad (6.10)$$

We now see that the factor $\exp[-iEt/\hbar]$ cancels from both sides of the equation, giving us

$$-\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} + V(x)\psi(x) = E\psi(x) \quad (6.11)$$

If we rearrange the terms, we end up with

$$\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} + (E - V(x))\psi(x) = 0 \quad (6.12)$$

which is the time independent Schrödinger equation. We note here that the quantity E , which we have identified as the energy of the particle, is a free parameter in this equation. In other words, at no stage has any restriction been placed on the possible values for E . Thus, if we want to determine the wave function for a particle with some specific value of E that is moving in the presence of a potential $V(x)$, all we have to do is to insert this value of E into the equation with the appropriate $V(x)$, and solve for the corresponding wave function. In doing so, we find, perhaps not surprisingly, that for different choices of E we get different solutions for $\psi(x)$. We can emphasize this fact by writing $\psi_E(x)$ as the solution associated with a particular value of E . But it turns out that it is not all quite as simple as this. To be physically acceptable, the wave function $\psi_E(x)$ must satisfy two conditions, one of which we have seen before namely that the wave function must be normalizable (see Eq. (5.3)), and a second, that the wave function and its derivative must be continuous. Together, these two requirements, the first founded in the probability interpretation of the wave function, the second in more esoteric mathematical necessities which we will not go into here and usually only encountered in somewhat artificial problems, lead to a rather remarkable property of physical systems described by this equation that has enormous physical significance: the quantization of energy.

¹A solution of this form can be shown to arise by the method of ‘the separation of variables’, a well known mathematical technique used to solve equations of the form of the Schrödinger equation.

The Quantization of Energy

At first thought it might seem to be perfectly acceptable to insert any value of E into the time independent Schrödinger equation and solve it for $\psi_E(x)$. But in doing so we must remain aware of one further requirement of a wave function which comes from its probability interpretation: to be physically acceptable a wave function must satisfy the normalization condition, Eq. (5.3)

$$\int_{-\infty}^{+\infty} |\Psi(x, t)|^2 dx = 1$$

for all time t . For the particular trial solution introduced above, Eq. (6.9):

$$\Psi(x, t) = \psi_E(x)e^{-iEt/\hbar} \quad (6.13)$$

the requirement that the normalization condition must hold gives, on substituting for $\Psi(x, t)$, the result²

$$\int_{-\infty}^{+\infty} |\Psi(x, t)|^2 dx = \int_{-\infty}^{+\infty} |\psi_E(x)|^2 dx = 1. \quad (6.14)$$

Since this integral must be finite, (unity in fact), we *must* have $\psi_E(x) \rightarrow 0$ as $x \rightarrow \pm\infty$ in order for the integral to have any hope of converging to a finite value. The importance of this with regard to solving the time dependent Schrödinger equation is that we must check whether or not a solution $\psi_E(x)$ obtained for some chosen value of E satisfies the normalization condition. If it does, then this is a physically acceptable solution, if it does not, then that solution *and the corresponding value of the energy* are not physically acceptable. The particular case of considerable physical significance is if the potential $V(x)$ is attractive, such as would be found with an electron caught by the attractive Coulomb force of an atomic nucleus, or a particle bound by a simple harmonic potential (a mass on a spring), or, as we have seen in Section 5.3, a particle trapped in an infinite potential well. In all such cases, we find that except for certain discrete values of the energy, the wave function $\psi_E(x)$ does not vanish, or even worse, diverges, as $x \rightarrow \pm\infty$. In other words, it is only for these discrete values of the energy E that we get physically acceptable wave functions $\psi_E(x)$, or to put it more bluntly, the particle can never be observed to have any energy other than these particular values, for which reason these energies are often referred to as the ‘allowed’ energies of the particle. This pairing off of allowed energy and normalizable wave function is referred to mathematically as $\psi_E(x)$ being an eigenfunction of the Schrödinger equation, and E the associated energy eigenvalue, a terminology that acquires more meaning when quantum mechanics is looked at from a more advanced standpoint.

So we have the amazing result that the probability interpretation of the wave function forces us to conclude that the allowed energies of a particle moving in a potential $V(x)$ are restricted to certain discrete values, these values determined by the nature of the potential. This is the phenomenon known as the quantization of energy, a result of quantum mechanics which has enormous significance for determining the structure of atoms, or, to go even further, the properties of matter overall. We have already seen an example of this quantization of energy in our earlier discussion of a particle in an infinitely deep potential

²Note that the time dependence has cancelled out because

$$|\Psi(x, t)|^2 = |\psi_E(x)e^{-iEt/\hbar}|^2 = |\psi_E(x)|^2|e^{-iEt/\hbar}|^2 = |\psi_E(x)|^2$$

since, for any complex number of the form $\exp(i\phi)$, we have $|\exp(i\phi)|^2 = 1$.

well, though we did not derive the results by solving the Schrödinger equation itself. We will consider how this is done shortly.

The requirement that $\psi(x) \rightarrow 0$ as $x \rightarrow \pm\infty$ is an example of a *boundary condition*. Energy quantization is, mathematically speaking, the result of a combined effort: that $\psi(x)$ be a solution to the time independent Schrödinger equation, and that the solution satisfy these boundary conditions. But both the boundary condition and the Schrödinger equation are derived from, and hence rooted in, the nature of the physical world: we have here an example of the unexpected relevance of purely mathematical ideas in formulating a physical theory.

Continuity Conditions There is one additional proviso, which was already mentioned briefly above, that has to be applied in some cases. If the potential should be discontinuous in some way, e.g. becoming infinite, as we have seen in the infinite potential well example, or having a finite discontinuity as we will see later in the case of the finite potential well, it is possible for the Schrödinger equation to have solutions that themselves are discontinuous. But discontinuous potentials do not occur in nature (this would imply an infinite force), and as we know that for continuous potentials we always get continuous wave functions, we then place the extra conditions that the wave function *and* its spatial derivative also must be continuous³. We shall see how this extra condition is implemented when we look at the finite potential well later.

Bound States and Scattering States But what about wave functions such as the harmonic wave function $\Psi(x, t) = A \exp[i(kx - \omega t)]$? These wave functions represent a particle having a definite energy $E = \hbar\omega$ and so would seem to be legitimate and necessary wave functions within the quantum theory. But the problem here, as has been pointed out before in Chapter 5, is that $\Psi(x, t)$ does *not* vanish as $x \rightarrow \pm\infty$, so the normalization condition, Eq. (6.14) cannot be satisfied. So what is going on here? The answer lies in the fact that there are two kinds of wave functions, those that apply for particles trapped by an attractive potential into what is known as a bound state, and those that apply for particles that are free to travel to infinity (and beyond), otherwise known as scattering states. A particle trapped in an infinitely deep potential well is an example of the former: the particle is confined to move within a restricted region of space. An electron trapped by the attractive potential due to a positively charged atomic nucleus is also an example – the electron rarely moves a distance more than ~ 10 nm from the nucleus. A nucleon trapped within a nucleus by attractive nuclear forces is yet another. In all these cases, the probability of finding the particle at infinity is zero. In other words, the wave function for the particle satisfies the boundary condition that it vanish at infinity. So we see that it is when a particle is trapped, or confined to a limited region of space by an attractive potential $V(x)$ (or $V(\mathbf{r})$ in three dimensions), we obtain wave functions that satisfy the above boundary condition, and hand in hand with this, we find that their energies are quantized. But if it should be the case that the particle is free to move as far as it likes in space, in other words, if it is not bound by any attractive potential, (or even repelled by a repulsive potential) then we find that the wave function need not vanish at infinity, and nor is its energy quantized. The problem of how to reconcile this with the normalization condition, and the probability interpretation of the wave function, is a delicate mathematical issue which we cannot hope to address here, but it can be done. Suffice to say that provided the wave function does not diverge at infinity (in

³The one exception is when the discontinuity is infinite, as in the case of the infinite potential well. In that case, only the wave function is required to be continuous.

other words it remains finite, though not zero) we can give a physical meaning of such states as being an idealized mathematical limiting case which, while it does not satisfy the normalization condition, can still be dealt with in, provided some care is taken with the physical interpretation, in much the same way as the bound state wave functions.

In order to illustrate how the time independent Schrödinger equation can be solved in practice, and some of the characteristics of its solutions, we will here briefly reconsider the infinitely deep potential well problem, already solved by making use of general properties of the wave function, in Section 5.3. We will then move on to looking at other simple applications.

6.2 Solving the Time Independent Schrödinger Equation

6.2.1 The Infinite Potential Well Revisited

Suppose we have a single particle of mass m confined to within a region $0 < x < L$ with potential energy $V = 0$ bounded by infinitely high potential barriers, i.e. $V = \infty$ for $x < 0$ and $x > L$. The potential experienced by the particle is then:

$$V(x) = 0 \quad 0 < x < L \quad (6.15)$$

$$= \infty \quad x \geq L; \quad x \leq 0 \quad (6.16)$$

In the regions for which the potential is infinite, the wave function will be zero, for exactly the same reasons that it was set to zero in Section 5.3, that is, there is zero probability of the particle being found in these regions. Thus, we must impose the boundary conditions

$$\psi(0) = \psi(L) = 0. \quad (6.17)$$

Meanwhile, in the region $0 < x < L$, the potential vanishes, so the time independent Schrödinger equation becomes:

$$-\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} = E\psi(x). \quad (6.18)$$

To solve this, we define a quantity k by

$$k = \sqrt{\frac{2mE}{\hbar^2}} \quad (6.19)$$

so that Eq. (6.18) can be written

$$\frac{d^2\psi(x)}{dx^2} + k^2\psi(x) = 0 \quad (6.20)$$

whose general solution is

$$\psi(x) = A \sin(kx) + B \cos(kx). \quad (6.21)$$

It is now that we impose the boundary conditions, Eq. (6.17), to give, first at $x = 0$:

$$\psi(0) = B = 0 \quad (6.22)$$

so that the solution is now

$$\psi(x) = A \sin(kx). \quad (6.23)$$

Next, applying the boundary condition at $x = L$ gives

$$\psi(L) = A \sin(kL) = 0 \quad (6.24)$$

which tells us that either $A = 0$, in which case $\psi(x) = 0$, which is not a useful solution (it says that there is no particle in the well at all!) or else $\sin(kL) = 0$, which gives an equation for k :

$$kL = n\pi, \quad n = 0, \pm 1, \pm 2, \dots \quad (6.25)$$

We exclude the $n = 0$ possibility as that would give us, once again $\psi(x) = 0$, and we exclude the negative values of n as they will merely reproduce the same set of solutions (except with opposite sign⁴) as the positive values. Thus we have

$$k_n = n\pi/L, \quad n = 1, 2, \dots \quad (6.26)$$

where we have introduced a subscript n . This leads to, on using Eq. (6.19),

$$E_n = \frac{\hbar^2 k_n^2}{2m} = \frac{n^2 \pi^2 \hbar^2}{2mL^2}, \quad n = 1, 2, \dots \quad (6.27)$$

as before in Section 5.3. Thus we see that the boundary conditions, Eq. (6.17), have the effect of restricting the values of the energy of the particle to those given by Eq. (6.27). The associated wave functions will be as in Section 5.3, that is we apply the normalization condition to determine A (up to an inessential phase factor) which finally gives

$$\begin{aligned} \psi_n(x) &= \sqrt{\frac{2}{L}} \sin(n\pi x/L) & 0 < x < L \\ &= 0 & x < 0, \quad x > L. \end{aligned} \quad (6.28)$$

6.2.2 The Finite Potential Well

The infinite potential well is a valuable model since, with the minimum amount of fuss, it shows immediately the way that energy quantization as potentials do not occur in nature. However, for electrons trapped in a block of metal, or gas molecules contained in a bottle, this model serves to describe very accurately the quantum character of such systems. In such cases the potential experienced by an electron as it approaches the edges of a block of metal, or as experienced by a gas molecule as it approaches the walls of its container are effectively infinite as far as these particles are concerned, at least if the particles have sufficiently low kinetic energy compared to the height of these potential barriers.

But, of course, any potential well is of finite depth, and if a particle in such a well has an energy comparable to the height of the potential barriers that define the well, there is the prospect of the particle escaping from the well. This is true both classically and quantum mechanically, though, as you might expect, the behaviour in the quantum mechanical case is not necessarily consistent with our classical physics based expectations. Thus we now proceed to look at the quantum properties of a particle in a finite potential well.

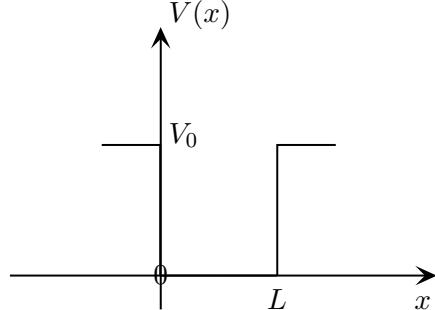


Figure 6.1: Finite potential well.

⁴The sign has no effect on probabilities as we always square the wave function.

In this case, the potential will be of the form

$$V(x) = 0 \quad 0 < x < L \quad (6.29)$$

$$= V \quad x \geq L \quad x \leq 0 \quad (6.30)$$

i.e. we have ‘lowered’ the infinite barriers to a finite value V . We now want to solve the time independent Schrödinger equation for this potential.

To do this, we recognize that the problem can be split up into three parts: $x \leq 0$ where the potential is V , $0 < x < L$ where the potential is zero and $x \geq 0$ where the potential is once again V . Therefore, to find the wave function for a particle of energy E , we have to solve three equations, one for each of the regions:

$$\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} + (E - V)\psi(x) = 0 \quad x \leq 0 \quad (6.31)$$

$$\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} + E\psi(x) = 0 \quad 0 < x < L \quad (6.32)$$

$$\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} + (E - V)\psi(x) = 0 \quad x \geq L. \quad (6.33)$$

The solutions to these equations take different forms depending on whether $E < V$ or $E > V$. We shall consider the two cases separately.

$E < V$

First define

$$k = \sqrt{\frac{2mE}{\hbar^2}} \quad \text{and} \quad \alpha = \sqrt{\frac{2m(V - E)}{\hbar^2}}. \quad (6.34)$$

Note that, as $V > E$, α will be a real number, as it is square root of a positive number. We can now write these equations as

$$\frac{d^2\psi(x)}{dx^2} - \alpha^2\psi(x) = 0 \quad x \leq 0 \quad (6.35)$$

$$\frac{d^2\psi(x)}{dx^2} + k^2\psi(x) = 0 \quad 0 < x < L \quad (6.36)$$

$$\frac{d^2\psi(x)}{dx^2} - \alpha^2\psi(x) = 0 \quad x \geq L. \quad (6.37)$$

Now consider the first of these equations, which will have as its solution

$$\psi(x) = Ae^{-\alpha x} + Be^{+\alpha x} \quad (6.38)$$

where A and B are unknown constants. It is at this point that we can make use of our boundary condition, namely that $\psi(x) \rightarrow 0$ as $x \rightarrow \pm\infty$. In particular, since the solution we are currently looking at applies for $x < 0$, we should look at what this solution does for $x \rightarrow -\infty$. What it does is diverge, because of the term $A \exp(-\alpha x)$. So, in order to guarantee that our solution have the correct boundary condition for $x \rightarrow -\infty$, we must have $A = 0$. Thus, we conclude that

$$\psi(x) = Be^{\alpha x} \quad x \leq 0. \quad (6.39)$$

We can apply the same kind of argument when solving Eq. (6.37) for $x \geq L$. In that case, the solution is

$$\psi(x) = Ce^{-\alpha x} + De^{\alpha x} \quad (6.40)$$

but now we want to make certain that this solution goes to zero as $x \rightarrow \infty$. To guarantee this, we must have $D = 0$, so we conclude that

$$\psi(x) = Ce^{-\alpha x} \quad x \geq L. \quad (6.41)$$

Finally, at least for this part of the argument, we look at the region $0 < x < L$. The solution of Eq. (6.36) for this region will be

$$\psi(x) = P \cos(kx) + Q \sin(kx) \quad 0 < x < L \quad (6.42)$$

but now we have no diverging exponentials, so we have to use other means to determine the unknown coefficients P and Q .

At this point we note that we still have four unknown constants B , P , Q , and C . To determine these we note that the three contributions to $\psi(x)$ do not necessarily join together smoothly at $x = 0$ and $x = L$. This awkward state of affairs has its origins in the fact that the potential is discontinuous at $x = 0$ and $x = L$ which meant that we had to solve three separate equations for the three different regions. But these three separate solutions cannot be independent of one another, i.e. there must be a relationship between the unknown constants, so there must be other conditions that enable us to specify these constants. The extra conditions that we impose, as discussed in Section 6.1.2, are that the wave function has to be a continuous function, i.e. the three solutions:

$$\psi(x) = Be^{\alpha x} \quad x \leq 0 \quad (6.43)$$

$$= P \cos(kx) + Q \sin(kx) \quad 0 < x < L \quad (6.44)$$

$$= Ce^{-\alpha x} \quad x \geq L. \quad (6.45)$$

should all ‘join up’ smoothly at $x = 0$ and $x = L$. This means that the first two solutions and their slopes (i.e. their first derivatives) must be the same at $x = 0$, while the second and third solutions and their derivatives must be the same at $x = L$. Applying this requirement at $x = 0$ gives:

$$B = P \quad (6.46)$$

$$\alpha B = kQ \quad (6.47)$$

and then at $x = L$:

$$P \cos(kL) + Q \sin(kL) = Ce^{-\alpha L} \quad (6.48)$$

$$-kP \sin(kL) + kQ \cos(kL) = -\alpha Ce^{-\alpha L}. \quad (6.49)$$

If we eliminate B and C from these two sets of equations we get, in matrix form:

$$\begin{pmatrix} \alpha & -k \\ \alpha \cos(kL) - k \sin(kL) & \alpha \sin(kL) + k \cos(kL) \end{pmatrix} \begin{pmatrix} P \\ Q \end{pmatrix} = 0 \quad (6.50)$$

and in order that we get a non-trivial solution to this pair of homogeneous equations, the determinant of the coefficients must vanish:

$$\begin{vmatrix} \alpha & -k \\ \alpha \cos(kL) - k \sin(kL) & \alpha \sin(kL) + k \cos(kL) \end{vmatrix} = 0 \quad (6.51)$$

which becomes, after expanding the determinant and rearranging terms:

$$\tan(kL) = \frac{2\alpha k}{k^2 - \alpha^2}. \quad (6.52)$$

Solving this equation for k will give the allowed values of k for the particle in this finite potential well, and hence, using Eq. (6.34) in the form

$$E = \frac{\hbar^2 k^2}{2m} \quad (6.53)$$

we can determine the allowed values of energy for this particle. What we find is that these allowed energies are finite in number, in contrast to the infinite potential well, but to show this we must solve this equation. This is made difficult to do analytically by the fact that this is a transcendental equation – it has no solutions in terms of familiar functions. However, it is possible to get an idea of what its solutions look like either numerically, or graphically. The latter has some advantages as it allows us to see how the mathematics conspires to produce the quantized energy levels. We can first of all simplify the mathematics a little by writing Eq. (6.52) in the form

$$\tan(kL) = \frac{2(\alpha/k)}{1 - (\alpha/k)^2} \quad (6.54)$$

which, by comparison with the two trigonometric formulae

$$\begin{aligned}\tan 2\theta &= \frac{2 \tan \theta}{1 - \tan^2 \theta} \\ \tan 2\theta &= \frac{2 \cot(-\theta)}{1 - \cot^2(-\theta)}\end{aligned}$$

we see that Eq. (6.52) is equivalent to the two conditions

$$\tan\left(\frac{1}{2}kL\right) = \frac{\alpha}{k} \quad (6.55)$$

$$\cot\left(-\frac{1}{2}kL\right) = -\cot\left(\frac{1}{2}kL\right) = \frac{\alpha}{k}. \quad (6.56)$$

The aim here is to plot the left and right hand sides of these two expressions as a function of k (well, actually as a function of $\frac{1}{2}kL$), but before we can do that we need to take account of the fact that the quantity α is given in terms of E by $\sqrt{2m(V - E)/\hbar^2}$, and hence, since $E = \hbar^2 k^2 / 2m$, we have

$$\frac{\alpha}{k} = \sqrt{\frac{V - E}{E}} = \sqrt{\left(\frac{k_0}{k}\right)^2 - 1}$$

where

$$k_0 = \sqrt{\frac{2mV}{\hbar^2}}. \quad (6.57)$$

As we will be plotting as a function of $\frac{1}{2}kL$, it is useful to rewrite the above expression for α/k as

$$\frac{\alpha}{k} = f\left(\frac{1}{2}kL\right) = \sqrt{\left(\frac{1}{2}k_0 L / \frac{1}{2}kL\right)^2 - 1}. \quad (6.58)$$

Thus we have

$$\tan\left(\frac{1}{2}kL\right) = f\left(\frac{1}{2}kL\right) \quad \text{and} \quad -\cot\left(\frac{1}{2}kL\right) = f\left(\frac{1}{2}kL\right). \quad (6.59)$$

We can now plot $\tan(\frac{1}{2}kL)$, $-\cot(\frac{1}{2}kL)$ and $f(\frac{1}{2}kL)$ as functions of $\frac{1}{2}kL$ for various values for k_0 . The points of intersection of the curve $f(\frac{1}{2}kL)$ with the tan and cot curves will then give the kL values for an allowed energy level of the particle in this potential.

This is illustrated in Fig. (6.2) where four such plots are given for different values of V . The important feature of these curves is that the number of points of intersection is finite, i.e. there are only a finite number of values of k that solve Eq. (6.52). Correspondingly, there will only be a finite number of allowed values of E for the particle, and there will always be at least one allowed value.

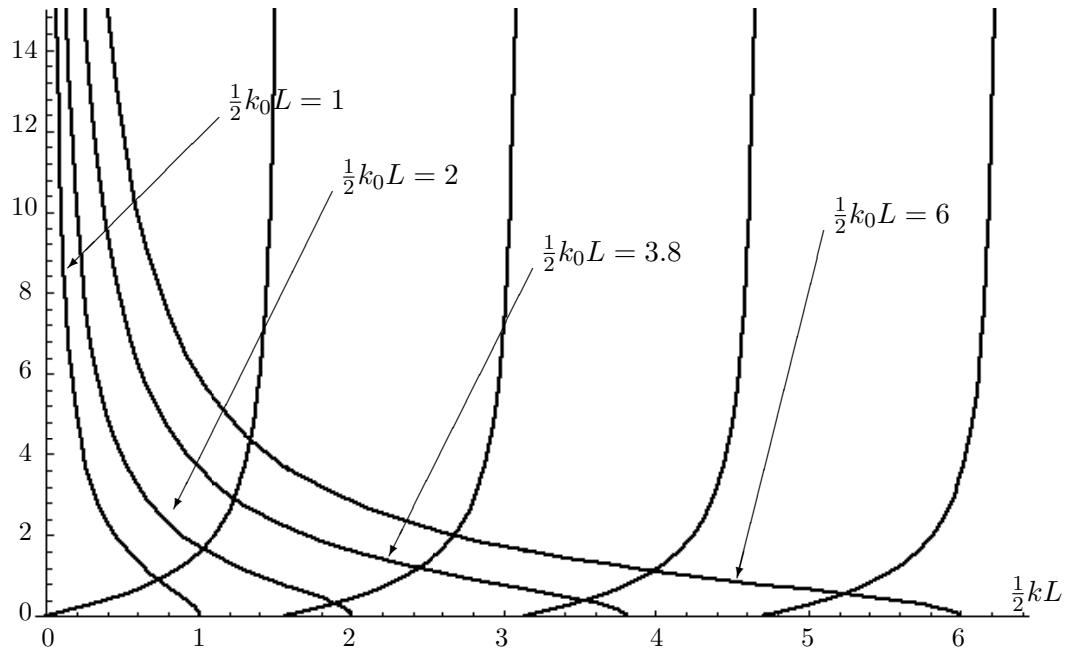


Figure 6.2: Graph to determine bound states of a finite well potential. The points of intersection are solutions to Eq. (6.52). The plots are for increasing values of V , starting with V lowest such that $\frac{1}{2}k_0L = 1$, for which there is only one bound state, slightly higher at $\frac{1}{2}k_0L = 2$, for which there are two bound states, slightly higher again for $\frac{1}{2}k_0L = 3.8$ where there are three bound states, and highest of all, $\frac{1}{2}k_0L = 6$ for which there is four bound states.

To determine the corresponding wave functions is a straightforward but complicated task. The first step is to show, by using Eq. (6.52) and the equations for B , C , P and Q that

$$C = e^{\alpha L} B \quad (6.60)$$

from which readily follows the solution

$$\psi(x) = Be^{\alpha x} \quad x \leq 0 \quad (6.61)$$

$$= B\left(\cos kx + \frac{\alpha}{k} \sin kx\right) \quad 0 < x < L \quad (6.62)$$

$$= Be^{-\alpha(x-L)} \quad x \geq L. \quad (6.63)$$

The constant B is determined by the requirement that $\psi(x)$ be normalized, i.e. that

$$\int_{-\infty}^{+\infty} |\psi(x)|^2 dx = 1. \quad (6.64)$$

which becomes:

$$|B|^2 \left[\int_{-\infty}^0 e^{-2\alpha x} dx + \int_0^L \left(\cos kx + \frac{\alpha}{k} \sin kx \right)^2 dx + \int_L^{+\infty} e^{-2\alpha(x-L)} dx \right] = 1. \quad (6.65)$$

After a somewhat tedious calculation that makes liberal use of Eq. (6.52), the result found is that

$$B = \frac{k}{k_0} \sqrt{\frac{\alpha}{\frac{1}{2}\alpha L + 1}}. \quad (6.66)$$

The task of determining the wave functions is then that of determining the allowed values of k from the graphical solution, or numerically, and then substituting those values into the above expressions for the wave function. The wave functions found a similar in appearance to the infinite well wave functions, with the big difference that they are non-zero outside the well. This is true even if the particle has the lowest allowed energy, i.e. there is a non-zero probability of finding the particle outside the well. This probability can be readily calculated, being just

$$P_{\text{outside}} = |B|^2 \left[\int_{-\infty}^0 e^{-2\alpha x} dx + \int_L^{+\infty} e^{-2\alpha(x-L)} dx \right] = \alpha^{-1} |B|^2 \quad (6.67)$$

6.2.3 Scattering from a Potential Barrier

The above examples are of *bound states*, i.e. wherein the particles are confined to a limited region of space by some kind of attractive or confining potential. However, not all potentials are attractive (e.g. two like charges repel), and in any case, even when there is an attractive potential acting (two opposite charges attracting), it is possible that the particle can be ‘free’ in the sense that it is not confined to a limited region of space. A simple example of this, classically, is that of a comet orbiting around the sun. It is possible for the comet to follow an orbit in which it initially moves towards the sun, then around the sun, and then heads off into deep space, never to return. This is an example of an unbound orbit, in contrast to the orbits of comets that return repeatedly, though sometimes very infrequently, such as Halley’s comet. Of course, the orbiting planets are also in bound states.

A comet behaving in the way just described – coming in from infinity and then ultimately heading off to infinity after bending around the sun – is an example of what is known as a scattering process. In this case, the potential is attractive, so we have the possibility of both scattering occurring, as well as the comet being confined to a closed orbit – a bound state. If the potential was repulsive, then only scattering would occur.

The same distinction applies in quantum mechanics. It is possible for the particle to be confined to a limited region in space, in which case the wave function must satisfy the boundary condition that

$$\psi(x) \rightarrow 0 \quad \text{as} \quad x \rightarrow \pm\infty.$$

As we have seen, this boundary condition is enough to yield the quantization of energy. However, in the quantum analogue of scattering, it turns out that energy is not quantized. This in part can be linked to the fact that the wave function that describes the scattering of a particle of a given energy does not decrease as $x \rightarrow \pm\infty$, so that the very thing that leads to the quantization of energy for a bound particle does not apply here.

This raises the question of what to do about the quantization condition, i.e. that

$$\int_{-\infty}^{+\infty} |\Psi(x, t)|^2 dx = 1.$$

If the wave function does not go to zero as $x \rightarrow \pm\infty$, then it is not possible for the wave function to satisfy this normalization condition – the integral will always diverge.

So how are we to maintain the claim that the wave function must have a probability interpretation if one of the principal requirements, the normalization condition, does not hold true? Strictly speaking, a wave function that cannot be normalized to unity is not physically permitted (because it is inconsistent with the probability interpretation of the wave function). Nevertheless, it is possible to retain, and work with, such wave functions, provided a little care is taken. The answer lies in interpreting the wave function so that $|\Psi(x, t)|^2 \propto \text{particle flux}^5$, though we will not be developing this idea to any extent here.

To illustrate the sort of behaviour that we find with particle scattering, we will consider a simple, but important case, which is that of a particle scattered by a potential barrier. This is sufficient to show the sort of things that can happen that agree with our classical intuition, but it also enables us to see that there occurs new kinds of phenomena that have no explanation within classical physics.

Thus, we will investigate the scattering problem of a particle of energy E interacting with a potential $V(x)$ given by:

$$\begin{aligned} V(x) &= 0 & x < 0 \\ V(x) &= V_0 & x > 0. \end{aligned} \quad (6.68)$$

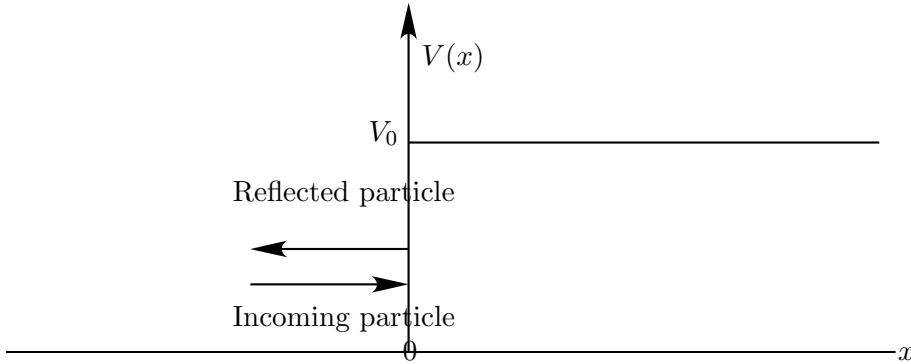


Figure 6.3: Potential barrier with particle of energy $E < V_0$ incident from the left. Classically, the particle will be reflected from the barrier.

In Fig. (6.3) is illustrated what we would expect to happen if a classical particle of energy $E < V_0$ were incident on the barrier: it would simply bounce back as it has insufficient energy to cross over to $x > 0$. Quantum mechanically we find that the situation is not so simple.

Given that the potential is as given above, the Schrödinger equation comes in two parts:

$$\begin{aligned} -\frac{\hbar^2}{2m} \frac{d^2\psi}{dx^2} &= E\psi & x < 0 \\ -\frac{\hbar^2}{2m} \frac{d^2\psi}{dx^2} + V_0\psi &= E\psi & x > 0 \end{aligned} \quad (6.69)$$

where E is, once again, the total energy of the particle.

⁵In more advanced treatments, it is found that the usual probability interpretation does, in fact, continue to apply, though the particle is described not by a wave function corresponding to a definite energy, but rather by a wave packet, though then the particle does not have a definite energy.

We can rewrite these equations in the following way:

$$\begin{aligned}\frac{d^2\psi}{dx^2} + \frac{2mE}{\hbar^2}\psi &= 0 & x < 0 \\ \frac{d^2\psi}{dx^2} - \frac{2m}{\hbar^2}(V_0 - E)\psi &= 0 & x > 0\end{aligned}\quad (6.70)$$

If we put

$$k = \frac{\sqrt{2mE}}{\hbar} \quad (6.71)$$

then the first equation becomes

$$\frac{d^2\psi}{dx^2} + k^2\psi = 0 \quad x < 0$$

which has the general solution

$$\psi = Ae^{ikx} + Be^{-ikx} \quad (6.72)$$

where A and B are unknown constants. We can get an idea of what this solution means if we reintroduce the time dependence (with $\omega = E/\hbar$):

$$\begin{aligned}\Psi(x, t) &= \psi(x)e^{-iEt/\hbar} = \psi(x)e^{-i\omega t} \\ &= Ae^{i(kx-\omega t)} + Be^{-i(kx+\omega t)} \\ &= \text{wave traveling to right} + \text{wave traveling to left}\end{aligned}\quad (6.73)$$

i.e. this solution represents a wave associated with the particle heading towards the barrier and a reflected wave associated with the particle heading away from the barrier. Later we will see that these two waves have the same amplitude, implying that the particle is perfectly reflected at the barrier.

In the region $x > 0$, we write

$$\alpha = \sqrt{2m(V_0 - E)/\hbar} > 0 \quad (6.74)$$

so that the Schrödinger equation becomes

$$\frac{d^2\psi}{dx^2} - \alpha^2\psi = 0 \quad x > 0 \quad (6.75)$$

which has the solution

$$\psi = Ce^{-\alpha x} + De^{\alpha x} \quad (6.76)$$

where C and D are also unknown constants.

The problem here is that the $\exp(\alpha x)$ solution grows exponentially with x , and we do not want wave functions that become infinite: it would essentially mean that the particle is forever to be found at $x = \infty$, which does not make physical sense. So we must put $D = 0$. Thus, if we put together our two solutions for $x < 0$ and $x > 0$, we have

$$\begin{aligned}\psi &= Ae^{ikx} + Be^{-ikx} & x < 0 \\ &= Ce^{-\alpha x} & x > 0.\end{aligned}\quad (6.77)$$

If we reintroduce the time dependent factor, we get

$$\Psi(x, t) = \psi(x)e^{-i\omega t} = Ce^{-\alpha x}e^{-i\omega t} \quad (6.78)$$

which is *not* a travelling wave at all. It is a stationary wave that simply diminishes in amplitude for increasing x .

We still need to determine the constants A, B , and C . To do this we note that for arbitrary choice of these coefficients, the wave function will be discontinuous at $x = 0$. For reasons connected with the requirement that probability interpretation of the wave function continue to make physical sense, we will require that the wave function *and* its first derivative both be continuous⁶ at $x = 0$.

These conditions yield the two equations

$$\begin{aligned} C &= A + B \\ -\alpha C &= ik(A - B) \end{aligned} \quad (6.79)$$

which can be solved to give

$$\begin{aligned} B &= \frac{ik + a}{ik - a} A \\ C &= \frac{2ik}{ik - \alpha} A \end{aligned} \quad (6.80)$$

and hence

$$\begin{aligned} \psi(x) &= Ae^{ikx} + \frac{ik + a}{ik - a} Ae^{-ikx} \quad x < 0 \\ &= \frac{2ik}{ik - \alpha} Ae^{-\alpha x} \quad x < 0. \end{aligned} \quad (6.81)$$

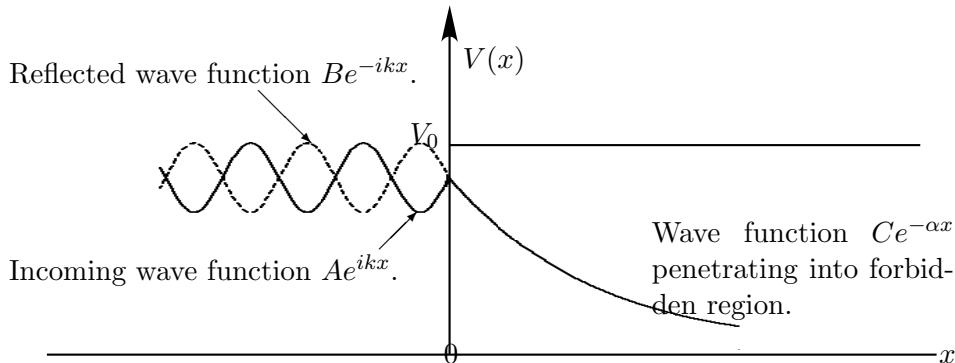


Figure 6.4: Potential barrier with wave function of particle of energy $E < V_0$ incident from the left (solid curve) and reflected wave function (dotted curve) of particle bouncing off barrier. In the classically ‘forbidden’ region of $x > 0$ there is a decaying wave function. Note that the complex wave functions have been represented by their real parts.

Having obtained the mathematical solution, what we need to do is provide a physical interpretation of the result.

⁶To properly justify these conditions requires introducing the notion of ‘probability flux’, that is the rate of flow of probability carried by the wave function. The flux must be such that the point $x = 0$, where the potential is discontinuous, does not act as a ‘source’ or ‘sink’ of probability. What this means, as is shown later, is that we end up with $|A| = |B|$, i.e. the amplitude of the wave arriving at the barrier is the same as the amplitude of the wave reflected at the barrier. If they were different, it would mean that there is more probability flowing into the barrier than is flowing out (or vice versa) which does not make physical sense.

First we note that we cannot impose the normalization condition as the wave function does not decrease to zero as $x \rightarrow -\infty$. But, in keeping with comments made above, we can still learn something from this solution about the behaviour of the particle.

Secondly, we note that the incident and reflected waves have the same ‘intensity’

$$\begin{aligned} \text{Incident intensity} &= |A|^2 \\ \text{Reflected intensity} &= |A|^2 \left| \frac{ik + \alpha}{ik - \alpha} \right|^2 = |A|^2 \end{aligned} \quad (6.82)$$

and hence they have the same amplitude. This suggests that the incident de Broglie wave is totally reflected, i.e. that the particle merely travels towards the barrier where it ‘bounces off’, as would be expected classically. However, if we look at the wave function for $x > 0$ we find that

$$\begin{aligned} |\psi(x)|^2 &\propto \left| \frac{2ik}{ik - \alpha} \right|^2 e^{-2\alpha x} \\ &= \frac{4k^2}{\alpha^2 + k^2} e^{-2\alpha x} \end{aligned} \quad (6.83)$$

which is an exponentially decreasing probability.

This last result tells us that there is a non-zero probability of finding the particle in the region $x > 0$ where, classically, the particle has no chance of ever reaching. The distance that the particle can penetrate into this ‘forbidden’ region is given roughly by $1/2\alpha$ which, for a subatomic particle can be a few nanometers, while for a macroscopic particle, this distance is immeasurably small.

The way to interpret this result is along the following lines. If we imagine that a particle is fired at the barrier, and we are waiting a short distance on the far side of the barrier in the forbidden region with a ‘catcher’s mitt’ poised to grab the particle then we find that *either* the particle hits the barrier and bounces off with the same energy as it arrived with, but with the opposite momentum – it never lands in the mitt, *or* it lands in the mitt and we catch it – it does not bounce off in the opposite direction. The chances of the latter occurring are generally very tiny, but it occurs often enough in microscopic systems that it is a phenomenon that is exploited, particularly in solid state devices. Typically this is done, not with a single barrier, but with a barrier of finite width, in which case the particle can penetrate through the barrier and reappear on the far side, in a process known as quantum tunnelling.

6.3 Expectation Value of Momentum

We can make use of Schrödinger’s equation to obtain an alternative expression for the expectation value of momentum given earlier in Eq. (5.13). This expression is

$$\langle p \rangle = m\langle v(t) \rangle = m \int_{-\infty}^{+\infty} x \left[\frac{\partial \Psi^*(x, t)}{\partial t} \Psi(x, t) + \Psi^*(x, t) \frac{\partial \Psi(x, t)}{\partial t} \right] dx. \quad (6.84)$$

We note that the appearance of time derivatives in this expression. If we multiply both sides by $i\hbar$ and make use of Schrödinger’s equation, we can substitute for these time derivatives to give

$$i\hbar\langle p \rangle = m \int_{-\infty}^{+\infty} x \left[\left\{ \frac{\hbar^2}{2m} \frac{\partial^2 \Psi^*(x, t)}{\partial x^2} - V(x) \Psi^*(x, t) \right\} \Psi(x, t) \right] dx \quad (6.85)$$

$$+ \Psi^*(x, t) \left[- \frac{\hbar^2}{2m} \frac{\partial^2 \Psi(x, t)}{\partial x^2} + V(x) \Psi(x, t) \right] dx. \quad (6.86)$$

The terms involving the potential cancel. The common factor $\hbar^2/2m$ can be moved outside the integral, while both sides of the equation can be divided through by $i\hbar$, yielding a slightly less complicated expression for $\langle p \rangle$:

$$\langle p \rangle = -\frac{1}{2}i\hbar \int_{-\infty}^{+\infty} x \left[\frac{\partial^2 \Psi^*(x, t)}{\partial x^2} \Psi(x, t) - \Psi^*(x, t) \frac{\partial^2 \Psi(x, t)}{\partial x^2} \right] dx. \quad (6.87)$$

Integrating both terms in the integrand by parts then gives

$$\begin{aligned} \langle p \rangle &= \frac{1}{2}i\hbar \int_{-\infty}^{+\infty} \left[\frac{\partial \Psi^*(x, t)}{\partial x} \frac{\partial x \Psi(x, t)}{\partial x} - \frac{\partial x \Psi^*(x, t)}{\partial x} \frac{\partial \Psi(x, t)}{\partial x} \right] dx \\ &\quad + \frac{1}{2}i\hbar \left[\frac{\partial \Psi^*(x, t)}{\partial x} \Psi(x, t) - \Psi^*(x, t) \frac{\partial \Psi(x, t)}{\partial x} \right]_{-\infty}^{+\infty} \end{aligned} \quad (6.88)$$

As the wave function vanishes for $x \rightarrow \pm\infty$, the final term here will vanish. Carrying out the derivatives in the integrand then gives

$$\langle p \rangle = \frac{1}{2}i\hbar \int_{-\infty}^{+\infty} \left[\frac{\partial \Psi^*(x, t)}{\partial x} \Psi(x, t) - \Psi^*(x, t) \frac{\partial \Psi(x, t)}{\partial x} \right] dx \quad (6.89)$$

Integrating the first term only by parts once again then gives

$$\langle p \rangle = -i\hbar \int_{-\infty}^{+\infty} \Psi^*(x, t) \frac{\partial \Psi(x, t)}{\partial x} dx + \frac{1}{2}i\hbar \Psi^*(x, t) \Psi(x, t) \Big|_{-\infty}^{+\infty}. \quad (6.90)$$

Once again, the last term here will vanish as the wave function itself vanishes for $x \rightarrow \pm\infty$ and we are left with

$$\langle p \rangle = -i\hbar \int_{-\infty}^{+\infty} \Psi^*(x, t) \frac{\partial \Psi(x, t)}{\partial x} dx. \quad (6.91)$$

This is a particularly significant result as it shows that the expectation value of momentum can be determined directly from the wave function – i.e. information on the momentum of the particle is contained within the wave function, along with information on the position of the particle. This calculation suggests making the identification

$$p \rightarrow -i\hbar \frac{\partial}{\partial x} \quad (6.92)$$

which further suggests that we can make the replacement

$$p^n \rightarrow \left(-i\hbar \frac{\partial}{\partial x} \right)^n \quad (6.93)$$

so that, for instance

$$\langle p^2 \rangle = -\hbar^2 \int_{-\infty}^{+\infty} \Psi^*(x, t) \frac{\partial^2 \Psi(x, t)}{\partial x^2} dx \quad (6.94)$$

and hence the expectation value of the kinetic energy of the particle is

$$\langle K \rangle = \frac{\langle p^2 \rangle}{2m} = -\frac{\hbar^2}{2m} \int_{-\infty}^{+\infty} \Psi^*(x, t) \frac{\partial^2 \Psi(x, t)}{\partial x^2} dx. \quad (6.95)$$

We can check this idea by turning to the classical formula for the total energy of a particle

$$\frac{p^2}{2m} + V(x) = E. \quad (6.96)$$

If we multiply this equation by $\Psi(x, t) = \psi(x) \exp(-iEt/\hbar)$ and make the replacement given in Eq. (6.94) we end up with

$$-\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} + V(x)\psi(x) = E\psi(x) \quad (6.97)$$

which is just the time independent Schrödinger equation. So there is some truth in the ad hoc procedure outlined above.

This association of the physical quantity p with the derivative i.e. is an example of a physical observable, in this case momentum, being represented by a differential operator. This correspondence between physical observables and operators is to be found throughout quantum mechanics. In the simplest case of position, the operator corresponding to position x is itself just x , so there is no subtleties in this case, but as we have just seen this simple state of affairs changes substantially for other observables. Thus, for instance, the observable quantity K , the kinetic energy, is represented by the differential operator

$$K \rightarrow \hat{K} = -\hbar^2 \frac{\partial^2}{\partial x^2}. \quad (6.98)$$

while the operator associated with the position of the particle is \hat{x} with

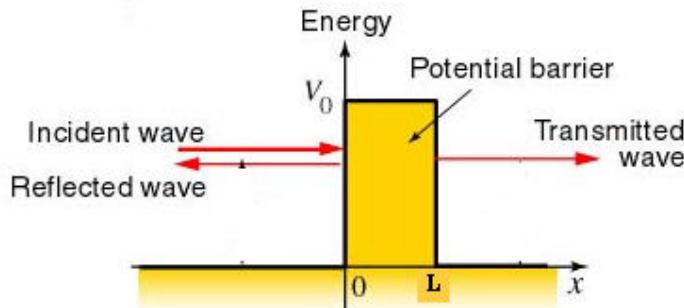
$$x \rightarrow \hat{x} = x. \quad (6.99)$$

In this last case, the identification is trivial.

Updated: 23rd May 2005 at 11:13am.

Tunnel Effect

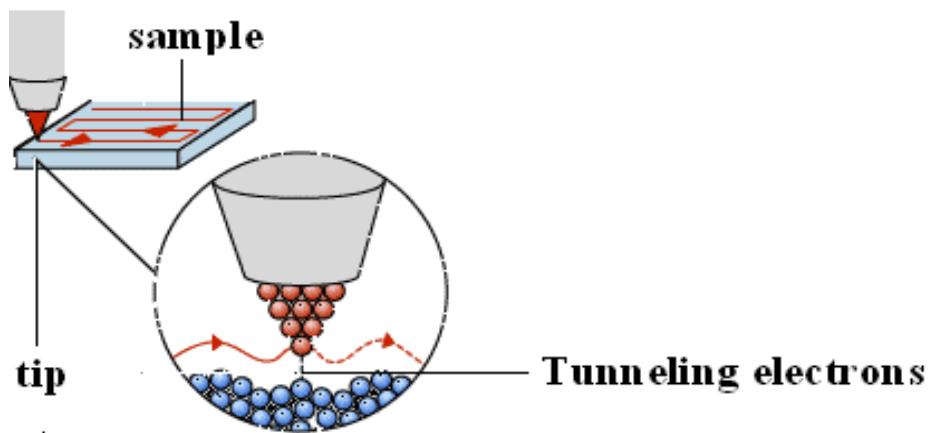
- Particle can penetrate into a potential wall (or barrier)



- Incident particle with energy E coming from left colloides the potential barrier.
- E is assumed to be lower than the top of the barrier
- In classical mechanics the particle is completely reflected by the potential barrier
- In quantum mechanics some probability will be reflected, other penetrates the barrier and pass through-right side region.
- Tunnel effect
- Electronic elements use the tunnel effect
- Principle of STM is based on the tunnel effect

Scanning Tunneling Microscope (STM)

- Type of electron microscope that shows 3D images of a sample
- Structure of the surface is studied using a stylus that scans the surface at a fixed distance from it



- An extremely fine conducting probe is held close to the sample.
- Electrons tunnel between the surface and the stylus, producing an electrical signal.
- Stylus is extremely sharp, the tip being formed by one single atom.
- Slowly scans across the surface at a distance of only an atom's diameter.
- Stylus is raised and lowered in order to keep the signal constant and maintain the distance.
- Smallest details of the surface is scanning.
- Vertical motion of the stylus makes it possible to study the structure of the surface atom by atom.
- A profile of the surface is created, and from that a computer-generated control map of the surface is produced.

Applications

- Used in both industrial and fundamental research to obtain atomic scale images of metal surfaces.
- It provides a 3D profile of the surface which is very useful for characterizing surface roughness and observing surface defects.
- Surface organic molecules can be studied and also their structures.

Scanning Tunneling Microscopy (STM)

Probing the Local Electronic Structure of a Sample's Surface

Scanning Tunneling Microscopy (STM) is one of the application modes for XE series SPM. STM is the ancestor of all scanning probe microscopes. It was invented in 1981 by Gerd Binnig and Heinrich Rohrer at IBM Zurich. Five years later they were awarded the Nobel prize in physics for its invention. The STM was the first instrument to generate real-space images of surface with, so called, atomic resolution (atomic lattice resolution to be precise).

The operation of STM and Conductive AFM is identical except that one uses a sharpened and conducting tip in STM instead of a conductive AFM cantilever as in Conductive AFM. A bias voltage is applied between the tip and the sample. When the tip is brought within about 10 Å of the sample, electrons from the sample begin to "tunnel" through the 10 Å gap into the tip or vice versa, depending upon the sign of the bias voltage as shown in Figure 1. The resulting tunneling current varies with tip-to-sample spacing, and both the sample and the tip must be conductors or semiconductors. Thus, STM cannot image insulating materials.

Figure 1.
Schematic diagram of the XE-series STM system

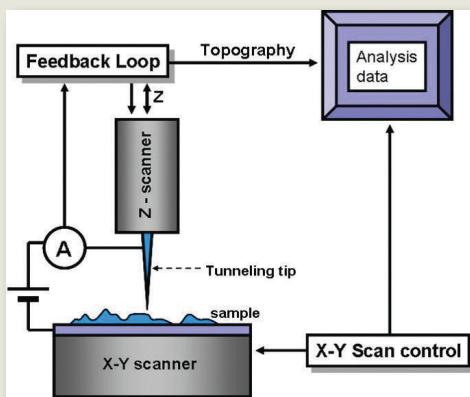
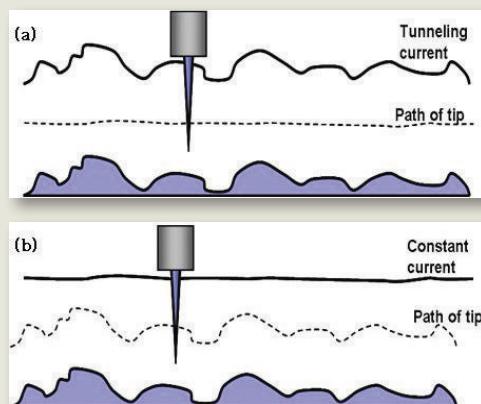


Figure 2.
Comparison of (a) constant-height and (b) constant-current mode for STM .



The tunneling current is an exponential function of distance. Based on quantum mechanics, the tunneling current (I_t) is,

$$I_t = e^{-kd}$$

where d is the distance between tip and sample surface.

If the separation between the tip and the sample changes by 10% (on the order of 1 Å), the tunneling current changes by an order of magnitude. This exponential dependence gives STMs their remarkable sensitivity. STMs can image the surface of the sample with sub-angstrom precision vertically, and atomic resolution laterally.

STM techniques encompass many methods: taking "topographic" (constant-current) images using different bias voltages and comparing them; taking current (constant-height) images at different heights; and ramping the bias voltage with the tip positioned over a feature of interest while recording the tunneling current. The last example results in current vs. voltage (I-V) curves characteristic of the electronic structure at a specific XY location on the sample surface. STMs can be set up to collect I-V curves at every point in a data set, providing a three-dimensional map of electronic structure. With a lock-in amplifier, dl/dV (conductivity) or dl/dz (work function) vs. V curves can be collected directly. All of these are ways of probing the local electronic structure of a surface using an STM.

The schematics of constant-height or constant-current mode are shown in Figure 2. In constant-height mode, the tip travels in a horizontal plane above the sample and the tunneling current varies depending on topography and the local surface electronic properties of the sample. The tunneling current measured at each location on the sample surface constitute the data set, the topography image in Figure 2 (a).

In constant-current mode, STMs use feedback to keep the tunneling current constant by adjusting the height of the scanner at each measurement point in Figure 2 (b). For example, when the system detects an increase in tunneling current, it adjusts the voltage applied to the Z-axis scanner to increase the distance between the tip and the sample. In constant-current mode, the motion of the scanner constitutes the data set. If the system keeps the tunneling current constant to within a few percent, the tip-to-sample distance will be constant to within a few hundredths of an angstrom. Each mode has advantages and disadvantages. Constant-height mode is faster because the system doesn't have to move the scanner up and down, but it provides useful information only for relatively smooth surfaces. Constant current mode can measure irregular surfaces with high precision, but the measurement takes more time.

As a first approximation, an image of the tunneling current maps the topography of the sample. More accurately, the tunneling current corresponds to the electronic density of states at the surface. STMs actually sense the number of filled or unfilled electron states near the Fermi surface, within an energy range determined by the bias voltage. Rather than measuring physical topography, it measures a surface of constant tunneling probability.

From a pessimist's viewpoint, the sensitivity of STM to local electronic structure can cause trouble if you are interested in mapping topography. For example, if an area of the sample has oxidized, the tunneling current will drop precipitously when the tip digs a hole in the surface.

From an optimist's viewpoint, however, the sensitivity of STM to electronic structure can be a tremendous advantage. Other techniques for obtaining information about the electronic properties of a sample detect and average the data originating from a relatively large area, a few microns to a few millimeters across. STM can be used as surface analysis tools that probe the electronic properties of the sample surface with atomic resolution. Figure 3 shows the (a) topography and (b) STM image of highly ordered (HOPG).

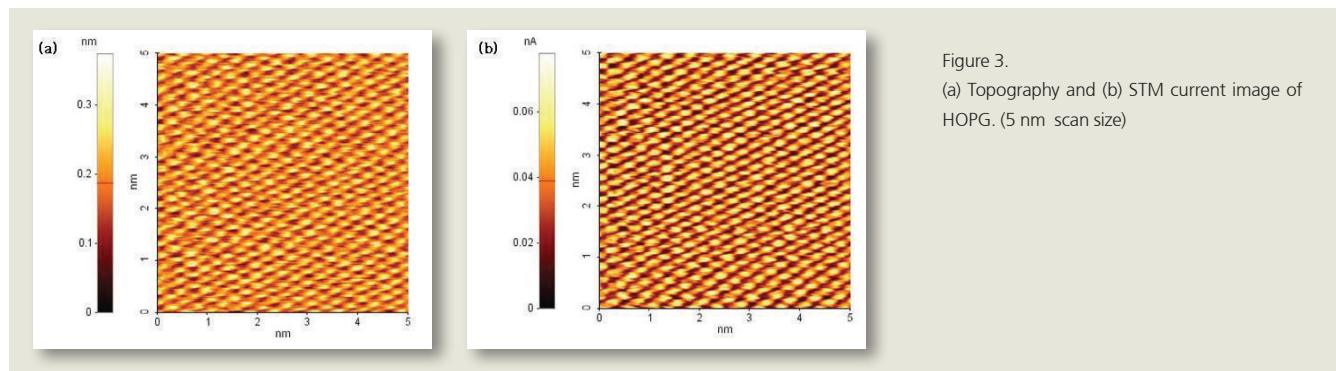


Figure 3.
(a) Topography and (b) STM current image of HOPG. (5 nm scan size)

There are two current amplifiers available for XE-STM modes; 'Internal STM' and 'External STM'. The 'Internal STM' mode refers to the STM mode that uses the fixed gain current amplifier in the head extension module. In the 'Internal STM' mode, the range of the measurable tunneling current is fixed because the gain of the amplifier is fixed. But in the external STM mode, measurable tunneling current range can be changed by varying the gain of the amplifier. The 'External STM' mode refers to the STM mode that uses the external low noise current amplifier with variable gain. (See "External Low Current Amplifier")

I/V spectroscopy mode supports the acquisition of Current (I) vs. Voltage (V) curves to investigate electrical properties of a sample surface. An I/V curve is a plot of the current as a function of the tip bias voltage applied to the sample.

Required Options

Internal
STM
STM Probes & STM Probe Holder
Head Extension module & Frame module

External

STM
STM Probes & STM Probe Holder
External Low Noise Current Amplifier
Head Extension Module & Frame Module

Nanoscience and Nanotechnology: An Introduction

**Junior Research Seminar
Spring 2004**

30 March 2004

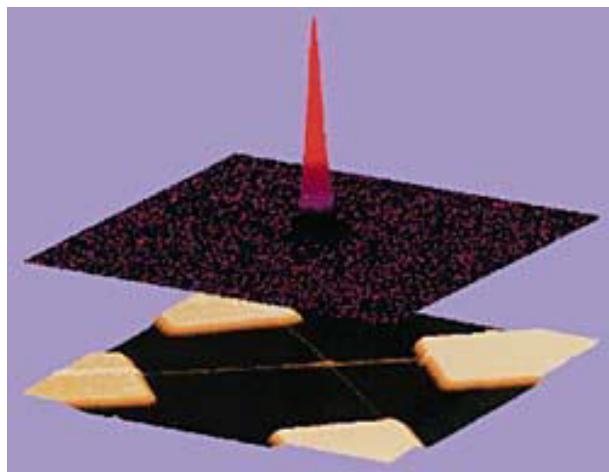
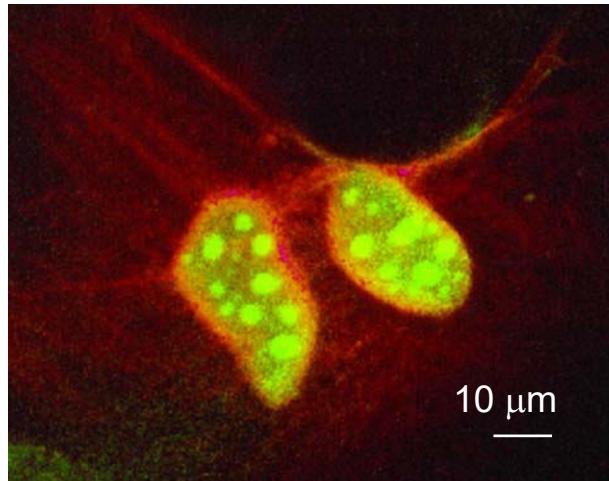


Why *Nano* Rather than Something Else?

- **A popular, understandable (but largely incorrect?) vision**
 - Drexler, Merkle, Joy, Golden, ...: little submarines; the “replicator”
 - The spice of apocalypse: “grey goo”
- **Legitimate scientific opportunity**
- **An inseparable mélange of accomplishment and hyperbole**
- **Unified support from the physical (and biological) sciences**
 - A common “story” for the physical sciences; strong advocates across competing disciplines
- **High potential for technological relevance**
 - Information; military/national security; biomedicine
- **A “New, new thing”: venture capital**



What is *Nano* (i.e. less than 100 nm)?



- Small aggregates/Single atoms
- Quantum phenomena (1-30 nm)
- New materials
 - Ultrahigh surface area
 - Low defects
 - New properties
 - Hierarchical structures: bottom-up and top-down
- Biological/cellular “machinery”
 - ATPase, chloroplast, ribosome...
 - Sensors smaller than a cell
- Large numbers of components
 - High-density technologies (e.g., memory)
 - Complex systems

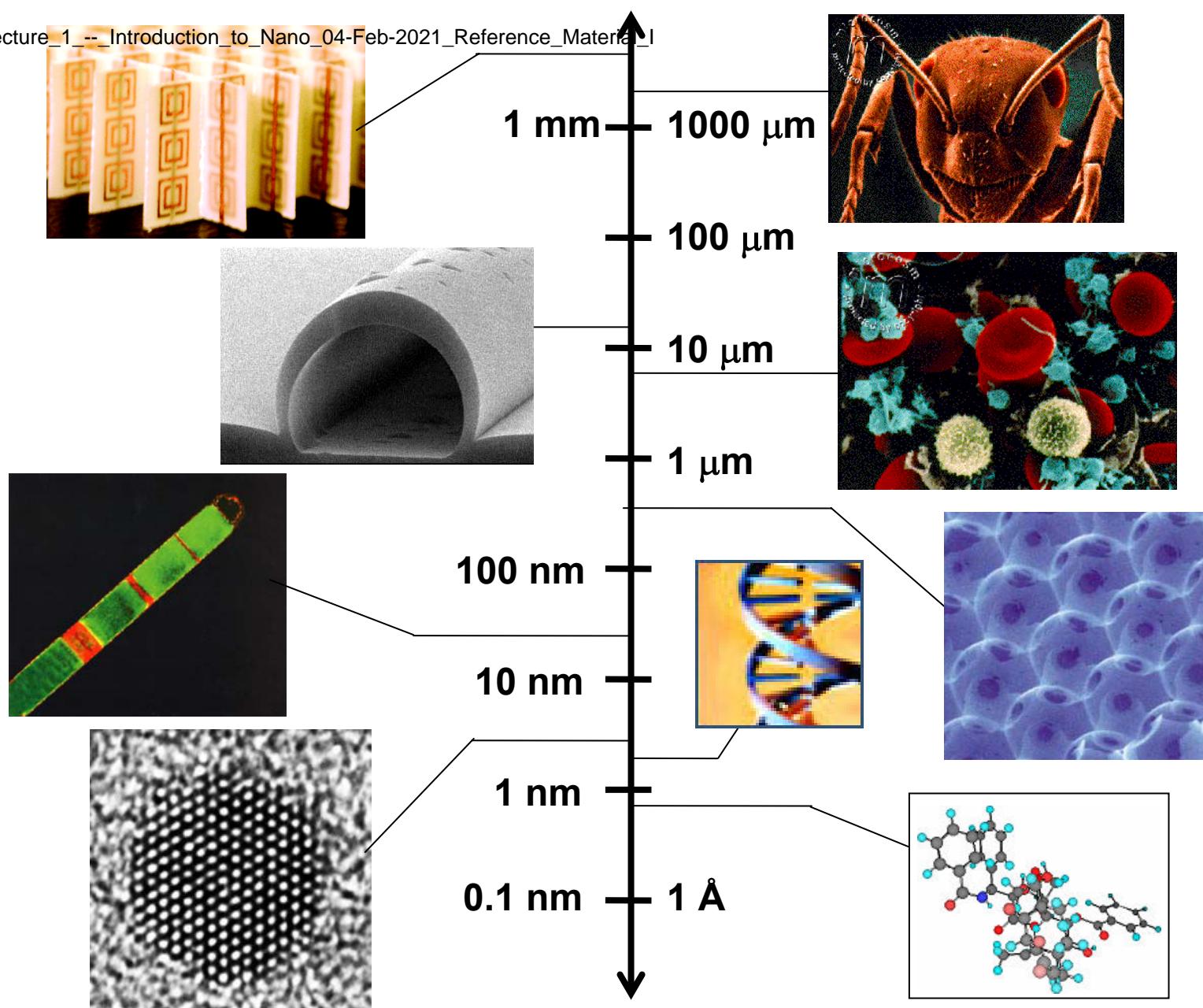
What is Nanotechnology?

Basic research on man-made structures that:

- Have at least one dimension of less than 100 nanometers;
- Are designed through a process that exhibits fundamental control over the physical and chemical attributes of the structures; and
- Can be combined to form larger structures.

* This definition is attributed to Mihail Roco, senior advisor to the National Nanotechnology Initiative.





Nanoscience and Nanotechnology

Science

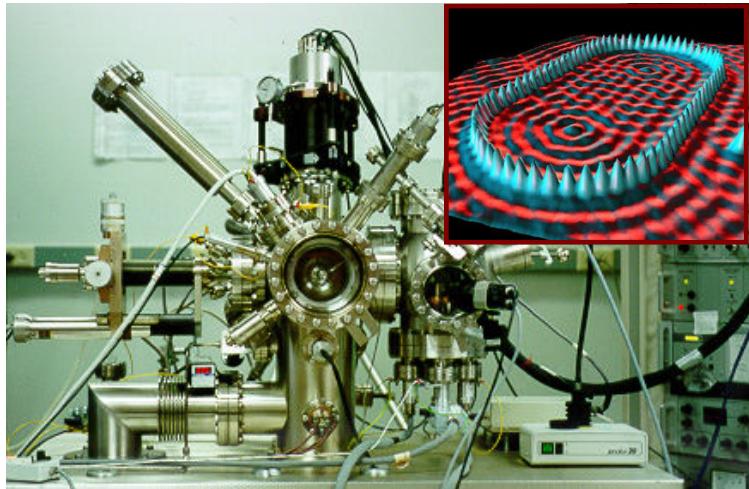
Technology

Patterning

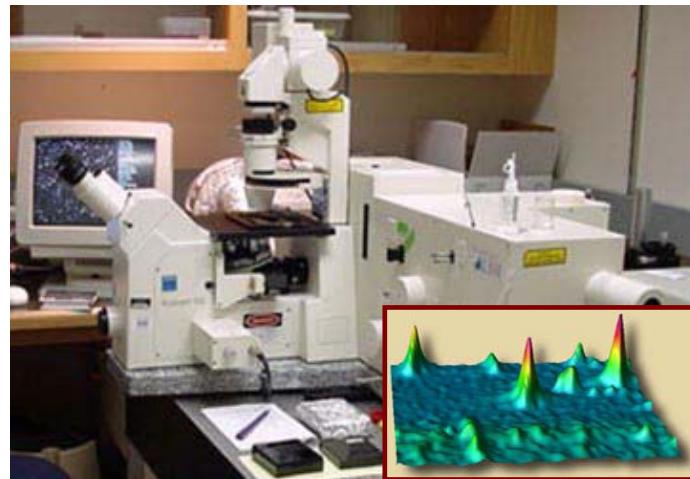
Societal implications



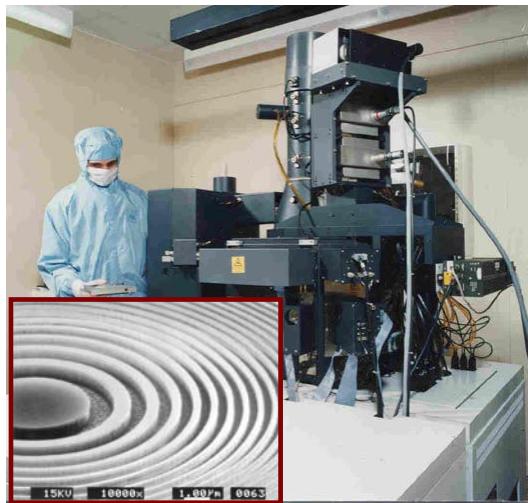
Science: New Tools



Eigler et al., IBM Research

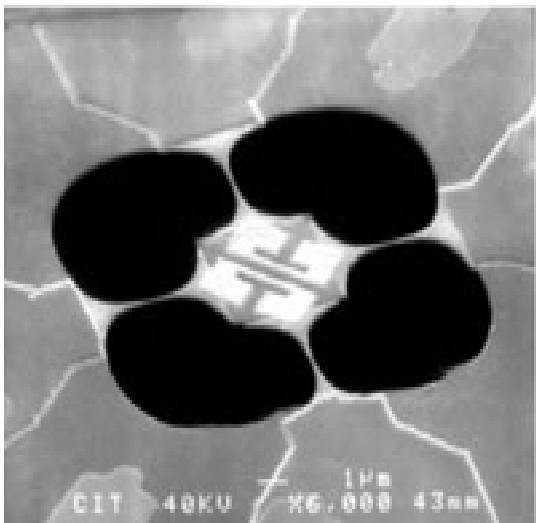


S. Block, Stanford



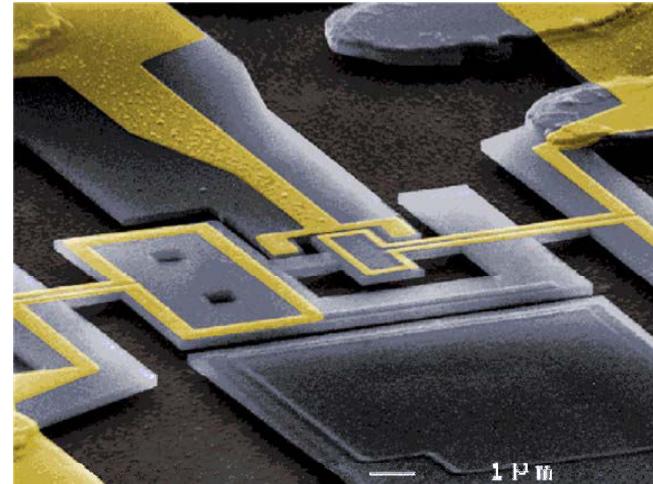
Science: New Physics

Single phonon counting

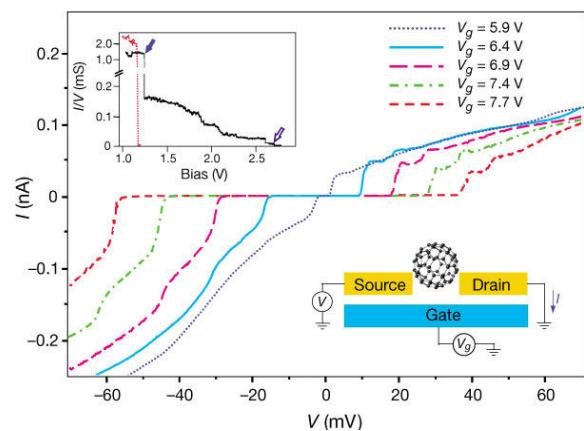


M. Roukes (Cal Tech)
Nature 404, 974 (2000)

Single electron detection through mechanical motion

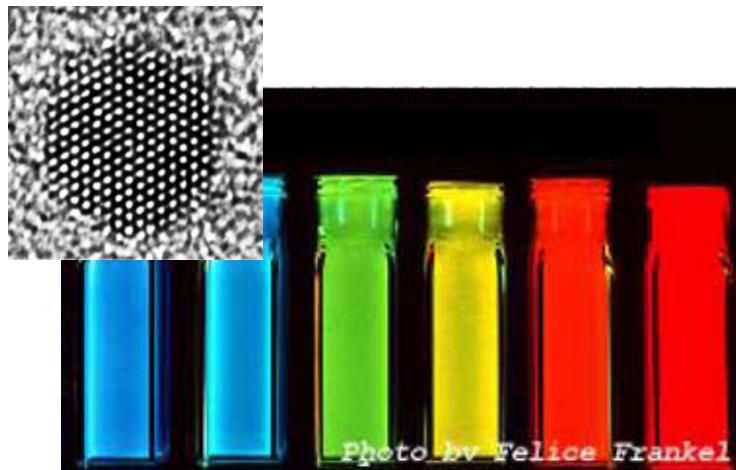


M. Roukes, Cal Tech

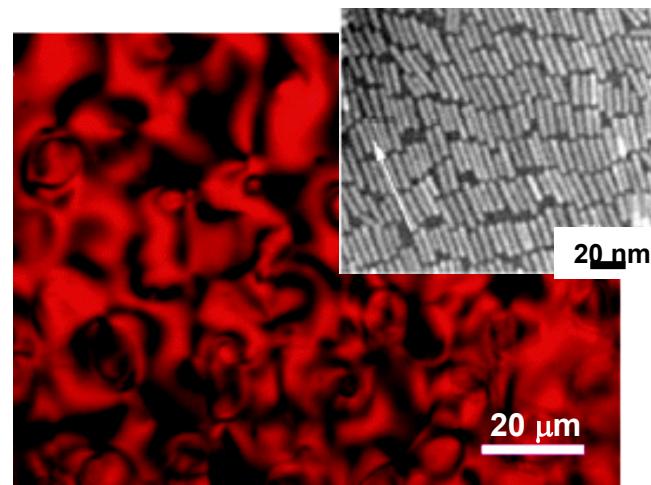


H. Park, P. Alivisatos, P. McEuen, U.C. Berkeley

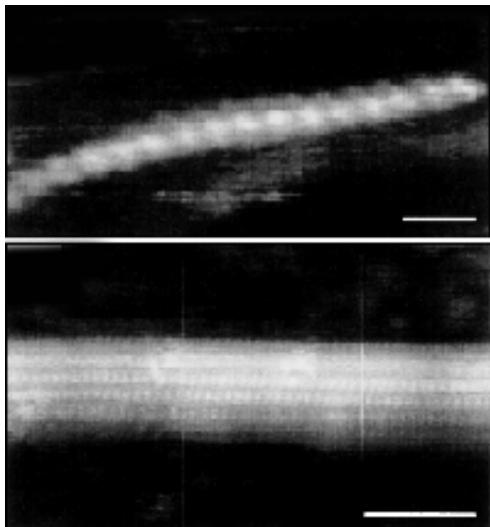
Science: New Chemistry



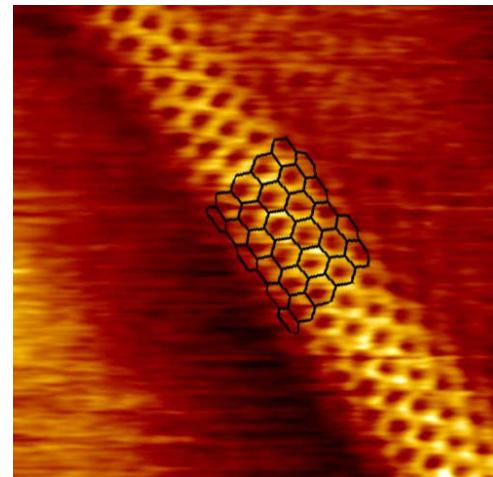
M.G. Bawendi (MIT)



A.P. Alivisatos (U.C. Berkeley)



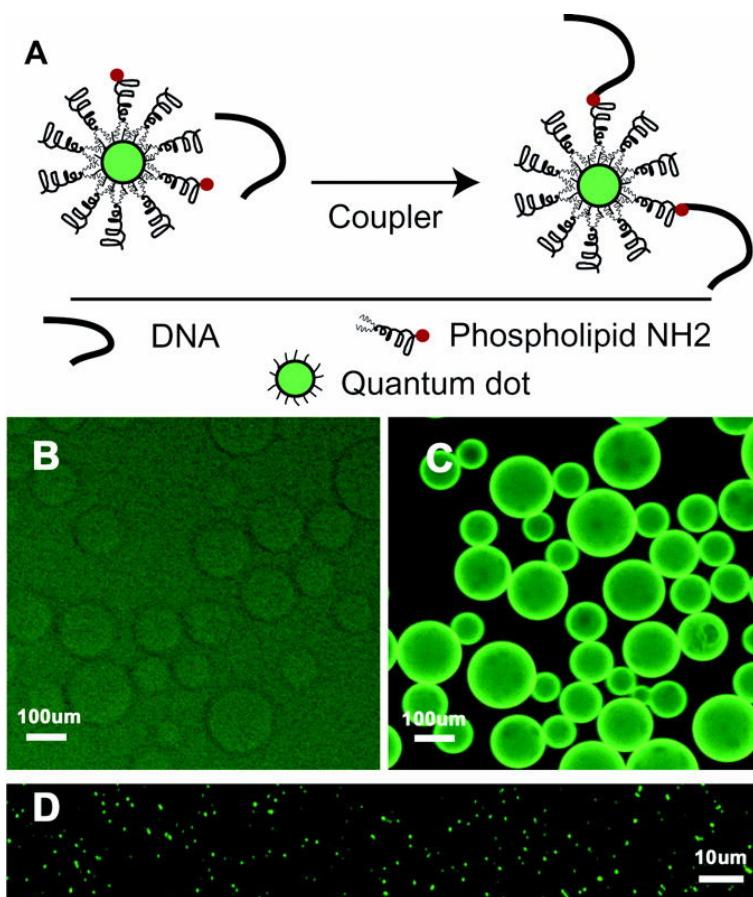
C.M. Lieber, *Phys. Rev. Lett.* 83, 5334 (1999)



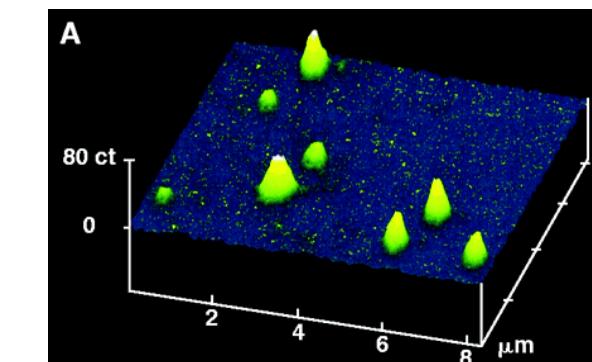
C.M. Lieber (Harvard), *Nature* 391, 62 (1998)



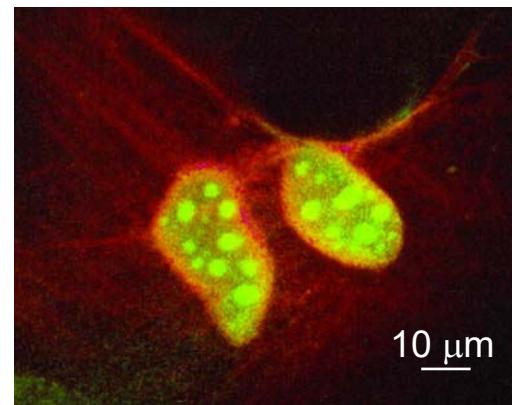
Science: New Biology



A. Libchaber (NEC) *Science* 298, 1759 (2002)



X.S. Xie (Harvard), *Science* 282, 1877 (1998)



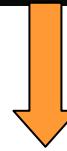
A.P. Alivisatos (U.C. Berkeley), *Science* 281, 2013 (1998)



Size-Dependent Properties



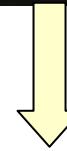
Ag Nanoprisms
~100 nm



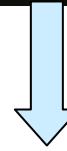
Au Spheres
~100 nm



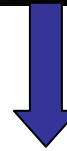
Au Spheres
~50 nm



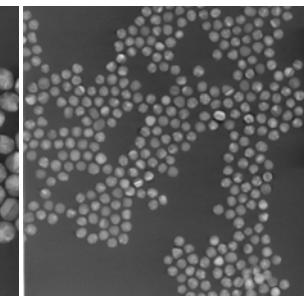
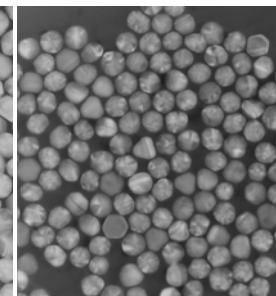
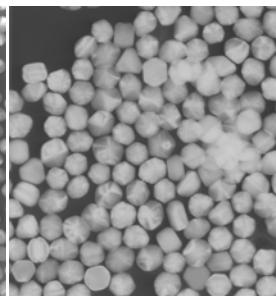
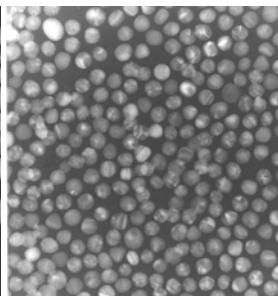
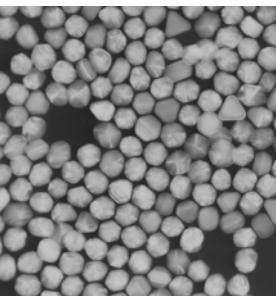
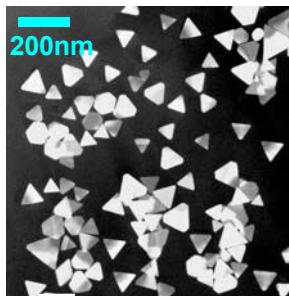
Ag Spheres
~120 nm



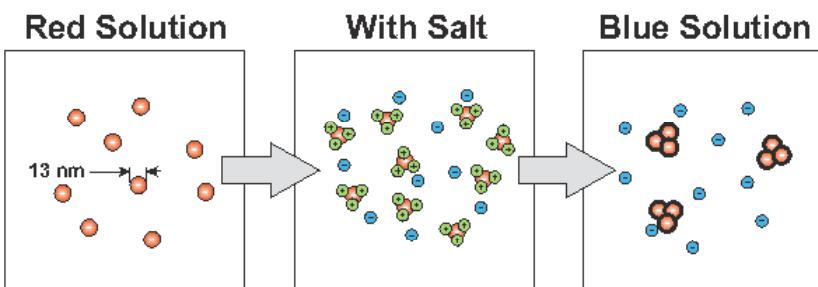
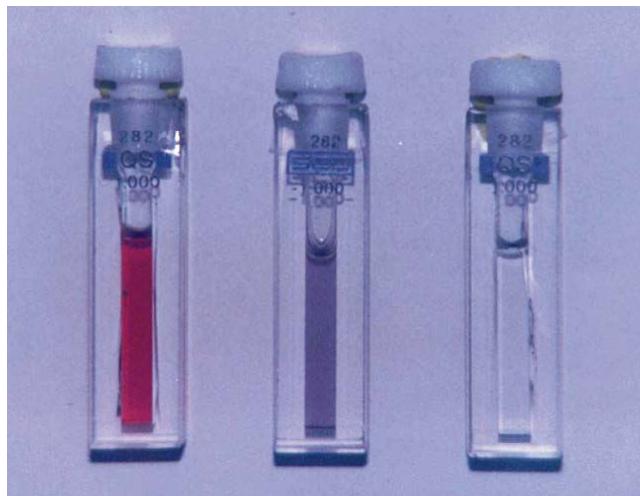
Ag Spheres
~80 nm



Ag Spheres
~40 nm



Bench-top Nanoscale Experiments: *Nanocrystal Synthesis*



Nanoscience and Nanotechnology

Science

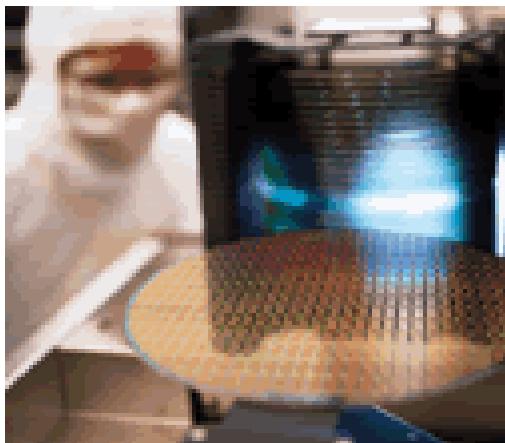
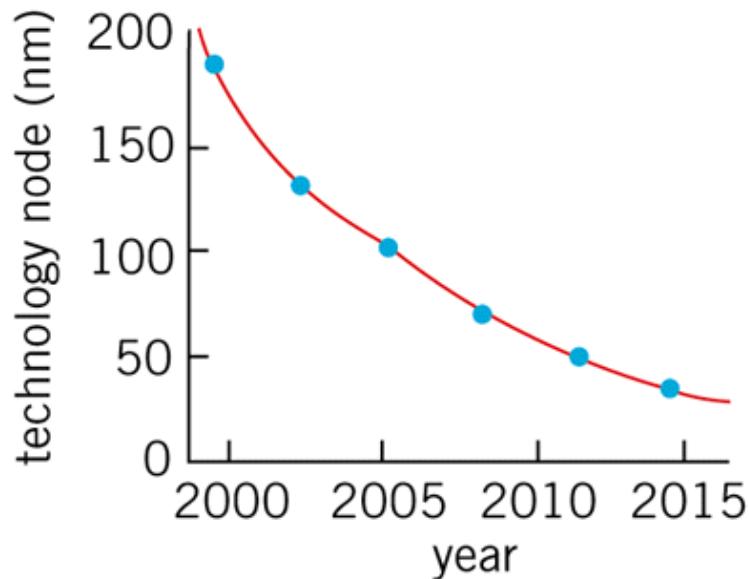
Technology

Patterning

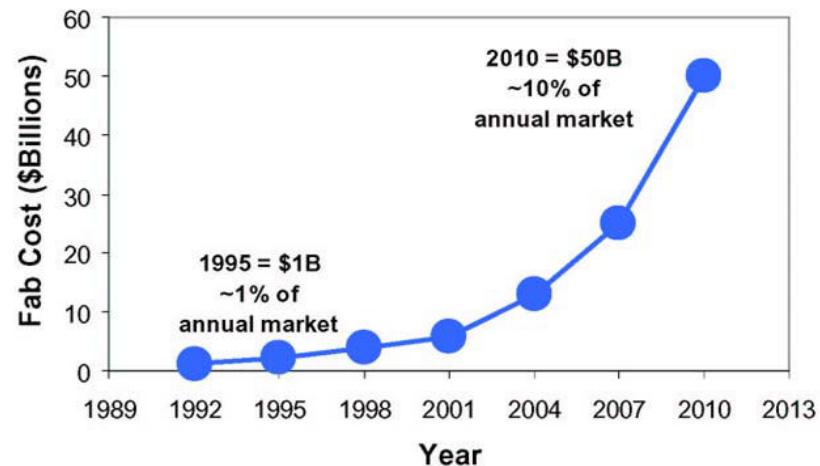
Societal implications



Revolution or Evolution?



Cost of New Fab



Evolution and/or Revolution in Information Systems?

Evolution

- Ultradense memory
- Faster microprocessors
- Customized/specialized microprocessors
- Portable systems
- Medium performance/
Low Cost Systems
 - Organic compounds as conductors and semiconductors

Revolution

- True Nano ICs (<10 nm); single molecule electronics
- Cellular automata
- Quantum computing
- Self-assembling /3D / Biomimetic systems
- Bio-hybrids

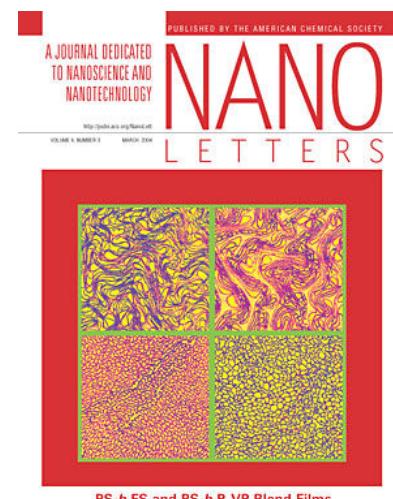


Revolutions in Technology

- ***Microelectronics and Information Technology***
 - *Silicon, the transistor, photolithography, the integrated circuit, displays, microprocessors and memory, software, the laser, optical fiber, the world-wide web*
- ***Biotechnology***
 - *DNA, sex, the double helix, restriction endonucleases, cloning, expression vectors, protein engineering, polymerase chain reaction, the cell cycle, oncogenes, apoptosis, genetically engineered foods*
- ***Nanotechnology***
 - Scanning probe microscopy, carbon nanotubes, nanowires, top-down nanofabrication, colloidal chemistry, structural biology
 - ?? Nano-IC, bio/IT interface, “applied quantum strangeness”/ quantum computation, self-assembly, molecular electronics, mechanical genomic surgery, sophisticated biomimicry, synthetic complexity ??



Nanotechnology in the News



Nanotechnology Start-up Companies



Converting the Promises of Nanotechnology into Reality



Leading-Edge Developer of Nanoclay Technologies for Plastics



Other Targets of Nanotechnology

- Advanced Materials
- Five minute health swab tests
- Interactive glucose sensing
- Human repair
- Genomic Medicine
- Optical Computers
- Self-monitoring food packaging
- Fully target drug delivery
- Heatless lights (LEDs)
- Efficient solar radiation capture



Nanotechnology Today

- **\$22 billion industry**
- **Heterogeneous catalysts**
 - Zeolite MCM-41 for higher octane gasoline
 - Cracking hydrocarbons
 - 40% of gas produced this way
- **Wilson Tennis Balls**
- **Oil of Olay and L’Oreal**
- **NanoPants (Nano-Care Khakis)**



Eddie Bauer



Worldwide Government Spending on Nanotechnology Today

FY 2003

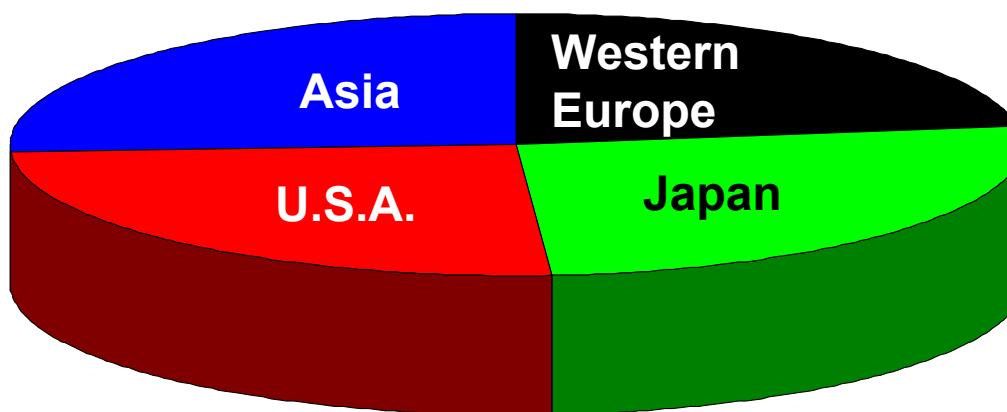
Western Europe: \$699 M

Japan: \$810 M

U.S.A.: \$774 M

Asia: \$800 M

Total Worldwide: \$3.083 B



Source: NSF
www.nanotechfoundation.org



US Government Outlook: 10-15 years

- **Need 1-2 million workers**
- **\$1 Trillion industry**
 - New materials (\$340 Billion)
 - Electronics (\$300 Billion)
 - Pharmaceuticals (\$180 Billion)
 - Transportation (\$70 Billion)



NANOMEDICINE ROADMAP INITIATIVE
Project Launch Meeting
NIH, Masur Auditorium, Building 10, Bethesda, MD USA
Tuesday, May 4, 2004



Nanoscience and Nanotechnology

Science

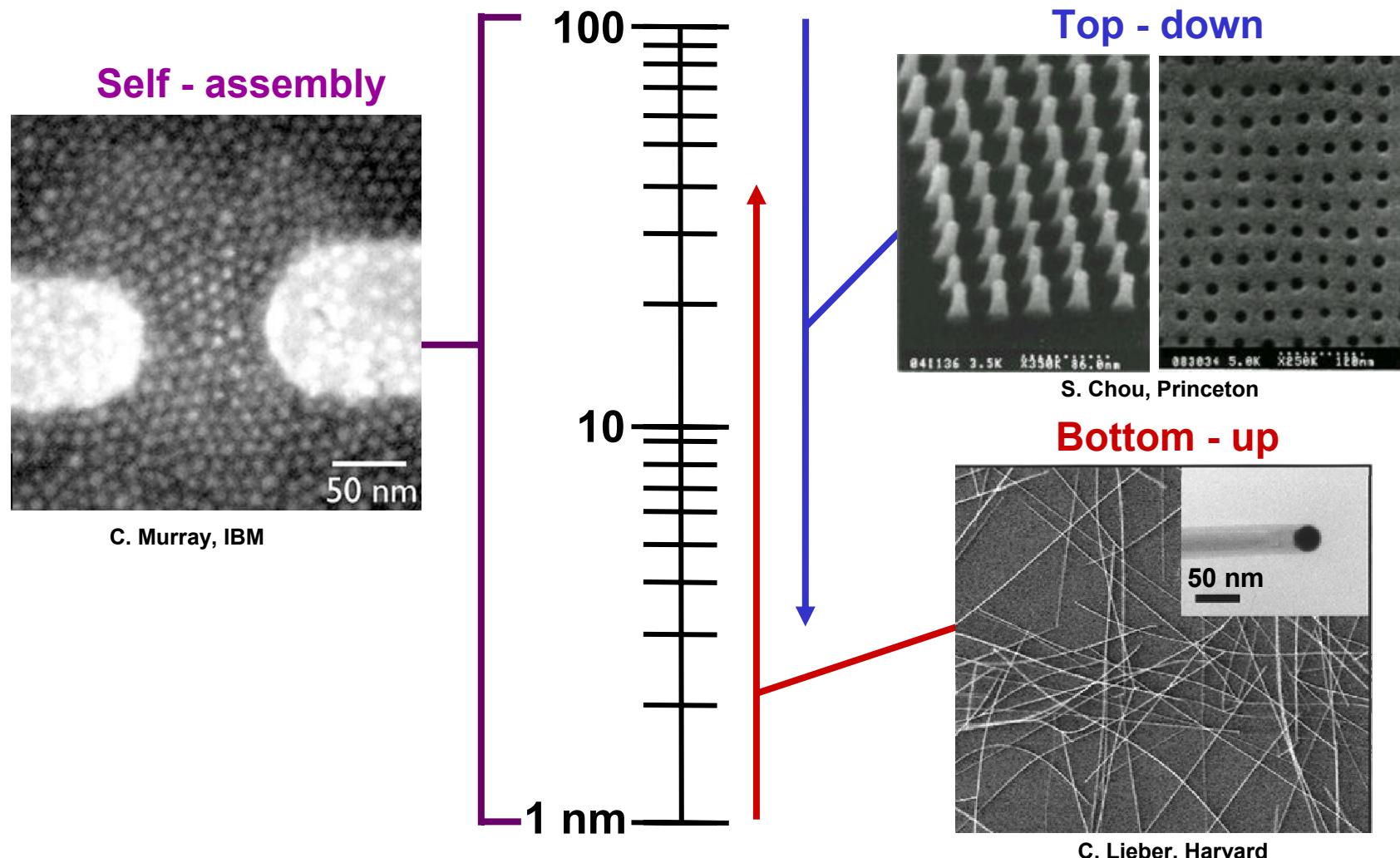
Technology

Patterning

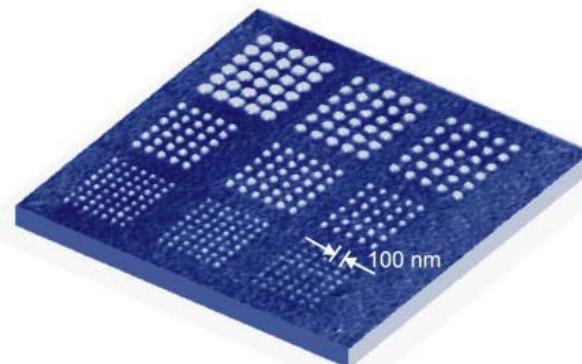
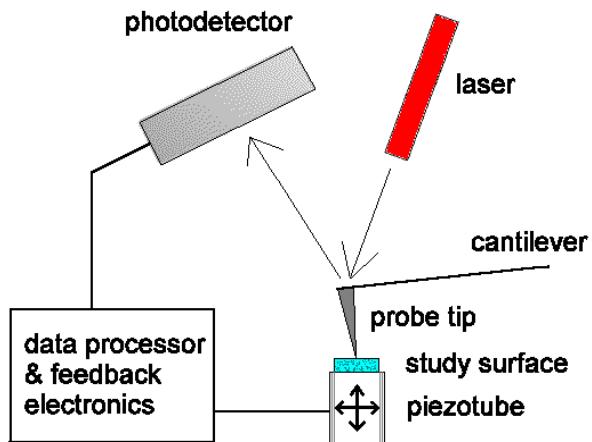
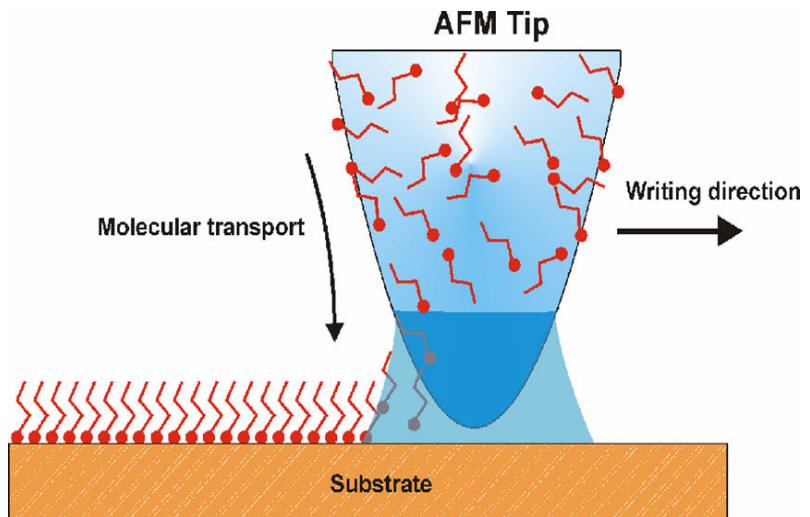
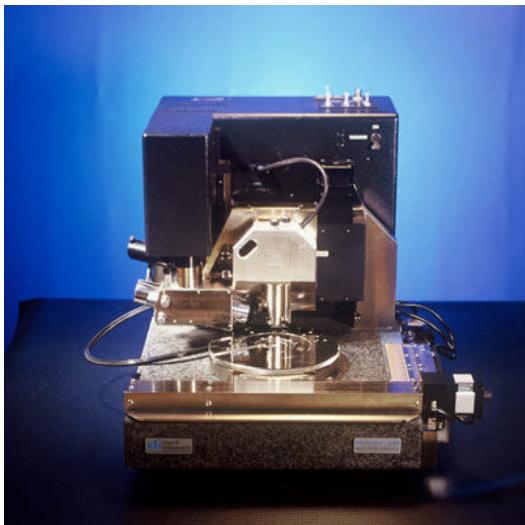
Societal implications



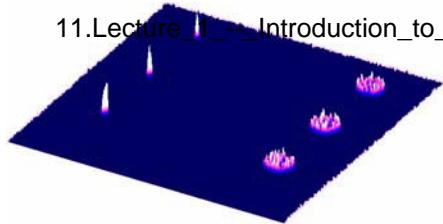
Nanoscale Patterning



Scanning Probe Lithography



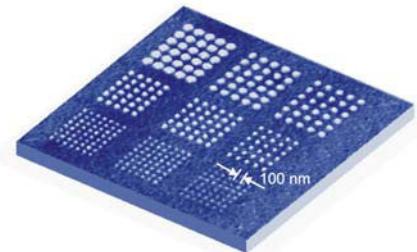
Chad Mirkin, Northwestern



Orthogonal Assembly of Nanoparticles

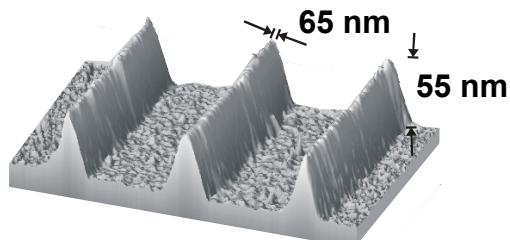
... people tell me about miniaturization, and how far it has progressed today. They tell me about electric motors that are the size of the nail on your small finger. And there is a device on the market... they tell me, by which you can write the Lord's Prayer on the head of a pin. But that's nothing; that's the most primitive, halting step in the direction I intend to discuss. It is a staggeringly small world that is below. In the year 2060, when they look back at this age, they will wonder why it was not until the year 1960 that anybody began seriously to move in this direction.

Richard P. Feynman, 1960



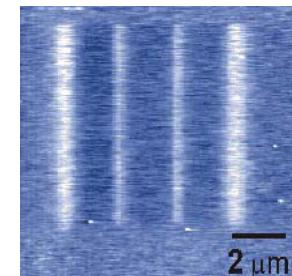
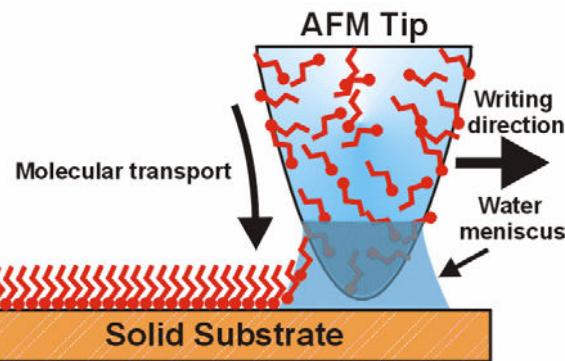
Combinatorial DPN Templates

Flexible Nanolithographic Capability

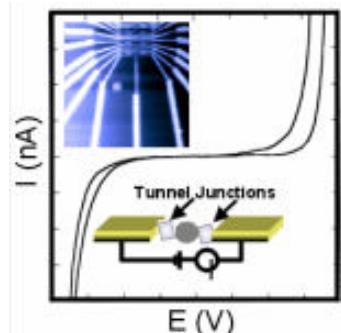


Silicon Nanostructures

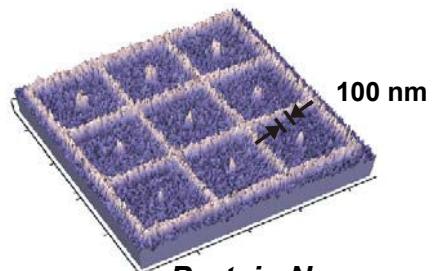
Dip-Pen Nanolithography



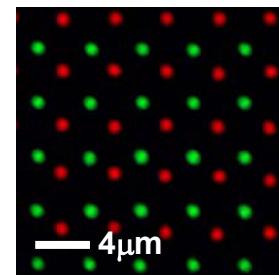
Polymer precursors



Bio-nanoelectrics



Protein Nanoarrays

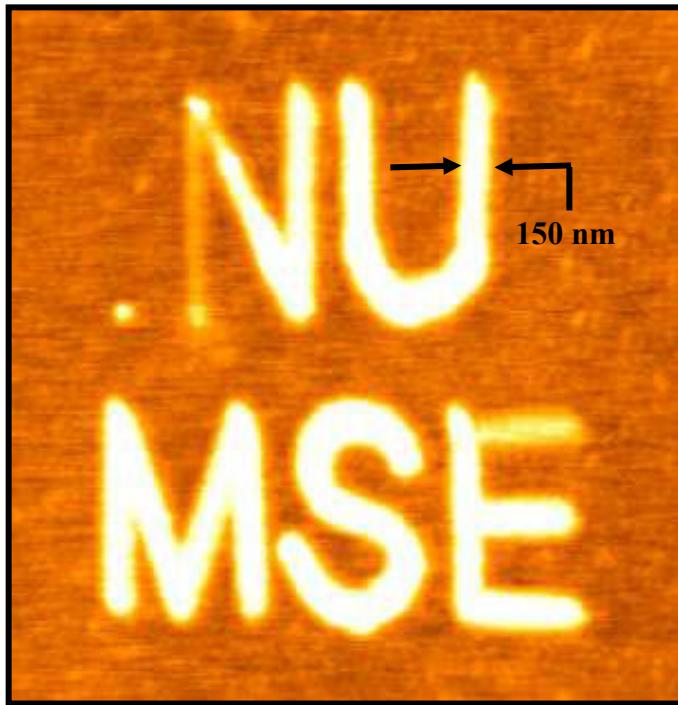


Ultrahigh Density DNA Arrays

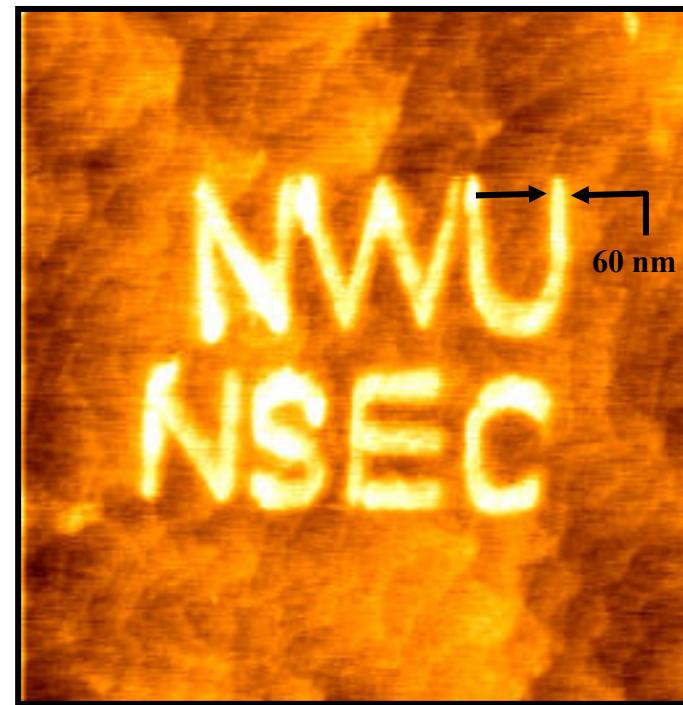
Chad Mirkin, Northwestern



AFM Field-Induced Oxidation



- Lines written on SiO_x
- Variability in line width
- No chemical contrast

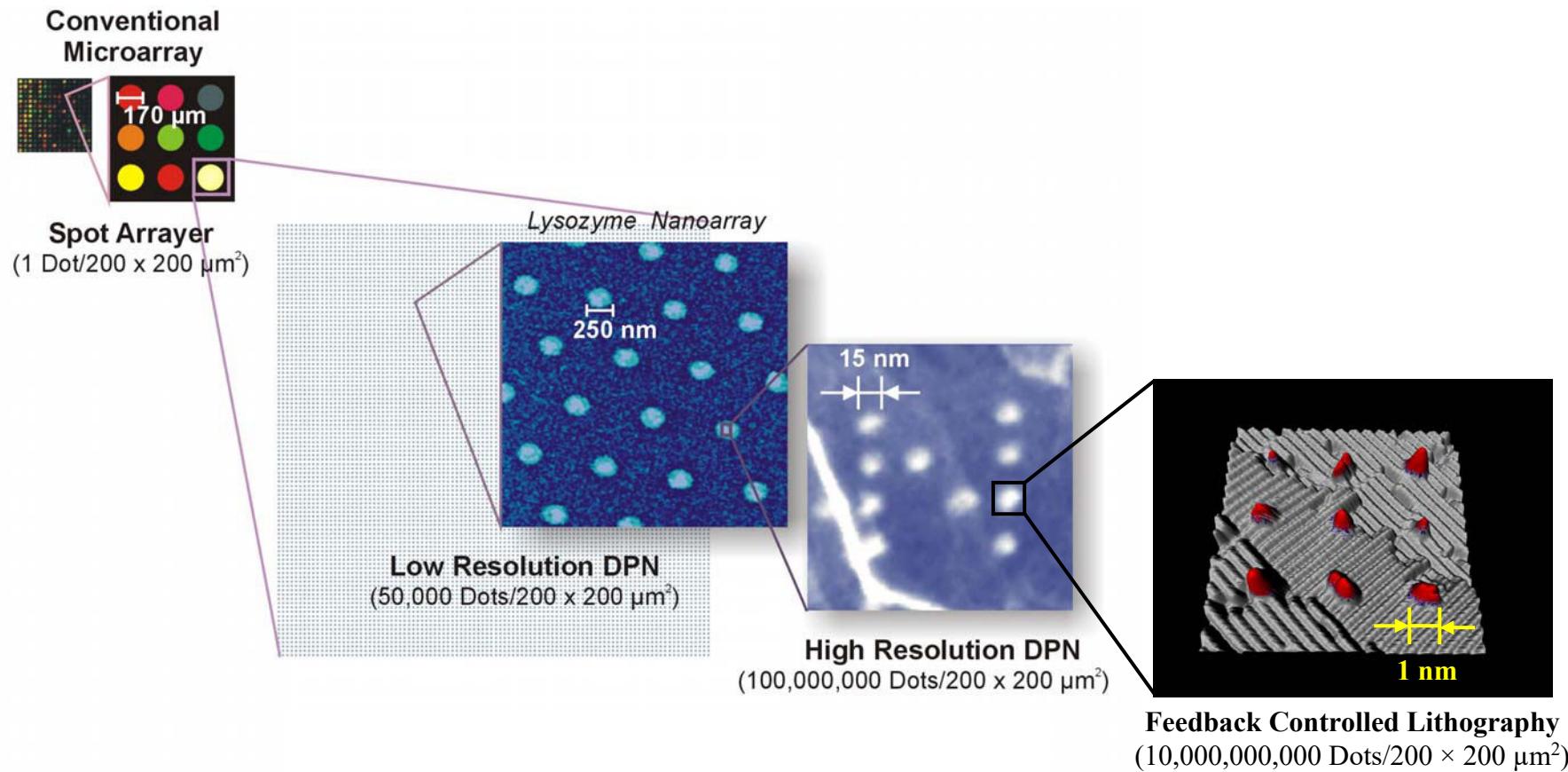


- Lines written at on $\text{Si}(111):\text{H}$
- Reproducibly narrow line width
- Hydrophilic/hydrophobic contrast

Mark Hersam, Northwestern



The Ultimate in High Density Arrays

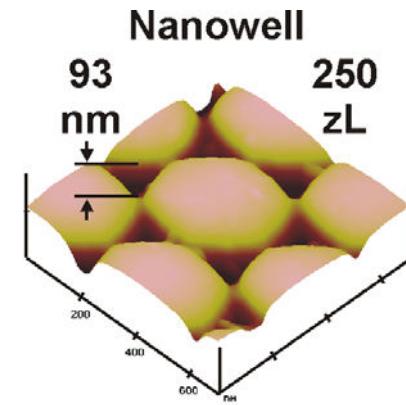
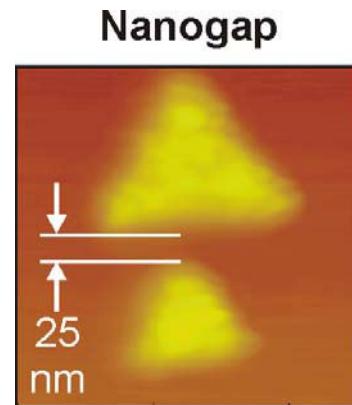
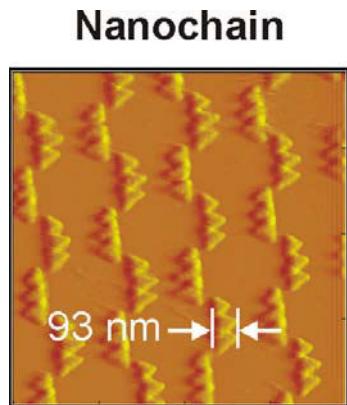


Biological Nanoarrays:

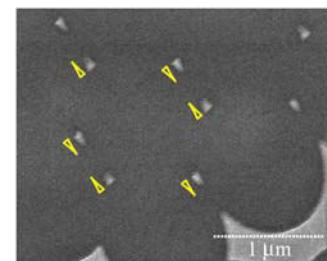
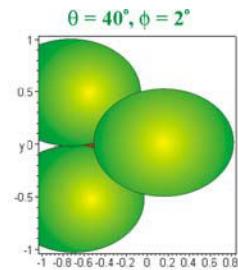
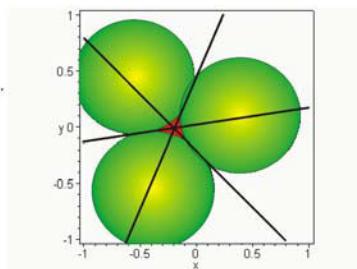
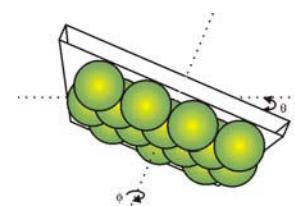
- More than just miniaturization with higher density
- New opportunities for biodetection and studying biorecognition



Bench-top Nanoscale Patterning: *Nanosphere Lithography*



Angle Resolved Nanosphere Lithography: Reduction of in-plane nanoparticle Size by a factor of 3 - 4

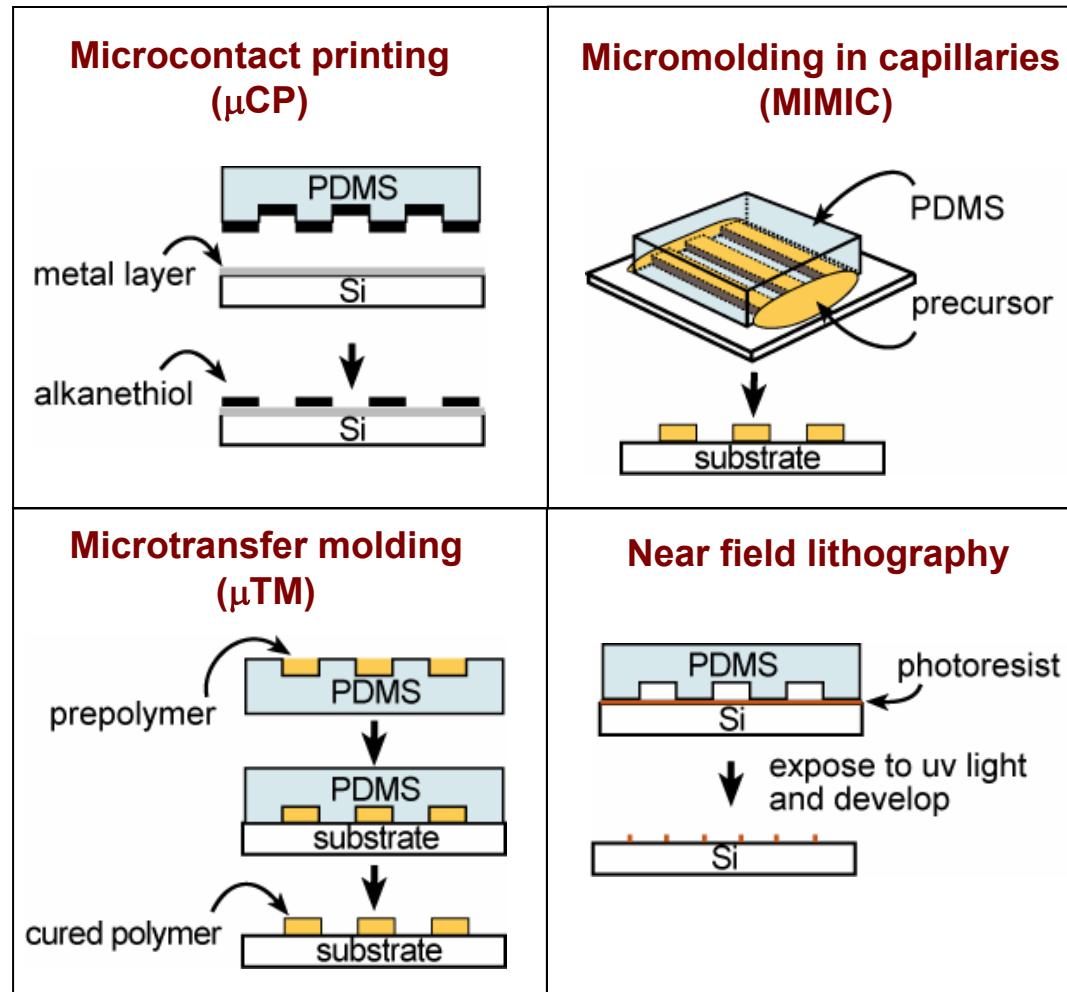
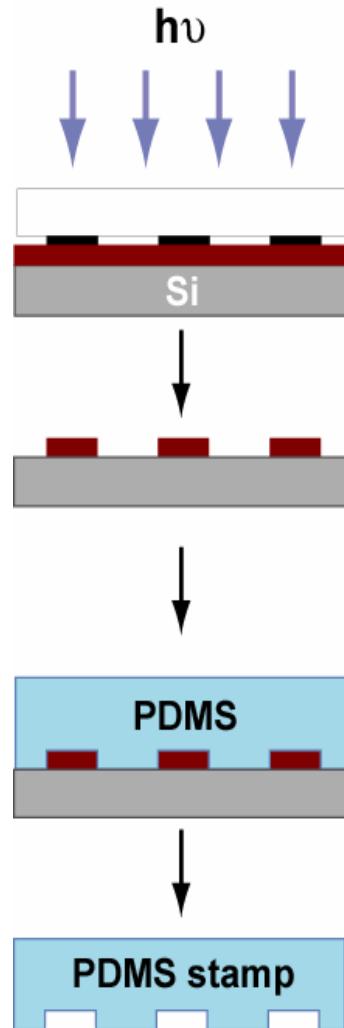


Rick Van Duyne, Northwestern



Bench-top Nanoscale Patterning:

Soft Lithography



Nanoscience and Nanotechnology

Science

Technology

Patterning

Societal implications



Issues in Policy

- **Research: What (and How) to Support?**
 - Balancing applied, fundamental, and exploratory research and development; National Initiatives
 - Nanobots/“Grey Goo”/Assemblers
 - Emphasis on “nano” may detract from “micro” (μ -fluidics, MEMS, μ -optics, μ -TAS...)
- **Commercialization: Involving Business**
 - Small Start-ups: Innovation
 - Large Businesses: Development, manufacturing, distribution
- **Education**
 - Can one pour new wine from old bottles?
- **Public Perception**
 - Hyperbole and Reality; Risk

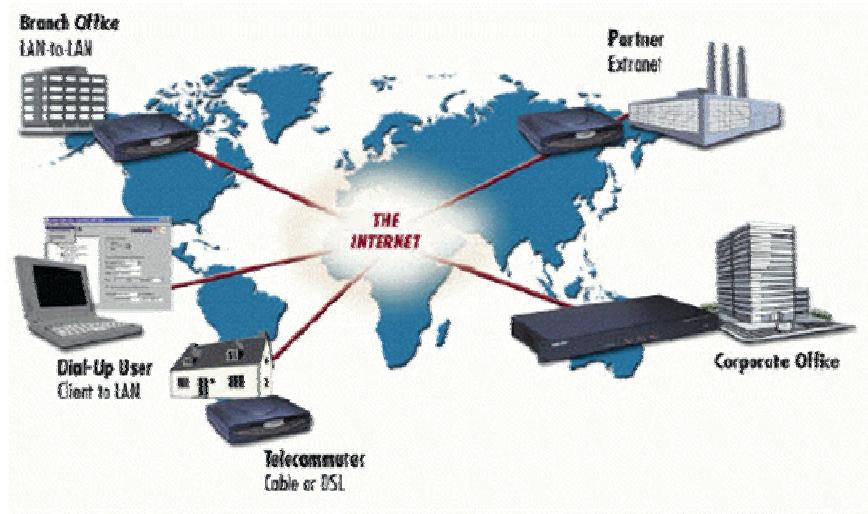


Impacts on Society

- **Portable Technology**
- **Strengthened Capitalism**
 - high-quality jobs — for better or worse
- **Improved National Security**
 - in an age of asymmetric warfare and terrorism: global surveillance; universal awareness
- **New Understanding of Nature**
 - complex systems, materials, biomachines, single molecules
- **Loss of Privacy**
 - very large databases; quantum computation; decryption; universal genomics.
- **Alienation**
 - “Any technology, if sufficiently advanced, is indistinguishable from magic”
- **Global Economic/Technological Segregation**



A Culture of Connectivity

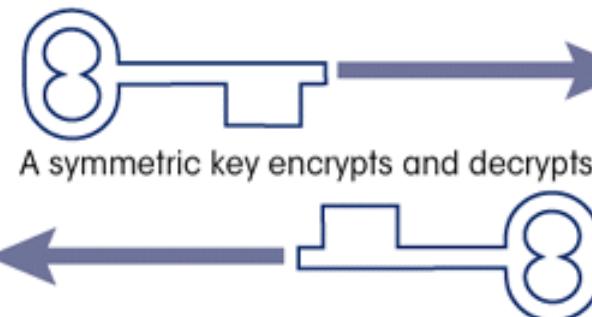


Privacy and Information



Plaintext

```
<html><head><title>  
Welcome to the NCSA  
</title></head>  
<body>Welcome to the  
National Center for  
Supercomputing  
Applications at the  
University of  
Illinois. <p><em>  
Administration and
```



Ciphertext

```
MIIBKjCB5VQQKECBkTEL  
MAkkMRswBhMCVVMxEcTC  
1NNVBAgTCKNhG1mb3Ju  
aWExFDASBgNVBAcTC1Nw  
cmLuZ2ZpZWxkMRswGQYD  
VQQKExJSYW5kb20gQ29y  
cG9yYXRpb24xIzAhBgNV  
BhMCVVMxEcTC1NNVBAgT  
CkNhG1mb3JuaWExFDAS  
BguZpb24i02MRsZpZWxxk
```

Information



- 2010: 15 Petabits (10^{16}) / \$250,000
- Human Genome: 10 Gigabits (10^{11})

For a few million dollars, one could store the complete genome of every American and European

...for several more, could add credit card records, telephone logs, travel history, etc...



Where Does *Nano* Stand?

- **Exciting science**
 - New discoveries are the current push
 - “Life” may be the ultimate inspiration
- **A high level of hyperbole**
- **Nanotechnology is already developing**
 - Electronics and materials
 - Contributions from discovery **and** development
- **A high potential for eventual impact on society**
 - Information, genomics, privacy

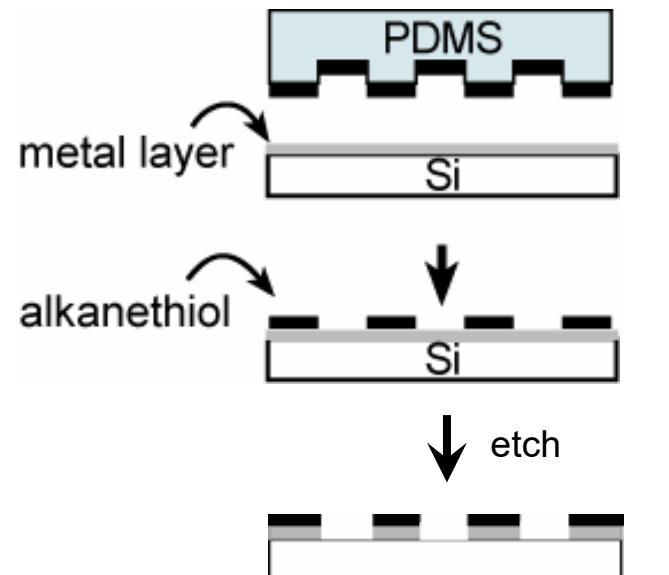
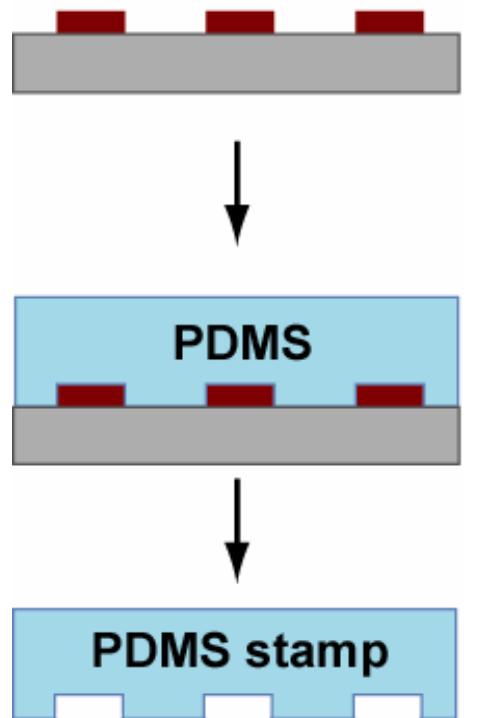


Lab 1: Introduction to Bench-top Nanoscale Tools

- **Making Poly(dimethylsiloxane) stamps**
- **Handling tweezers**
- **Cutting silicon wafers**
- **Microcontact printing**
- **Wet chemical etching**

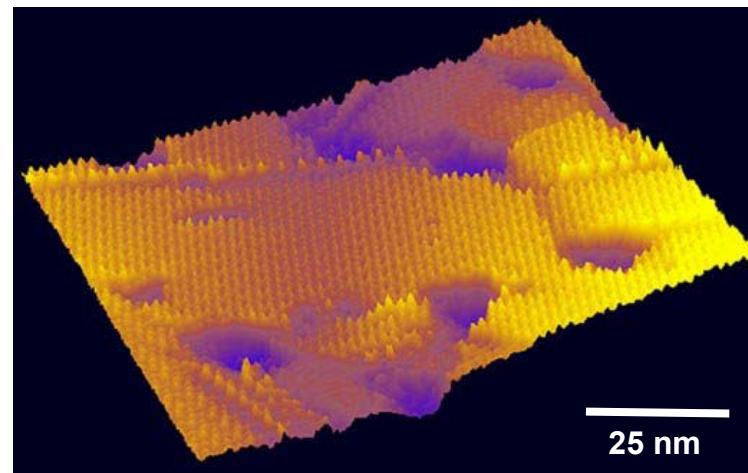
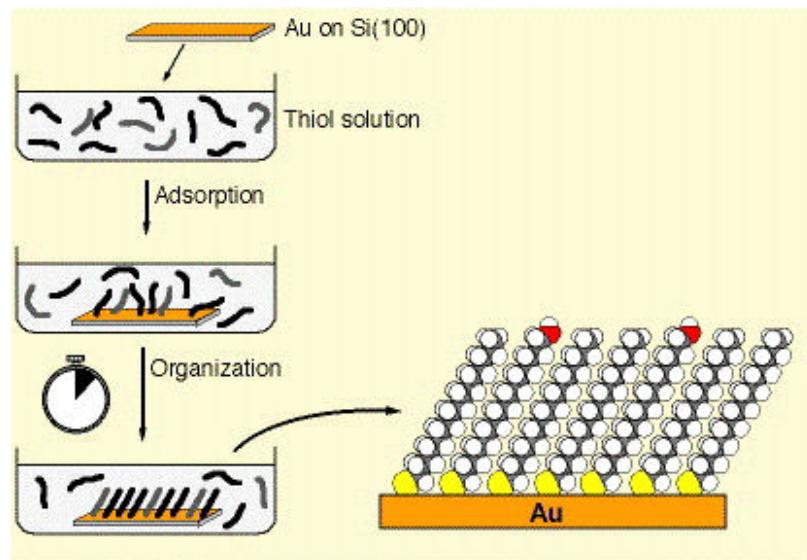


Micro-contact Printing and Etching



Self-assembled monolayers (SAMs)

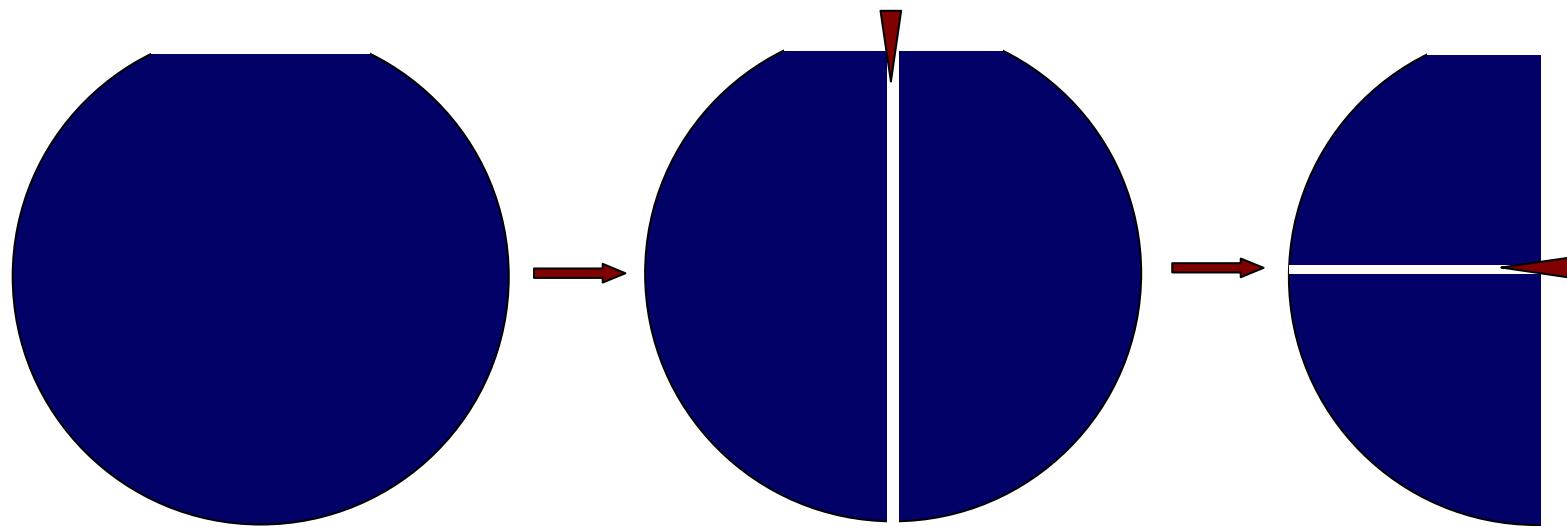
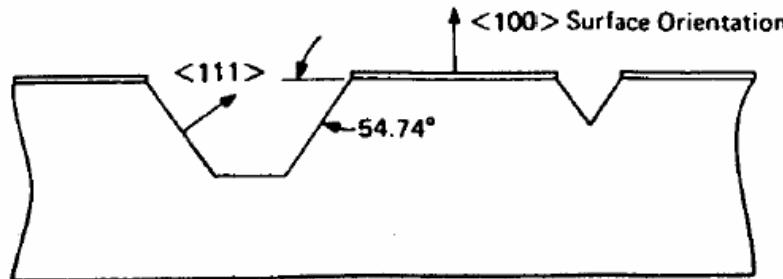
- **Alkanethiols on Au, Ag, Cu**
 - Sulfur chemisorbs to noble metal surface
 - Chains with C₁₆₋₁₈ pack with fewer defects
- **Directed chemical assembly**

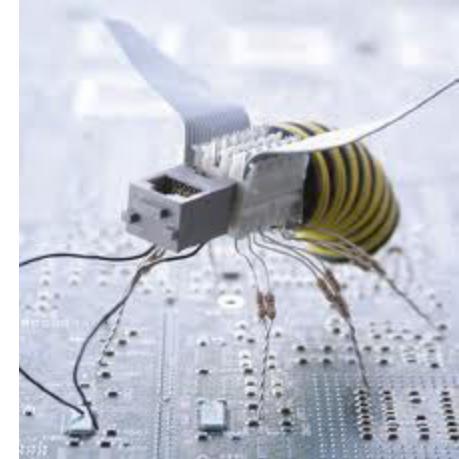
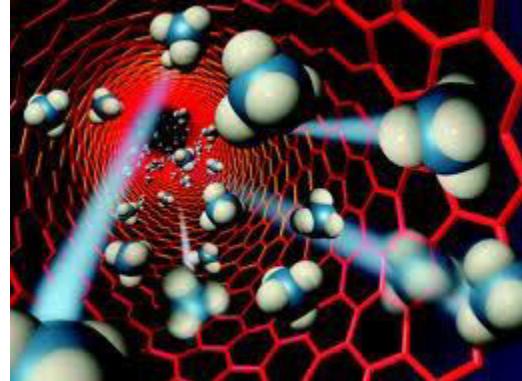
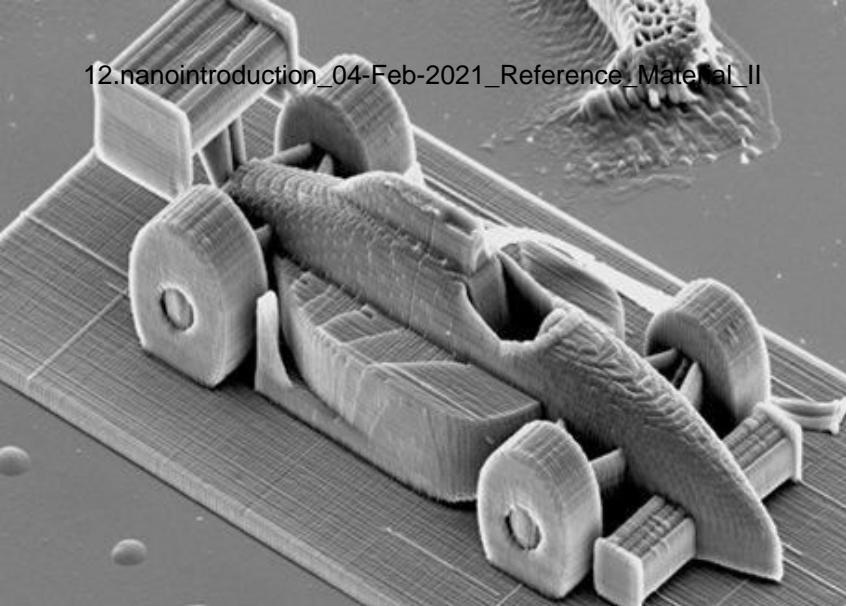


<http://www.ifm.liu.se/Appiphys/ftir/sams.html>

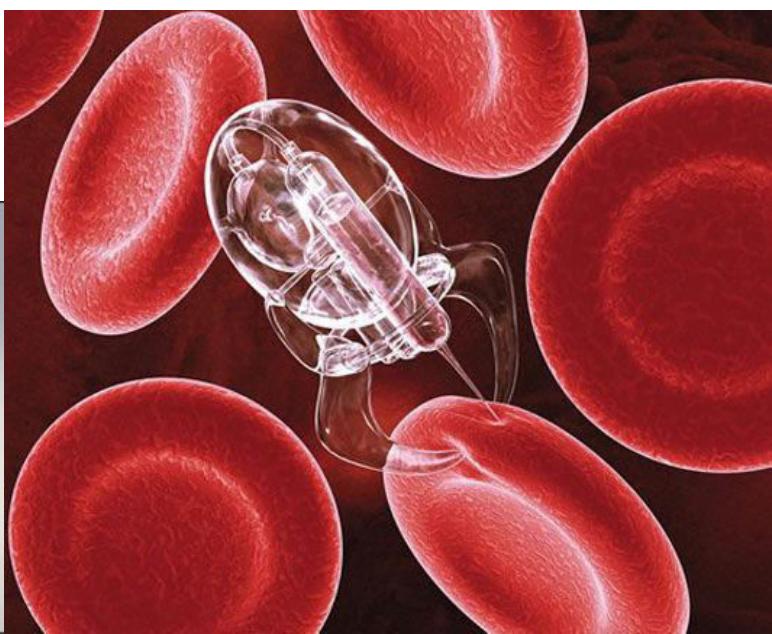
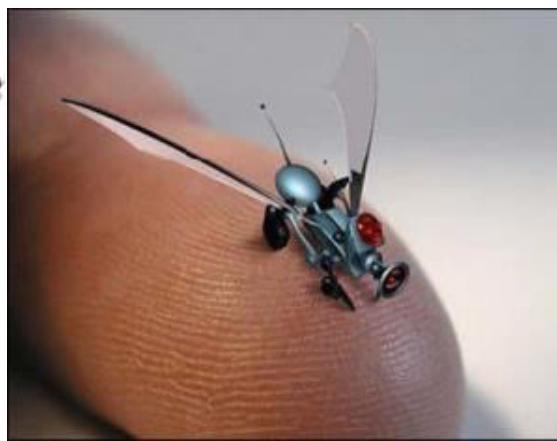
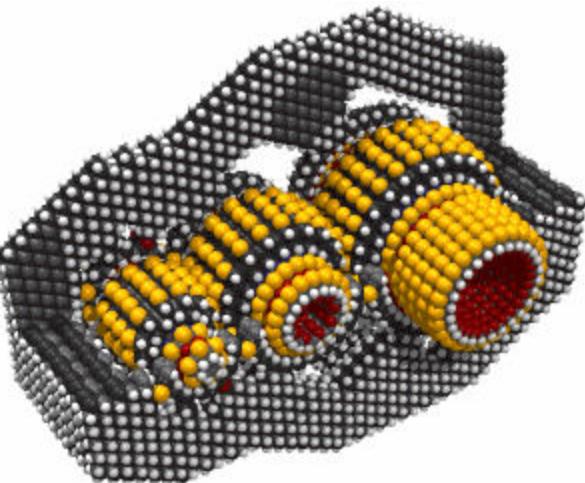


Cutting Silicon Wafers





NANO TECHNOLOGY



What is Nano??

- *Nano is the flavor of the day.*
- *Nano means one billionth or 10^{-9}*
- *Nano refers to the scale of nanometers.*
- *This is the scale of molecules, proteins, and other nano-objects that are the topics of this course.*
- *The Nanoscale involves the range from approximately 100 nm to 1 nm.*

A few more familiar examples may convince you of the difficult in imaging the size of nano-objects

A single strand of human hair is around 50,000 nm in diameter

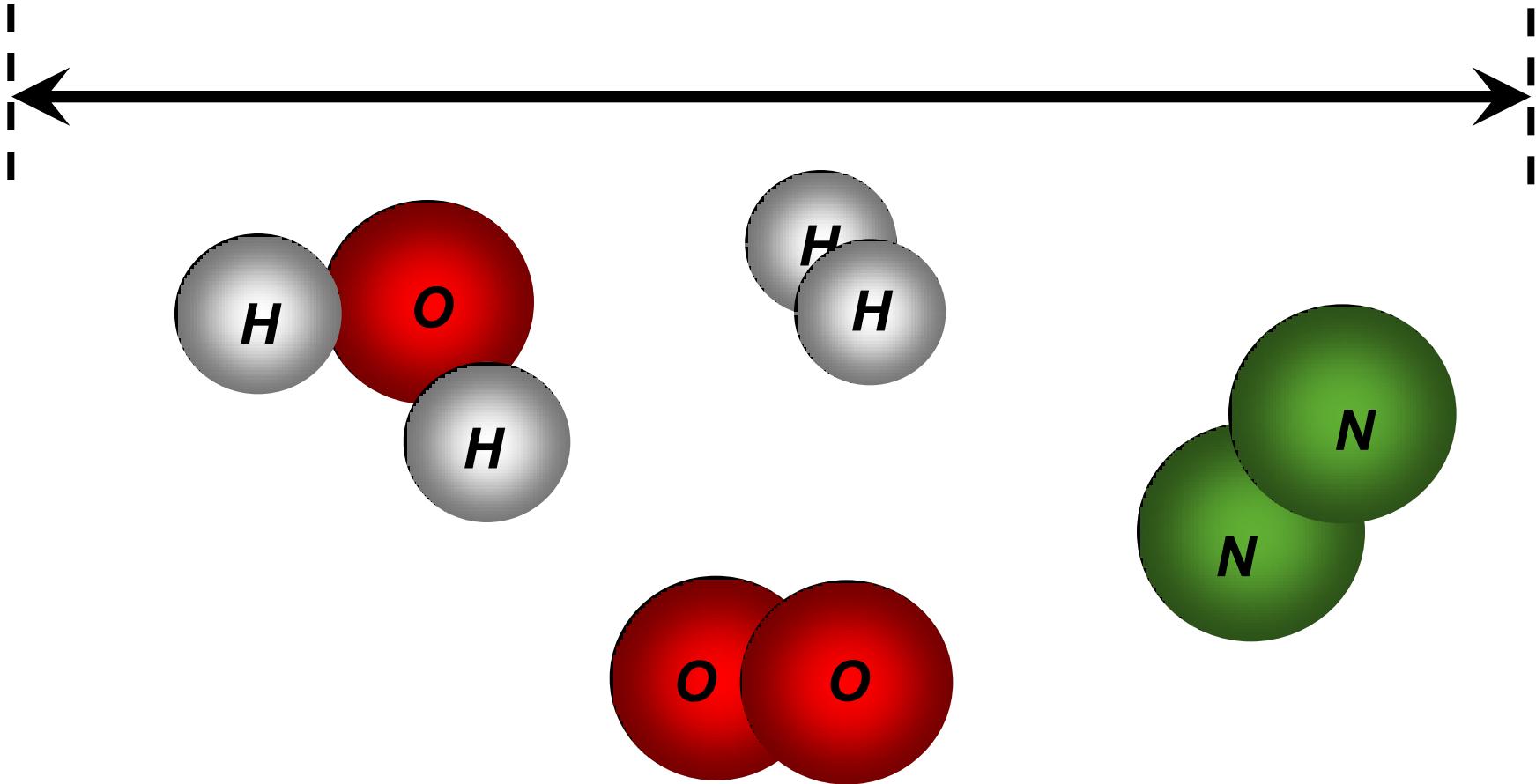
The population of India is one billion or >100 crore. Each Indian – you or me – is nano in comparison with population in India

10 Hydrogen atoms in line

Nanometer is the scale used to measure objects in the nanoworld

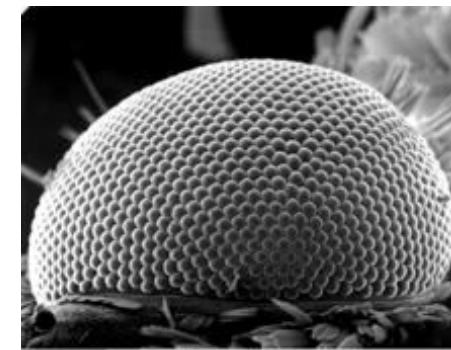


One Nanometer



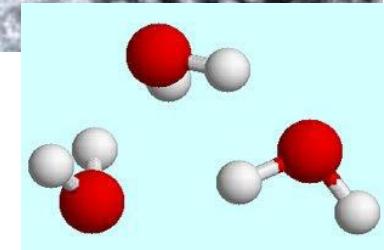
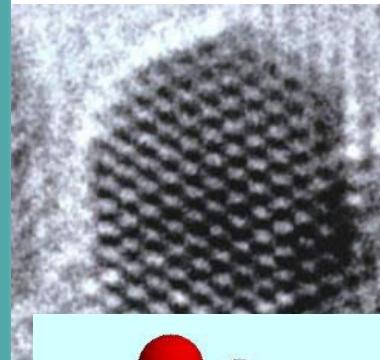
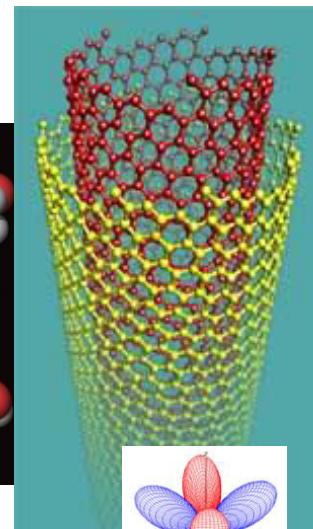
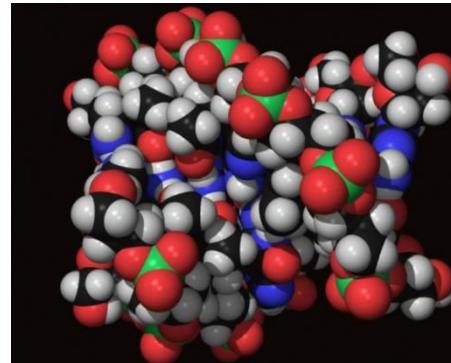
Also, 1 Angstrom (\AA) = 0.1 nm

Distances between atoms in molecules are measured in Angstroms (\AA)



“Macroscopic” Scale

Nanoscale



Subatomic scale:
Nuclear Physics

Partical Physics

atoms

Molecular / Atomic Scale

Examples of Nanoscale Structures

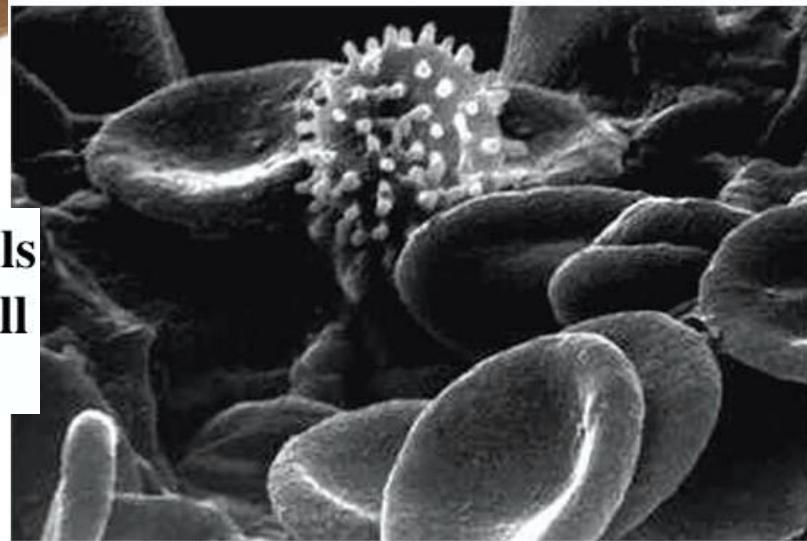
Human hair

~50–150 μm wide



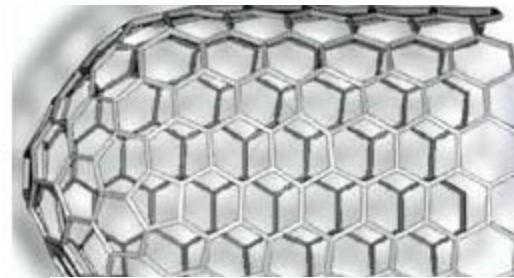
**Red blood cells
with white cell**

~2–5 μm



DNA

~2.5 nm diameter

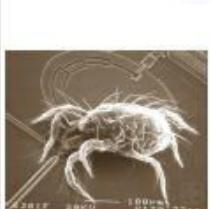


Carbon nanotube

~2 nm in diameter

The Scale of Things – Nanometers and More

Things Natural



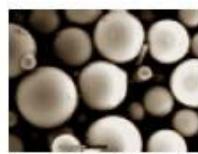
Dust mite
200 μm



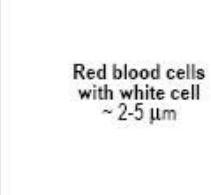
Ant
~5 mm



Human hair
~60-120 μm wide



Fly ash
~10-20 μm



Red blood cells
with white cell
~2-5 μm



Things Manmade



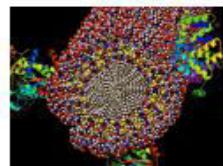
Head of a pin
1-2 mm



MicroElectroMechanical (MEMS) devices
10-100 μm wide



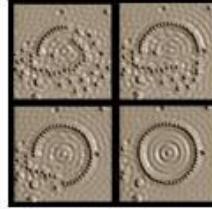
Pollen grain
Red blood cells
Zone plate x-ray "lens"
Outer ring spacing ~35 nm



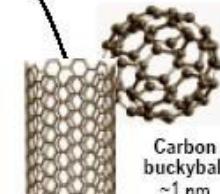
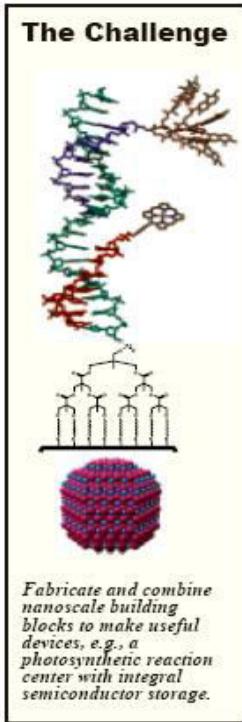
Self-assembled,
Nature-inspired structure
Many 10s of nm



Nanotube electrode



Quantum corral of 48 iron atoms on copper surface
positioned one at a time with an STM tip
Corral diameter 14 nm



Carbon buckyball
~1 nm
diameter

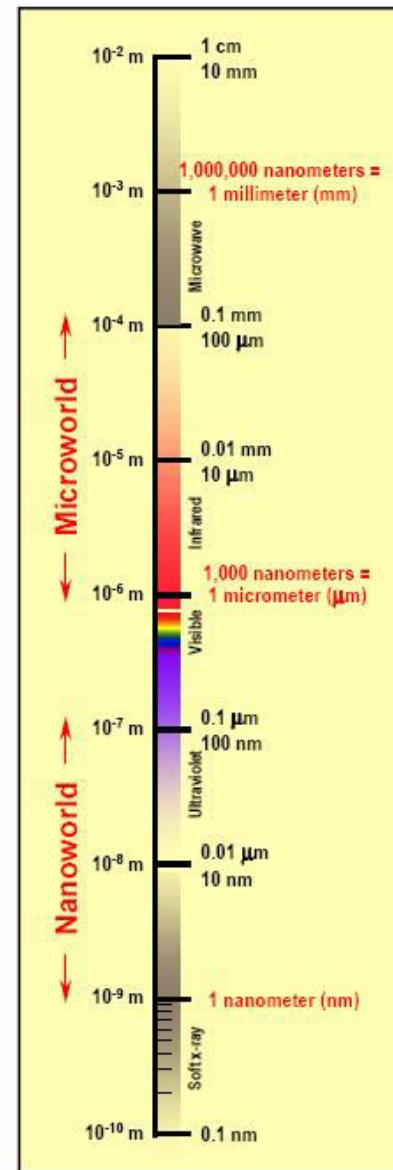


Figure 10.3: The Scale of Things Natural and Manmade. Source: DOE

Above the line:

We can still use light-based “Micro-fabrication” techniques

And even though they were developed for electronics,

they are now also applied to making all sorts of micro things!

Below the line:

NO longer able to use Micro-fabrication

Replacement would be called “Nanofabrication” or “Nanotechnology”

What is Nano Science??

- Synthesis the novel nano materials
- Study the all the properties (chemical, physical and electrical etc,)
- Characterize in different ways

“So where does Nanotechnology fit in the curriculum?

On one hand, it is not Physics, Biology, or Chemistry.

On the other hand, it is all of them !

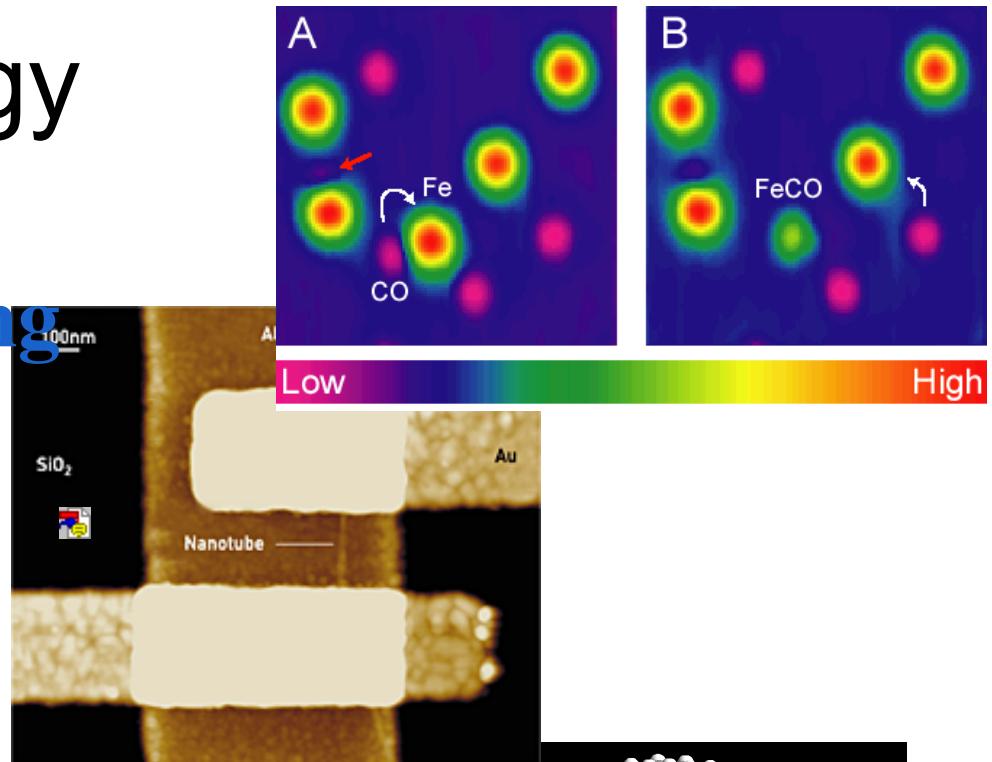
Is nanotechnology a subject of its own, or is it just a way of thinking about other subjects?”

What is Nanotechnology?

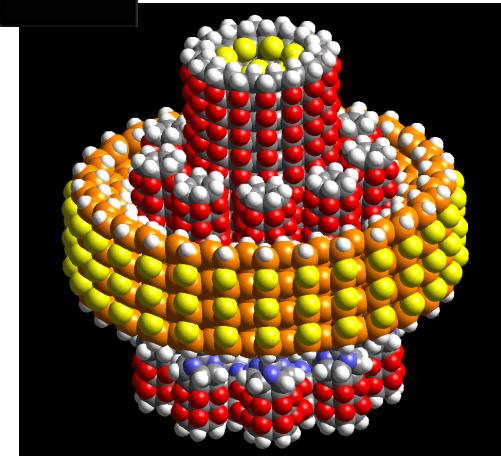
- *It is science and engineering performed at the nanoscale, i.e., at the molecular and atomic scale.*
- *The result is that materials have different properties and behave in unique ways at this small scale.*
- *At this small scale, the effects of quantum physics become more prominent than classical physics.*
- *Nanotechnology provides mankind the ability to make things the way nature has been doing it for eons; atom by atom from the bottom up.*

Nanoscience vs. Nanotechnology

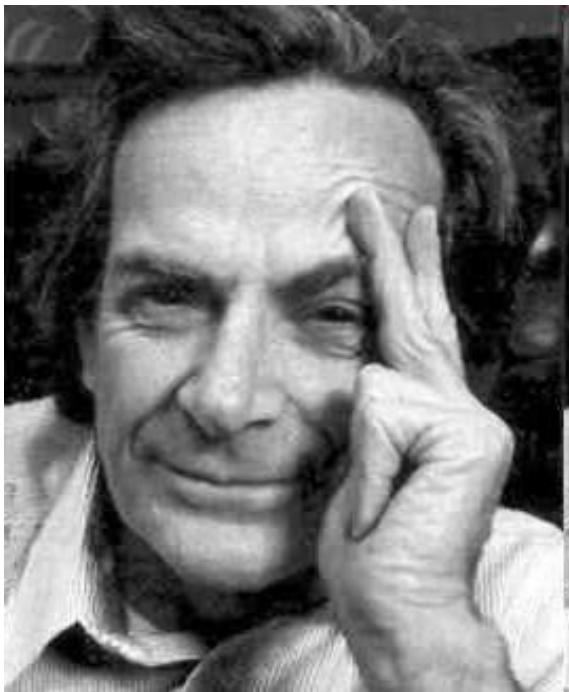
Nanoscience: exploring
and studying the
properties of the
nanoscale



Applying the unique
properties of the
nanoscale to technology



ORIGIN OF NANOSCIENCE



Richard P. Feynman

“There is a plenty of room at the bottom”

**(Lecture in 1959 at the annual meeting
of the American Physical Society)**

**“I would like to describe a field in which
little has been done but in which an
enormous amount can be done in
principle”**



**“Why cannot we write the entire 24
volumes of the Encyclopaedia Britannica
on the head of a pin ?”**



Professor Norio Taniguchi was the first person to use the term 'nanotechnology' in 1974.

"Nano-technology' mainly consists of the processing of, separation, consolidation, and deformation of materials by one atom or by one molecule."

Moore's Law

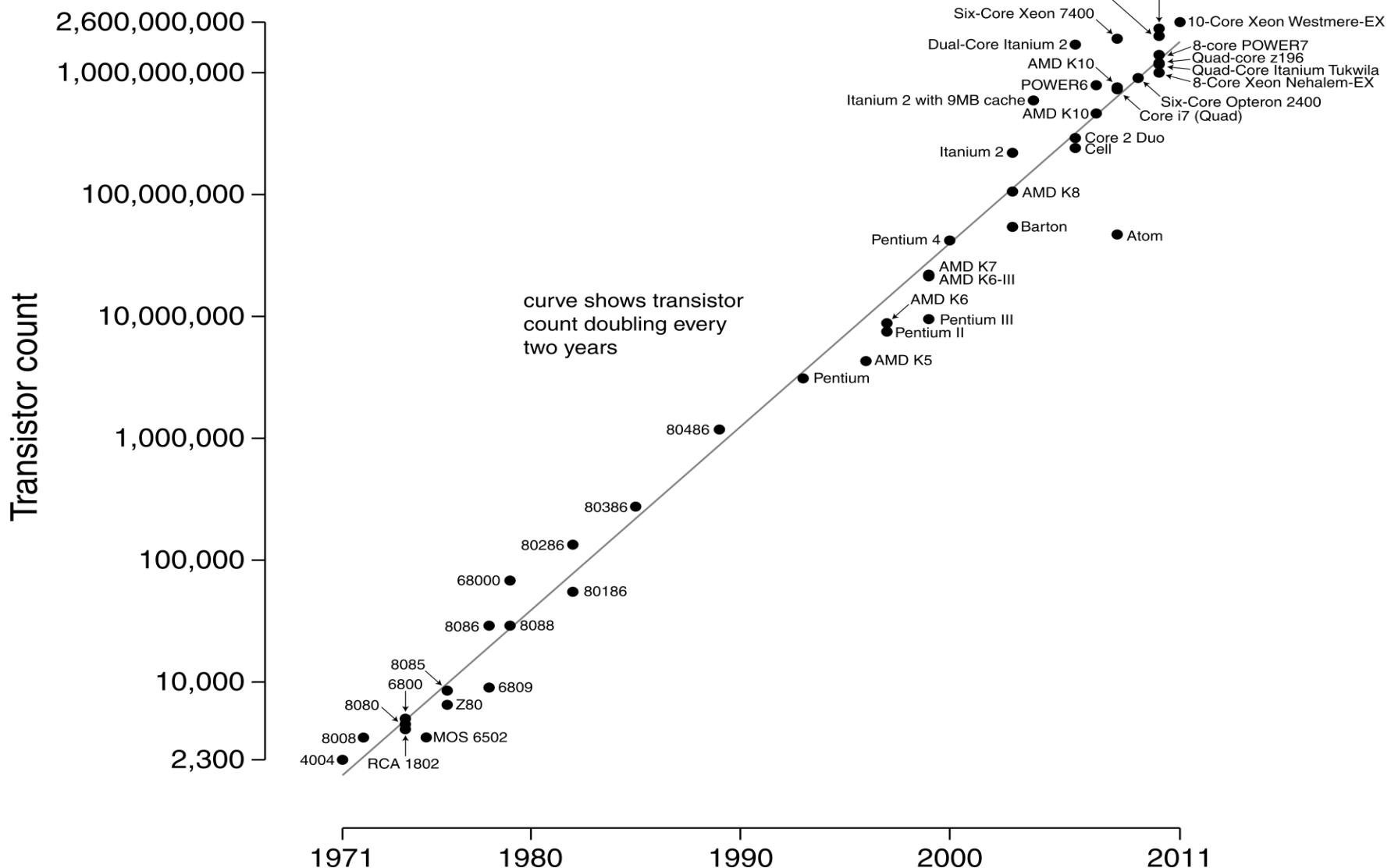
Gordon Moore, one of the founders of Intel, made the observation in 1965 that the number of transistors per circuit would double every year, through the decade following that year. Later he changed this to 24 months. There were only 30 transistors on an integrated circuit at that time.

We must, however, remember that Moore's law is neither a scientific law nor a law of nature. It is only a prophetic statement.

Moore's law has been the name given to everything that changes exponentially.

Microprocessor Transistor Counts 1971-2011 & Moore's Law

Shrinking Transistors to the Nanoscale



Nanotechnology is interdisciplinary and impacts many application

- Physics
- Chemistry
- Biology
- Materials Science
- Polymer Science
- Electrical Engineering
- Chemical Engineering
- Mechanical Engineering
- Medicine
- Electronics
- Materials
- Health/Biotech
- Chemical
- Environmental
- Energy
- Aerospace
- Automotive
- Security
- Forest products

INTRODUCTION TO NANOMATERIALS

Richard P. Feynman



The Lycurgus Cup, when illuminated from outside, it appears green and when illuminated from within the cup, it glows red (glass; British Museum; 4th century A.D. From the site, <http://www.thebritishmuseum.ac.uk>)

Nano Structure and Device

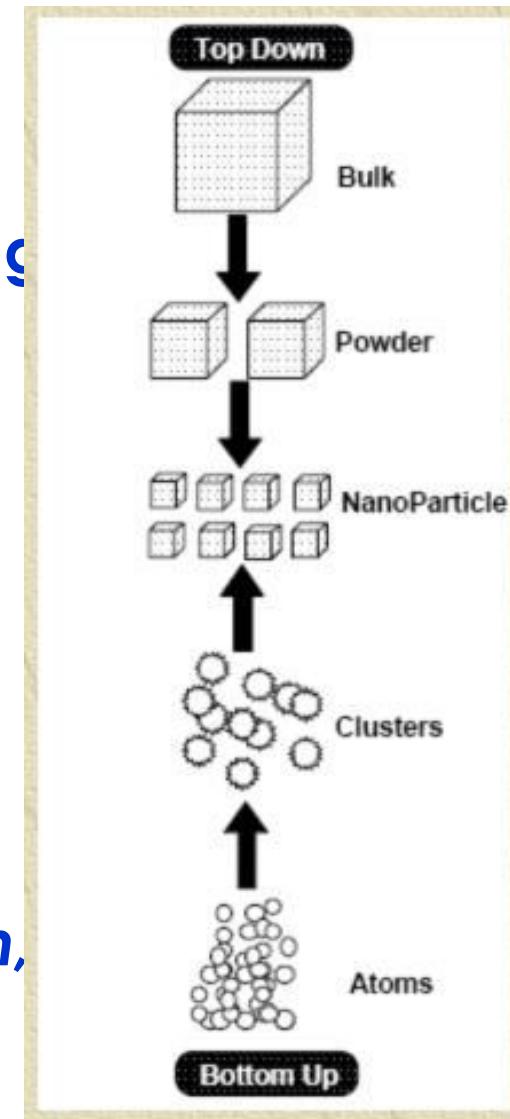
- Nano structure and device can be accomplished by two approaches.

-- “Bottom Up” method where small building blocks are produced and assembled into larger structures.

examples: chemical synthesis, laser trapping, self assembly, colloidal aggregation, etc

-- “Top Down” method where Large object are modified to give smaller features.

examples: film deposition and growth, nano imprint / lithography, etching technology, mechanical polishing



Different types of Nanomaterials

QUANTUM DOT (Zero-dimensional objects)

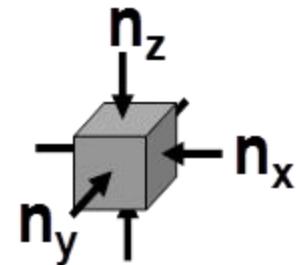
Nanoparticles or nanocrystals

Small nanoparticles are often called quantum dots

Nanodimensions \sim 1-50 nm in size

In this systems where the electrons are confined in their motion in all three directions

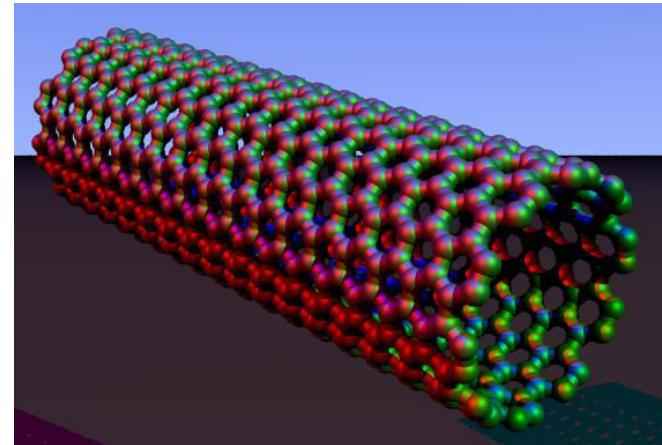
Metal, metal oxides, semiconducting and magnetic materials can be prepared



QUANTUM WIRE

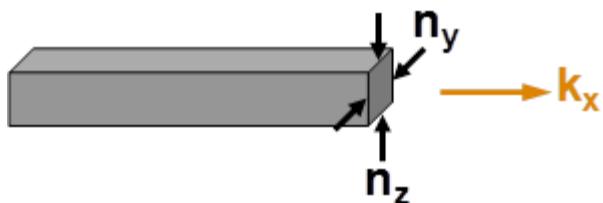
One dimensional materials

Long (several microns in length) but with diameters only a few nanometers

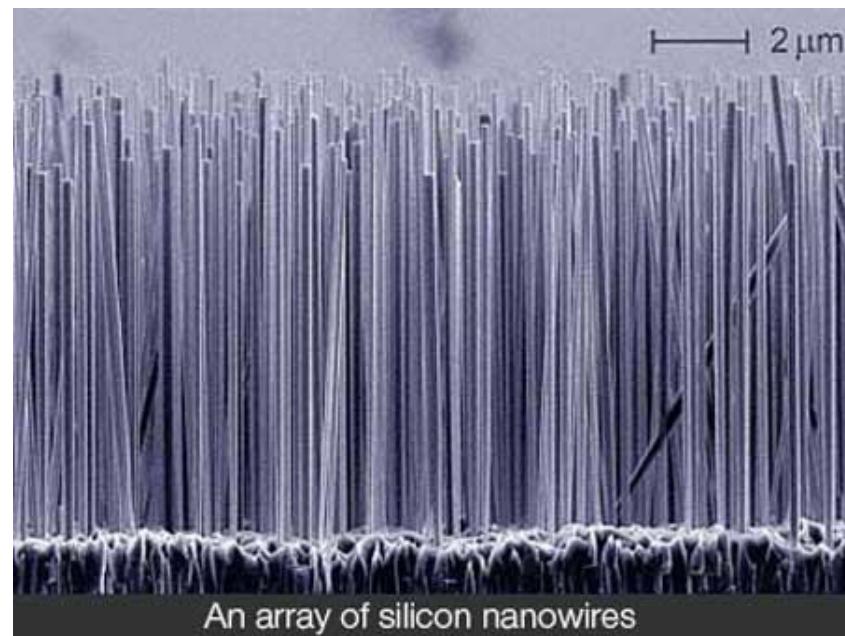


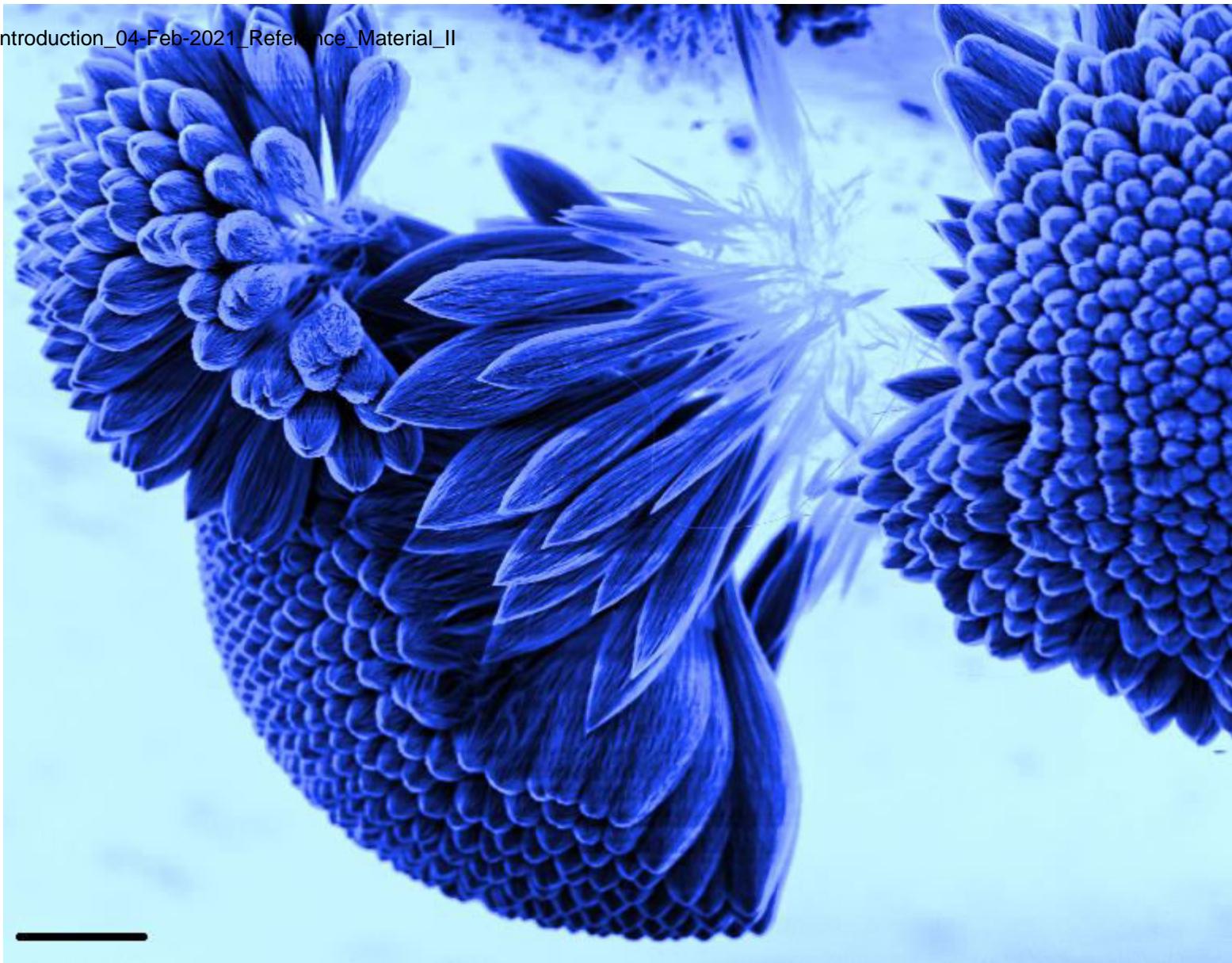
Nanowires and nanotubes belong to this category

Metals, oxides and other materials



In this system where electrons are free to move in one direction and confined in the other two directions.

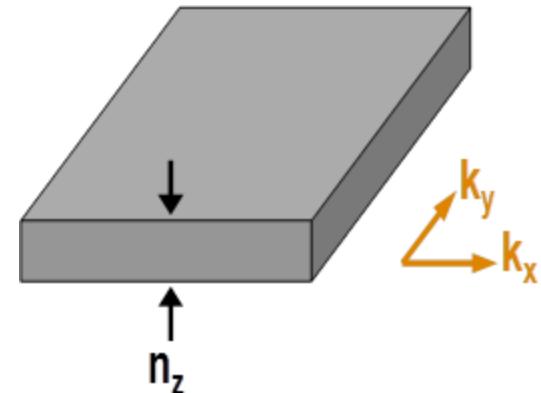




**Dr.K.JEGANATHAN, Associate Prof. and Coordinator
Center for NanoScience and NanoTechnology
Bharathidasan University, Trichy -24**

QUANTUM WELL (Two dimensional materials)

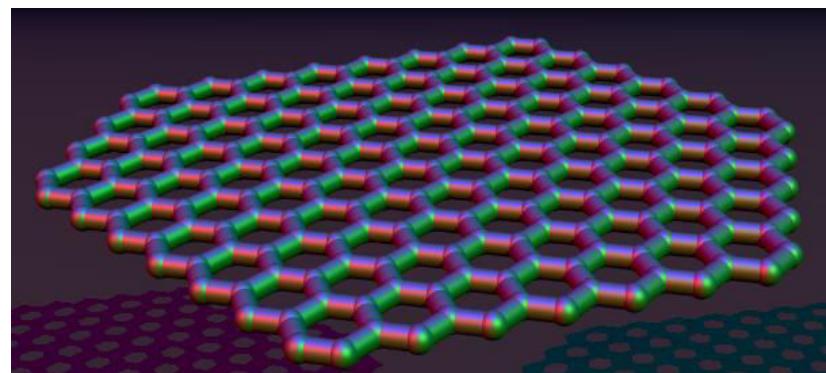
Area (several microns square) but with thickness only a few nanometers



Nanofilms and nanosheets or nanowall belong to this category

e.g. Graphite

In this type of systems the electron are confined in their motion in one direction while they move as in the corresponding bulk material in the other two directions.



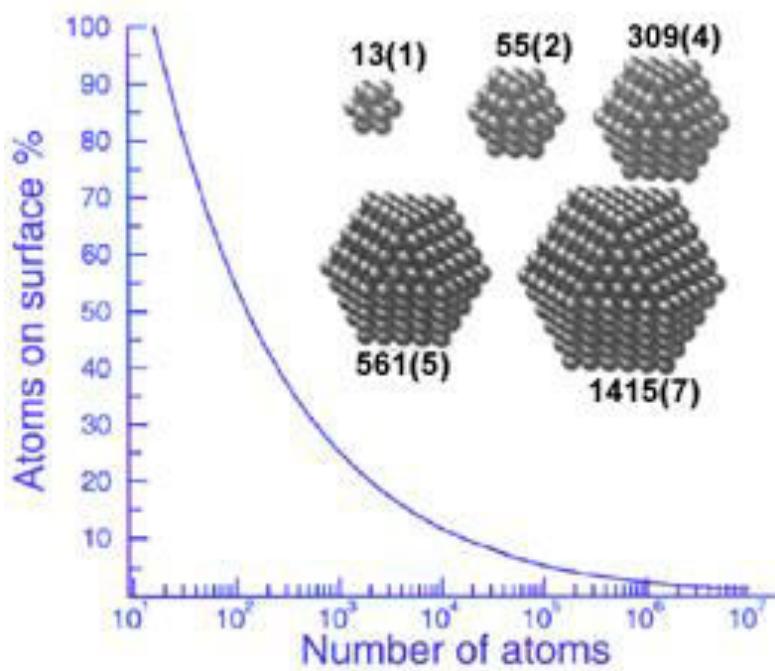
Physical properties of nanomaterials

Atoms on a small scale behave like nothing on a large scale

Main feature – size determines the properties of nanomaterials

In macro-scale – properties do not change much with size

In nano-scale – enormous changes



The Surface Area - if you go on reducing the size of a material to a very very small particle, say 1 nm then all the atoms constituting the particle will be on the surface.

-if you increase the size 1 nm to 5 nm, the number of atoms on the surface will decrease.

SURFACE AREA TO VOLUME RATIO

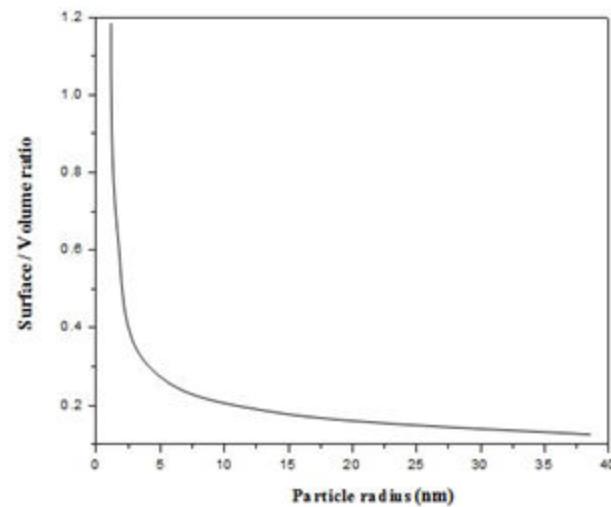
In a nanoparticle, the amount of surface area the particle has is larger compared to its volume.

This means there are more atoms on the surface of the particle than in the middle of it, and that makes them the most important.

Surface atoms act differently to atoms inside a particle, so when there are more surface atoms than inside atoms the way they behave dominates the whole behaviour of the particle.

The opposite is also true, when the particle is bigger it has a large volume compared to its surface area and the number of atoms inside the particle is much higher than the number of atoms on the outside (the surface) of the particle.

What the inside atoms are doing is the most important thing and the behaviour of the particle will be decided by them.



So what difference does it make?

How surface atoms and inside atoms behave can be very different.

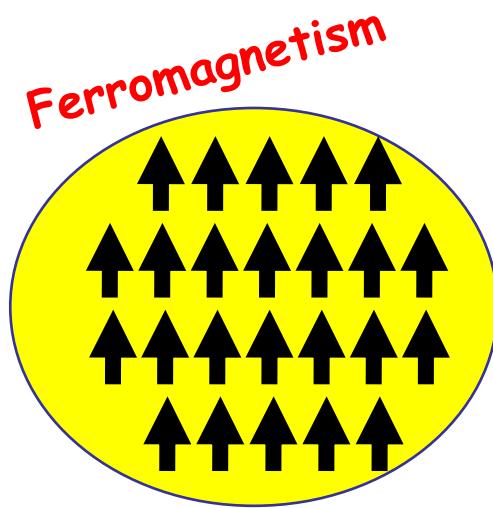
This means that when we get a very small piece of material, with comparatively large numbers of surface atoms, the material can act very differently to what we are used to (aluminium nanoparticles explode!).

In nanotechnology we are making use of particles with lots of surface atoms and the fact that this makes them behave differently - it allows us to do new and exciting things.

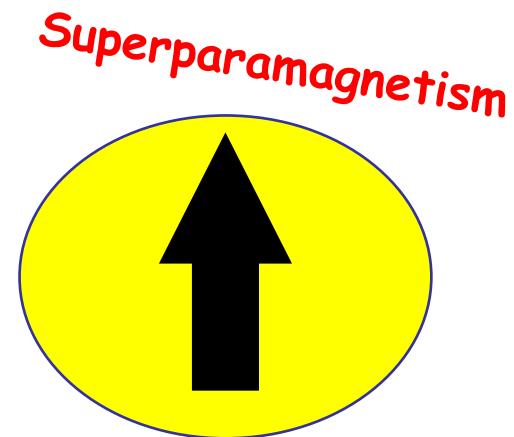
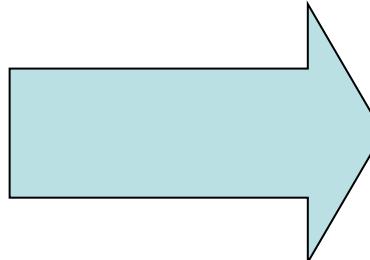
Changes in properties

Magnetic properties

Magnetic properties of nanostructured materials are distinctively different from that of bulk materials. Ferromagnetism of bulk materials disappears and transfers to superparamagnetism in the nanometer scale due to the huge surface energy.



N spins, each
of moment μ



'Superspin' of
moment $N\mu$

Optical properties of nanomaterials can be significantly different from bulk crystals. E.g. The optical absorption peak of a semiconductor nanoparticle shifts to short wavelength, due to an increased band gap. The colour of metallic nanoparticles may change with their sizes due to surface plasmon resonance.



Bulk gold

shine as a metal;

Chemically not reactive
(make jewel)

small particle of gold

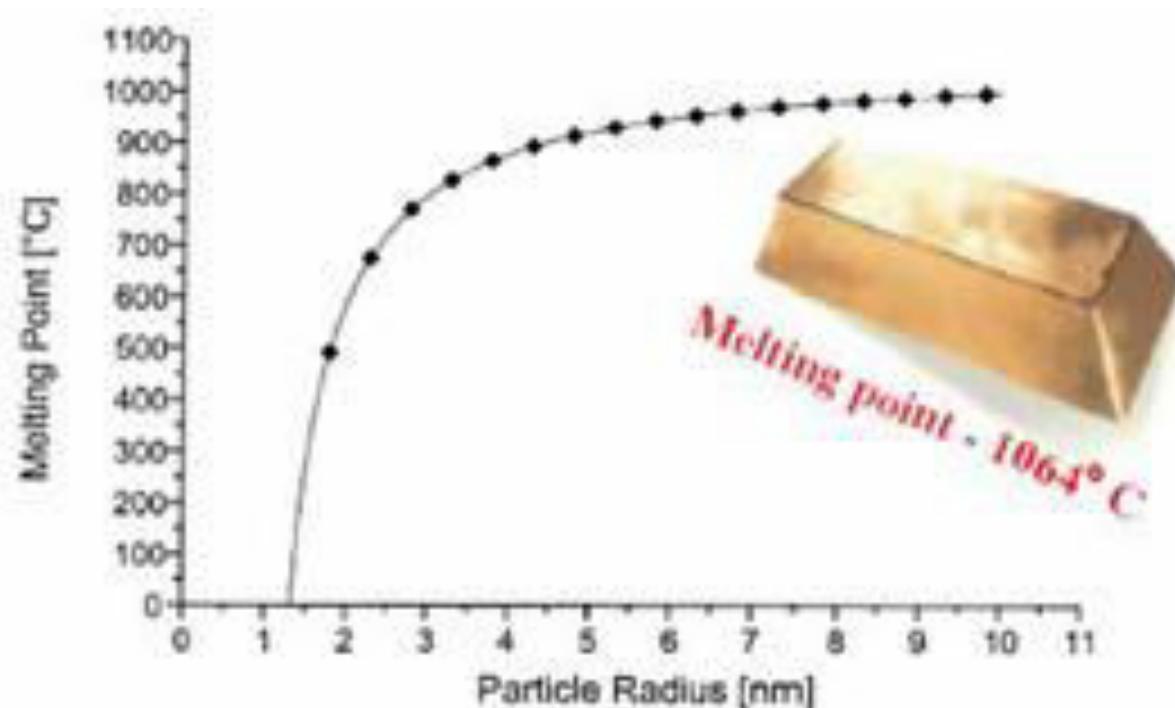
no metallic, don't shine

Reactive

Thermal properties

12 Nanomaterials 04 Feb 2021 - Ferreira - Materials

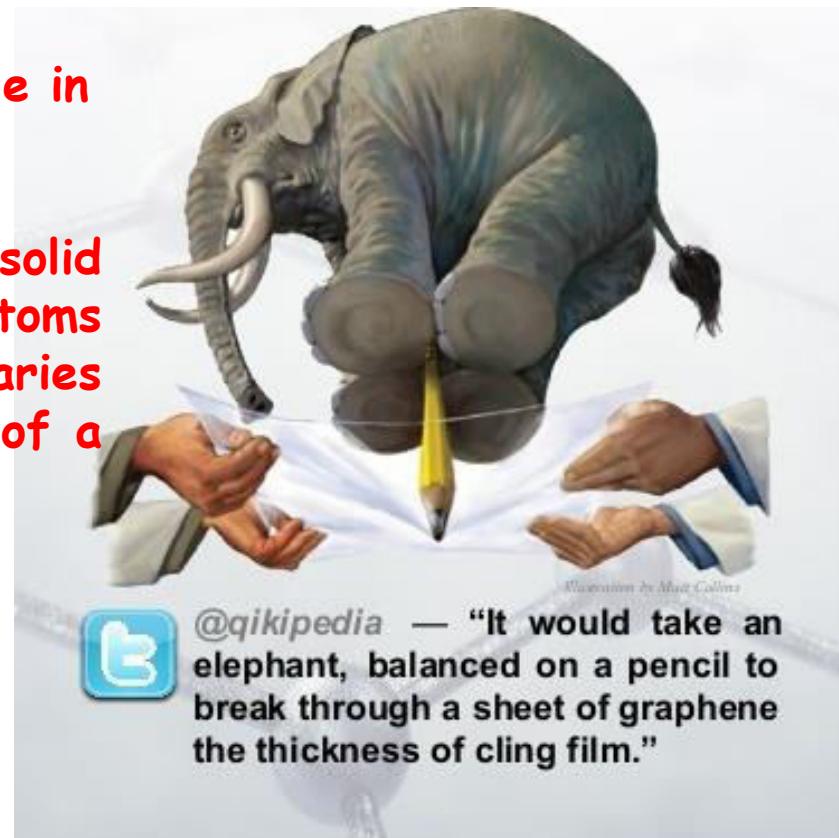
Nanomaterials may have a significantly lower melting point or phase transition temperature and appreciably reduced lattice constants, due to a huge fraction of surface atoms in the total amount of atoms



Mechanical properties of nanomaterials may reach the theoretical strength, which are **one or two** orders of magnitude **higher than** that of single crystals in the bulk form. The enhancement in mechanical strength is due to the **reduced probability of defects**.

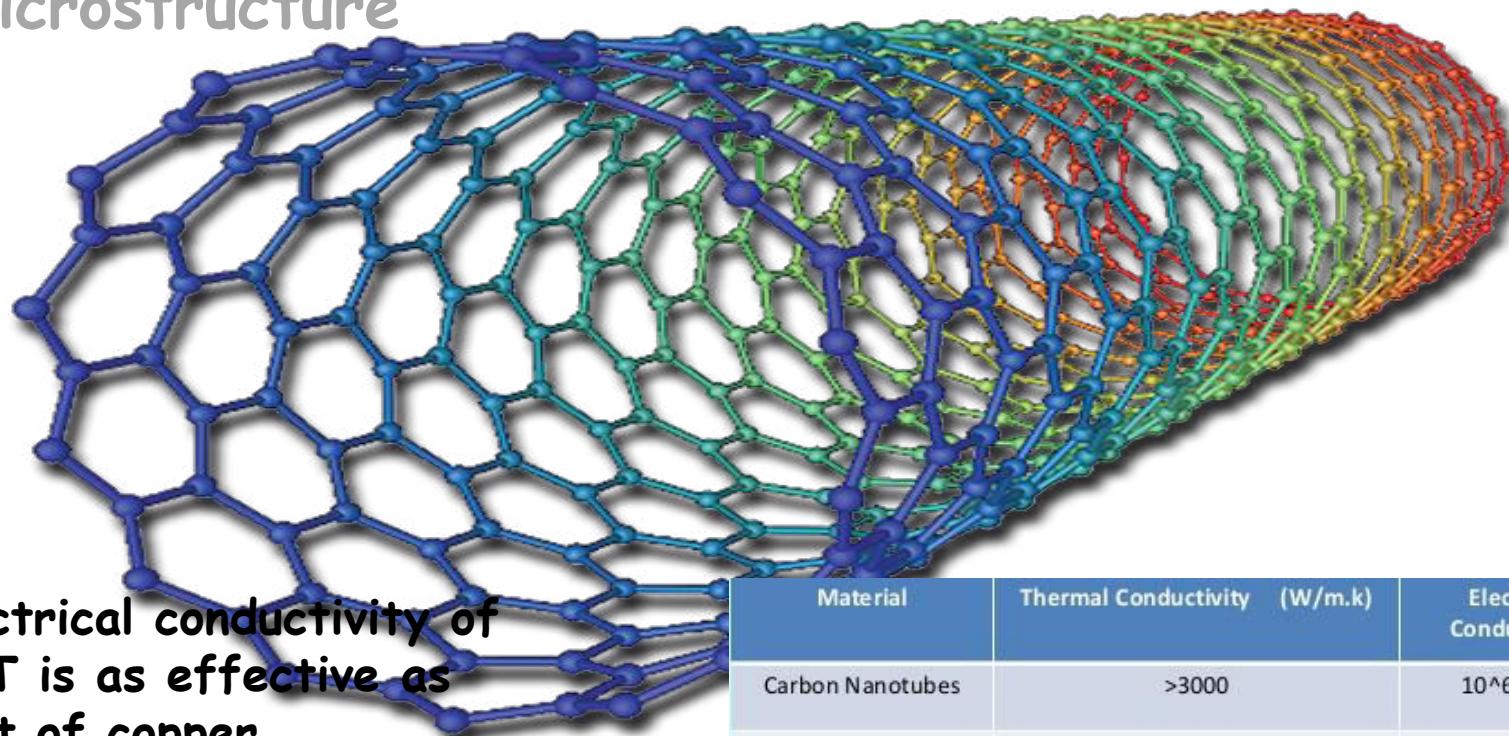
Grain boundaries play a significant role in the materials properties

As the grain size d of the solid decreases, the proportion of atoms located at or near grain boundaries relative to those within the interior of a crystalline grain, scales as $1/d$.



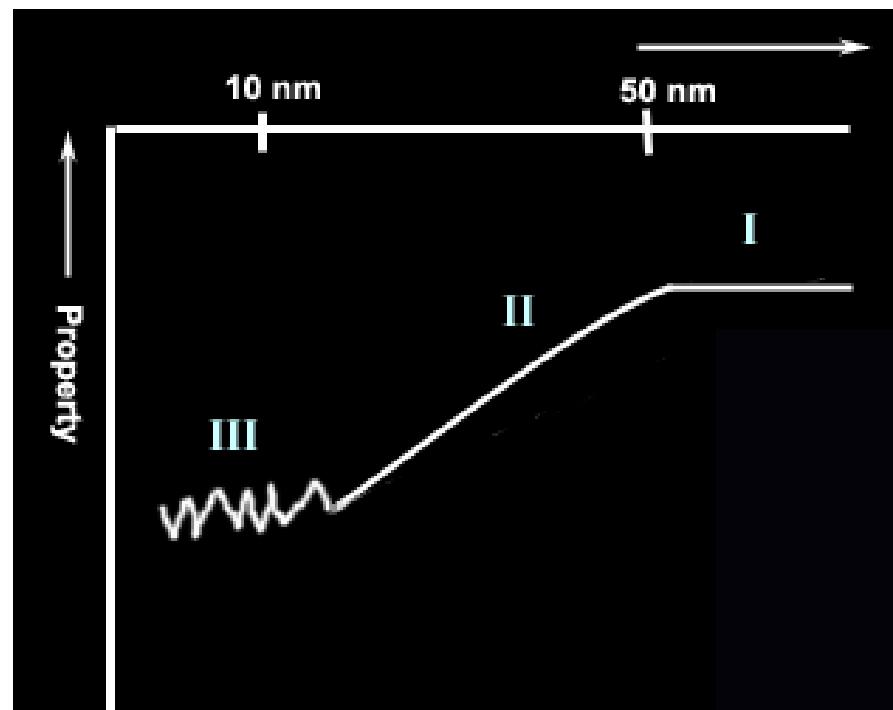
@qikipedia — “It would take an elephant, balanced on a pencil to break through a sheet of graphene the thickness of cling film.”

Electrical conductivity decreases with a reduced dimension due to increased surface scattering. However, electrical conductivity of nanomaterials could also be enhanced appreciably, due to the better ordering in microstructure



Very high current carrying capacity

Material	Thermal Conductivity (W/m.k)	Electrical Conductivity
Carbon Nanotubes	>3000	$10^6\text{-}10^7$
Copper	400	$6\text{*}10^7$
Carbon Fiber – Pitch	1000	$2\text{-}8.5\text{*}10^6$
Carbon Fiber – PAN	8-105	$6.5\text{-}14\text{*}10^6$



In region I (50 nm and above) ,
properties will be similar to those the bulk

II (10 – 50 nm),
their properties vary linearly with size

III (very small),
we get some unusual and new properties

These are due to quantum effects

Any heat treatment increases the diffusion of impurities, intrinsic structural defects and dislocations, and one can easily push them to the nearby surface. chemical stability would be enhanced.

Bulk gold does not exhibit catalytic properties
Au nanocrystal is an excellent low temperature catalyst.

- Therefore, if we can control the processes that make a nanoscopic material, then we can control the material's properties.

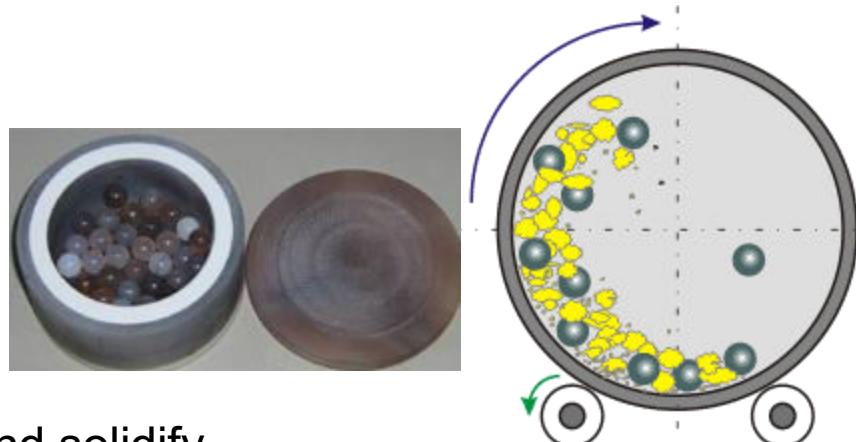
Characteristics of nanomaterials that distinguish them from bulk materials

- large fraction of surface atoms
- high surface energy
- spatial confinement
- reduced imperfections

Methods of preparing nanomaterials

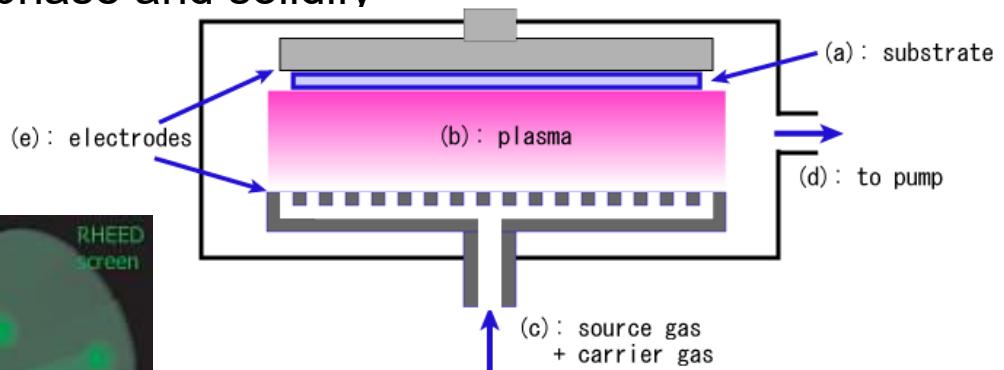
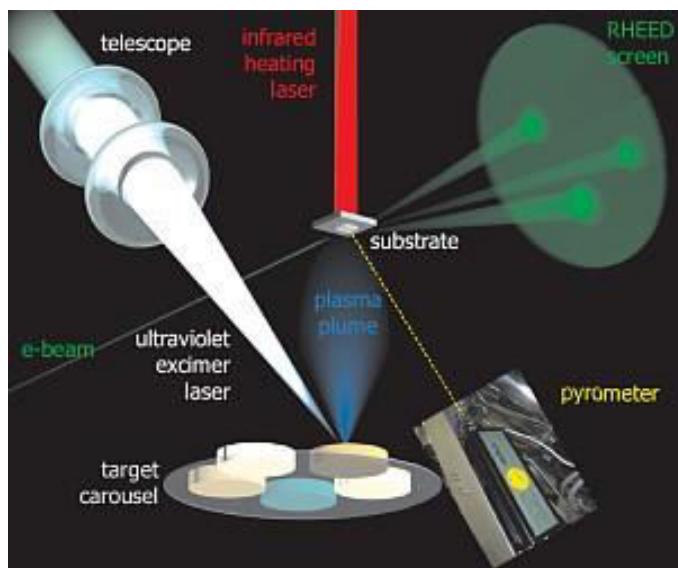
Physical and Chemical methods

Physical Method -
Grind materials Using ball milling



Evaporate materials to gas phase and solidify

Laser

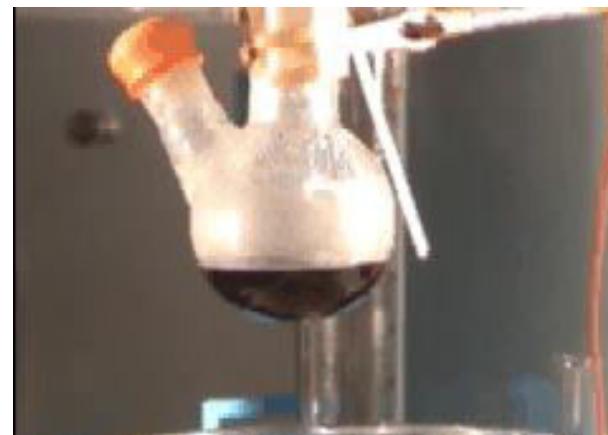


Chemical methods (most powerful)

React a metal salt with a alcohol or
some other reducing agent

Metal compound along with a reagent
in a boiling solvent (sealed vessel)

Under this condition,
many kinds of nano-particles are formed



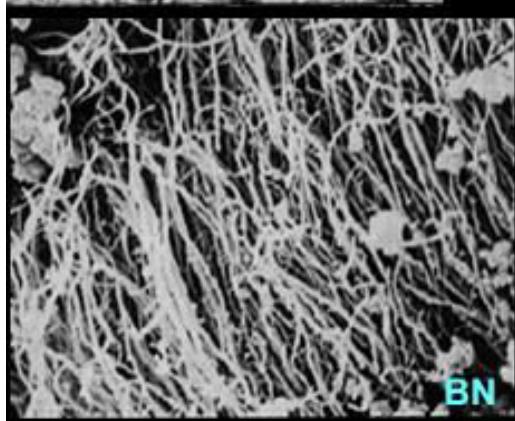
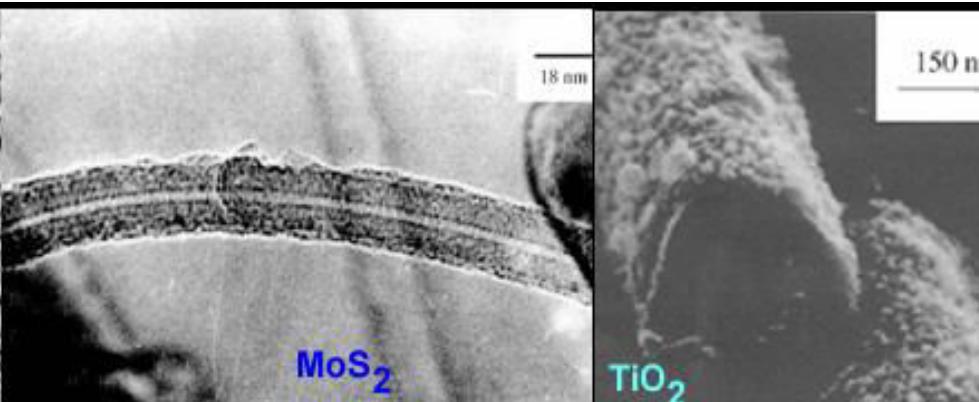
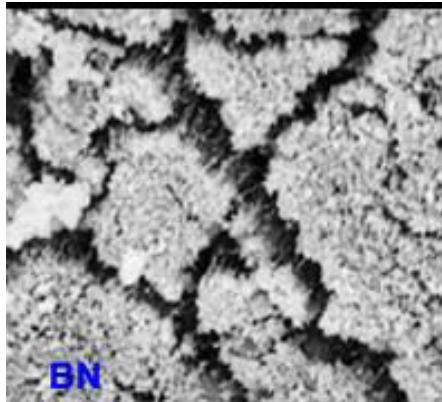
Hydrothermal and solvothermal methods

Water used as a boiling solvent - Hydrothermal

Organic boiling solvent like a hydrocarbon - solvothermal

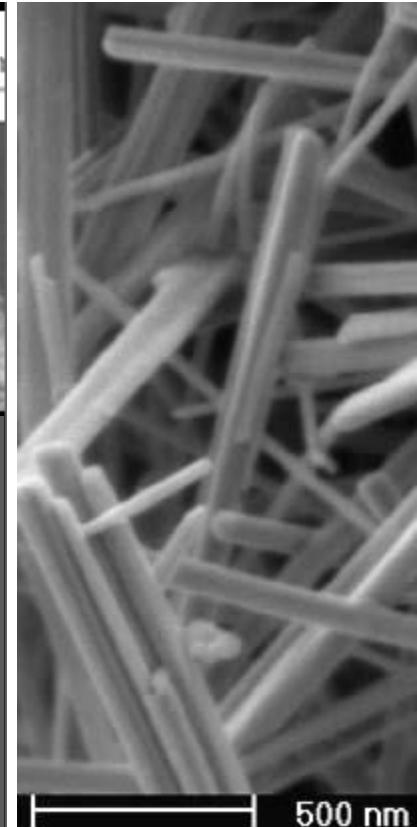
For example, if we take a metal acetate and heat in a boiling hydrocarbon, we get metal or metal oxide nanoparticles

Today we have reached a level where can make nanomaterial of any compound in any shape we desired



Carbon nanotubes have been known since 1991.

Now it is possible to make nanotubes of metal sulfides, metal oxides and other materials



SEM image of WS₂ nanotubes

Nanotubes, therefore, need not be of carbon alone

Physical properties of nanomaterials

Atoms on a small scale behave like nothing on a large scale

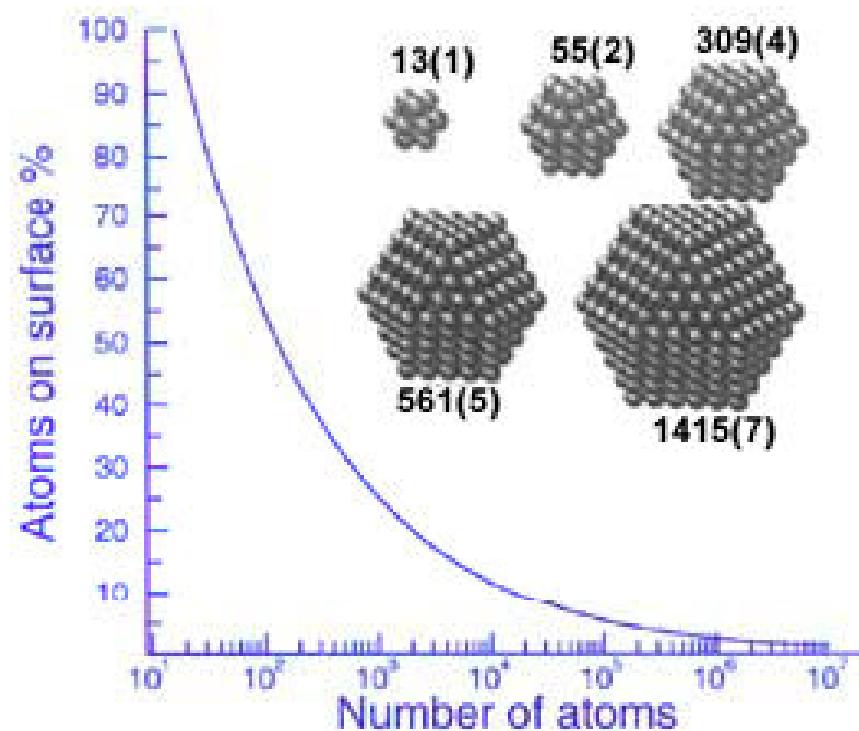
Main feature – size determines the properties of nanomaterials

In macro-scale – properties do not change much with size

In nano-scale – enormous changes

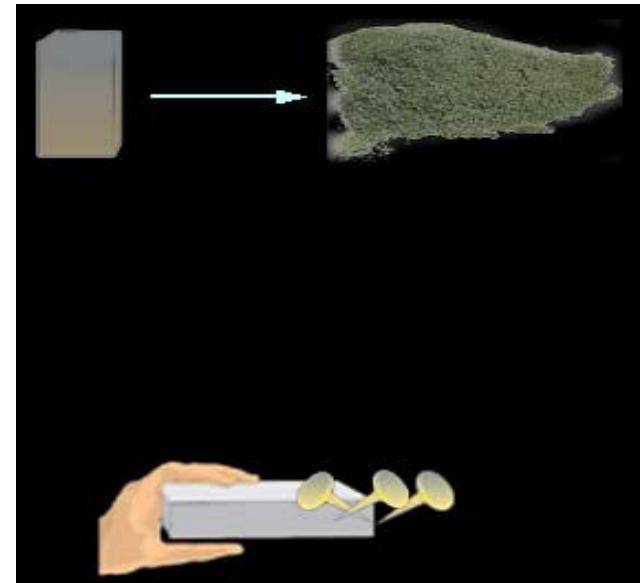
The Surface Area - if you go on reducing the size of a material to a very very small particle, say 1 nm then all the atoms constituting the particle will be on the surface.

-if you increase the size 1 nm to 5 nm, the number of atoms on the surface will decrease.



Changes in properties

1. Magnetic properties of nanostructured materials are distinctively different from that of bulk materials. Ferromagnetism of bulk materials disappears and transfers to superparamagnetism in the nanometer scale due to the huge surface energy.



Bulk gold

shine as a metal;

Chemically not reactive
(make jewel)

2. Optical properties of nanomaterials can be significantly different from bulk crystals. E.g. The optical absorption peak of a semiconductor nanoparticle shifts to short wavelength, due to an increased band gap. The colour of metallic nanoparticles may change with their sizes due to surface plasmon resonance.

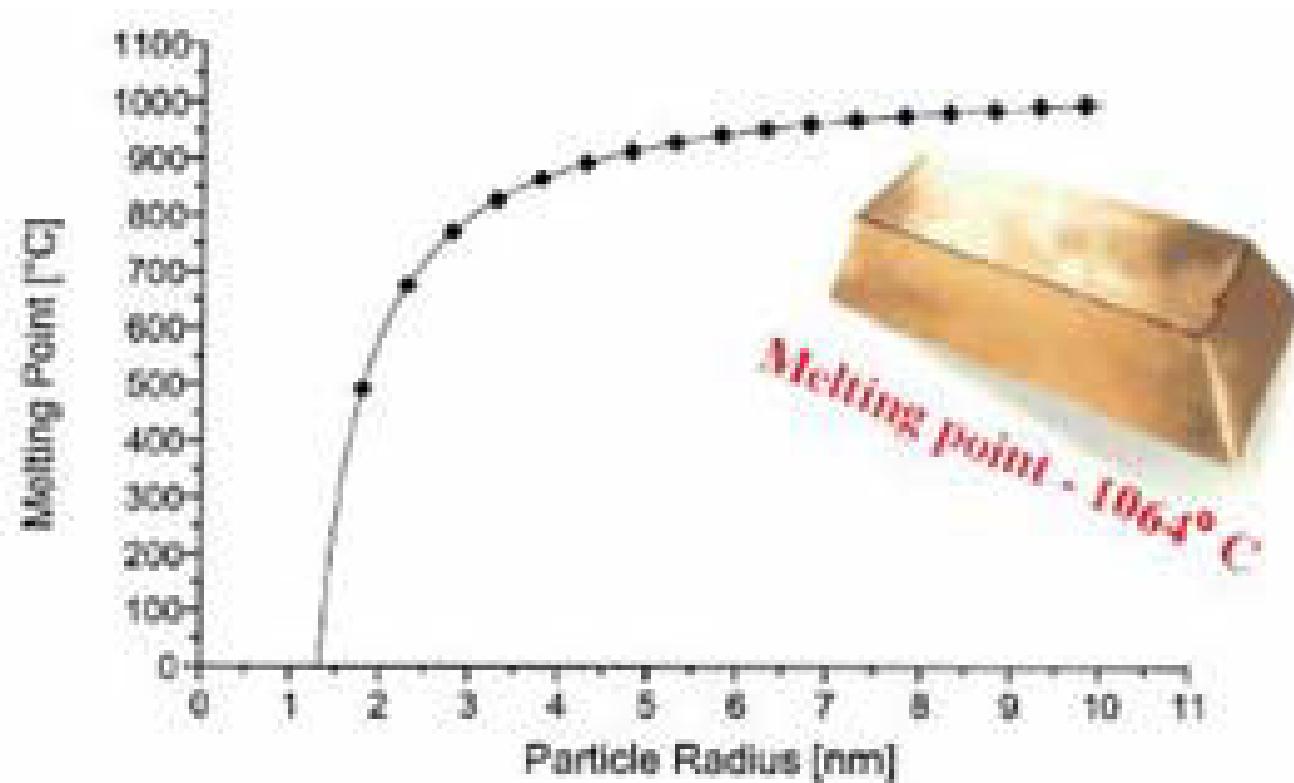
small particle of gold
no metallic, don't shine

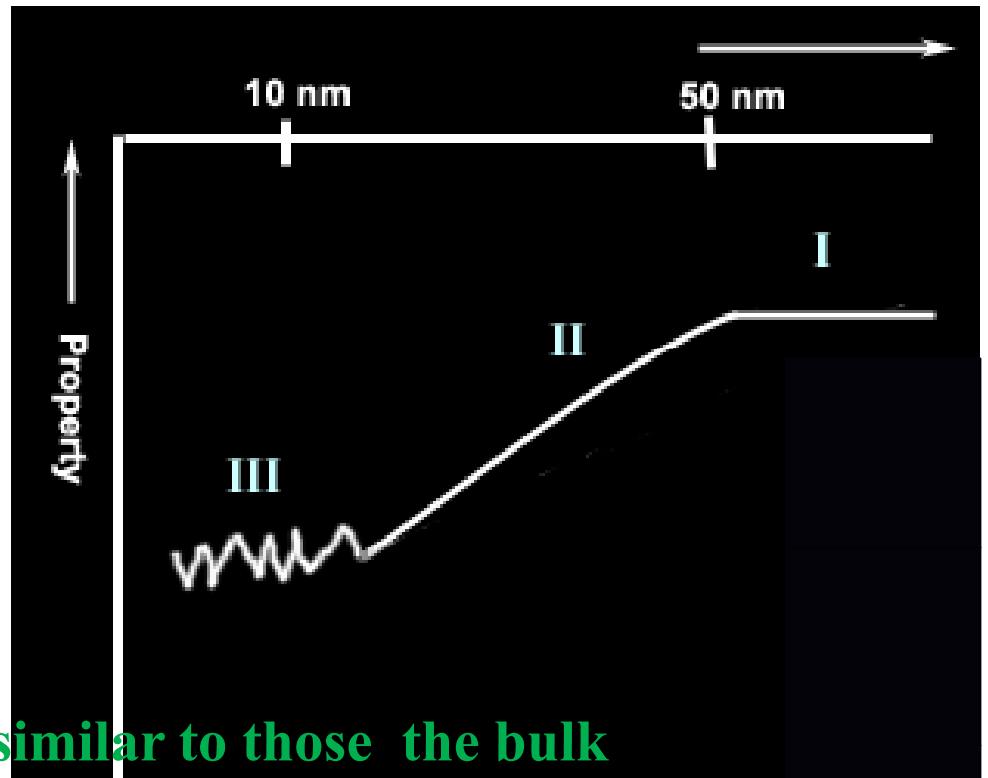
Reactive



Changes in properties

3. Thermal Properties of Nanomaterials may have a significantly lower melting point or phase transition temperature and appreciably reduced lattice constants, due to a huge fraction of surface atoms in the total amount of atoms





In region I (50 nm and above) ,
properties will be similar to those the bulk

II (10 – 50 nm),
their properties vary linearly with size

III (very small),
we get some unusual and new properties

These are due to quantum effects

4.Mechanical properties of nanomaterials may reach the theoretical strength, which are **one or two** orders of magnitude **higher than** that of single crystals in the bulk form. The enhancement in mechanical strength is due to the **reduced probability of defects.**

5.Electrical conductivity decreases with a reduced dimension due to increased surface scattering. However, electrical conductivity of nanomaterials could also be enhanced appreciably, due to the better ordering in microstructure, e.g. polymeric fibrils.

Any heat treatment increases the diffusion of impurities, intrinsic structural defects and dislocations, and one can easily push them to the nearby surface. chemical stability would be enhanced.

Bulk gold does not exhibit catalytic properties
Au nanocrystal is an excellent low temperature catalyst.

- Therefore, if we can control the processes that make a nanoscopic material, then we can control the material's properties.

Characteristics of nanomaterials that distinguish them from bulk materials

- large fraction of surface atoms
- high surface energy
- spatial confinement
- reduced imperfections

PHYSICS FOR ENGINEERS

Course Code: PHY1701 Engineering Physics

Lecture : 3 per week L T P J C : 3 0 2 0 4

Theory

Mod 1: Introduction to Modern Physics (6)

Mod 2: Applications of Quantum Physics (5)

Mod 3: Nanophysics (5)

Mod 4: Laser Principles and Engineering Applications (6)

Mod 5: Electromagnetic Theory and Applications (6)

Mod 6: Propagation of EM waves in Optical fibers (6)

Mod 7: Optoelectronic Devices & Applications of Optical fibers (6)

Mod 8: Special Theory of Relativity (5)



VIT[®]
UNIVERSITY
(Estd. u/s 3 of UGC Act 1956)

Module 3: Nanophysics

1. Introduction to Nano-materials.
2. Moore's law.
3. Properties of Nano-materials.
4. Quantum confinement.
5. Quantum well, wire & dot.
6. Carbon Nano-tubes (CNT).
7. Applications of nanotechnology in industry.

Module 3: Nanophysics

1. Introduction to Nano-materials.
2. Moore's law.
3. Properties of Nano-materials.
4. Quantum confinement.
5. Quantum well, wire & dot.
6. Carbon Nano-tubes (CNT).
7. Applications of nanotechnology in industry.

BOOKS:

The Essential Understandings of Nanoscience and
Nanotechnology, J. Pradeep, Tata McGraw Hill, 2007

NANO

Nano

A prefix meaning 10^{-9} or one billionth

Nanometer

- One billionth (10^{-9}) of a meter
- Hydrogen atom 0.04 nm
- Proteins ~ 1-20 nm
- Feature size of computer chips
90 nm (in 2005)
- Diameter of human hair ~ 10
 μm

The Scale of Things – Nanometers and More



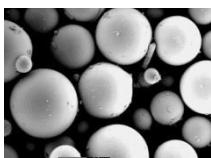
Dust mite
200 μm



Ant
~5 mm

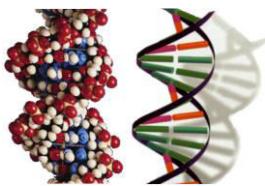
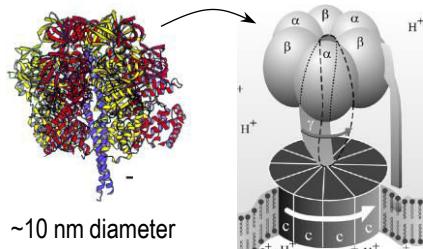
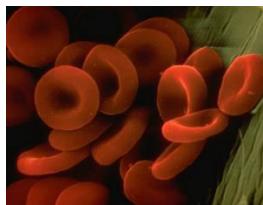


Human hair
~60-120 μm wide

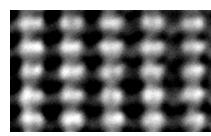


Fly ash
~10-20 μm

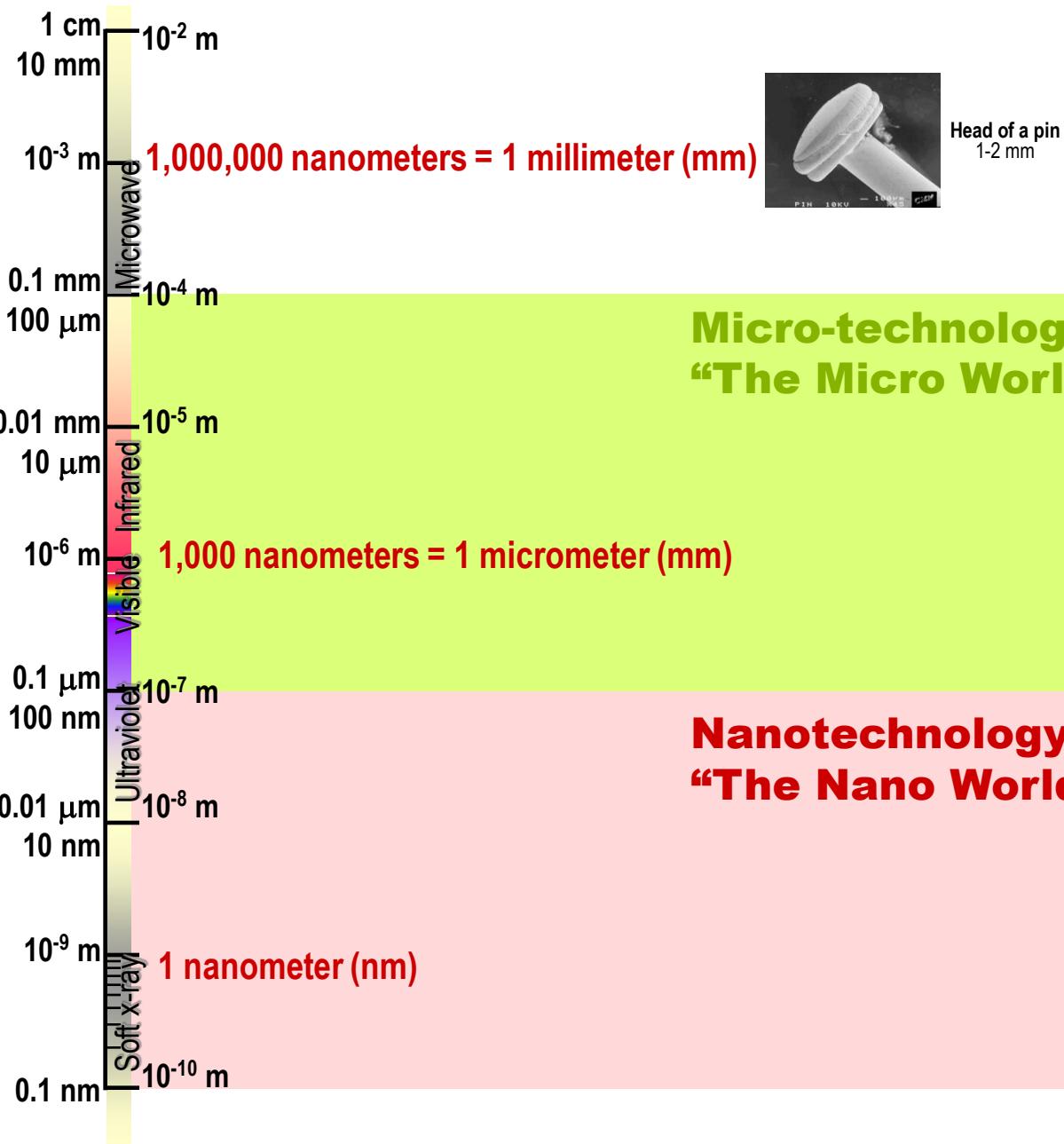
Red blood cells
(~7-8 μm)



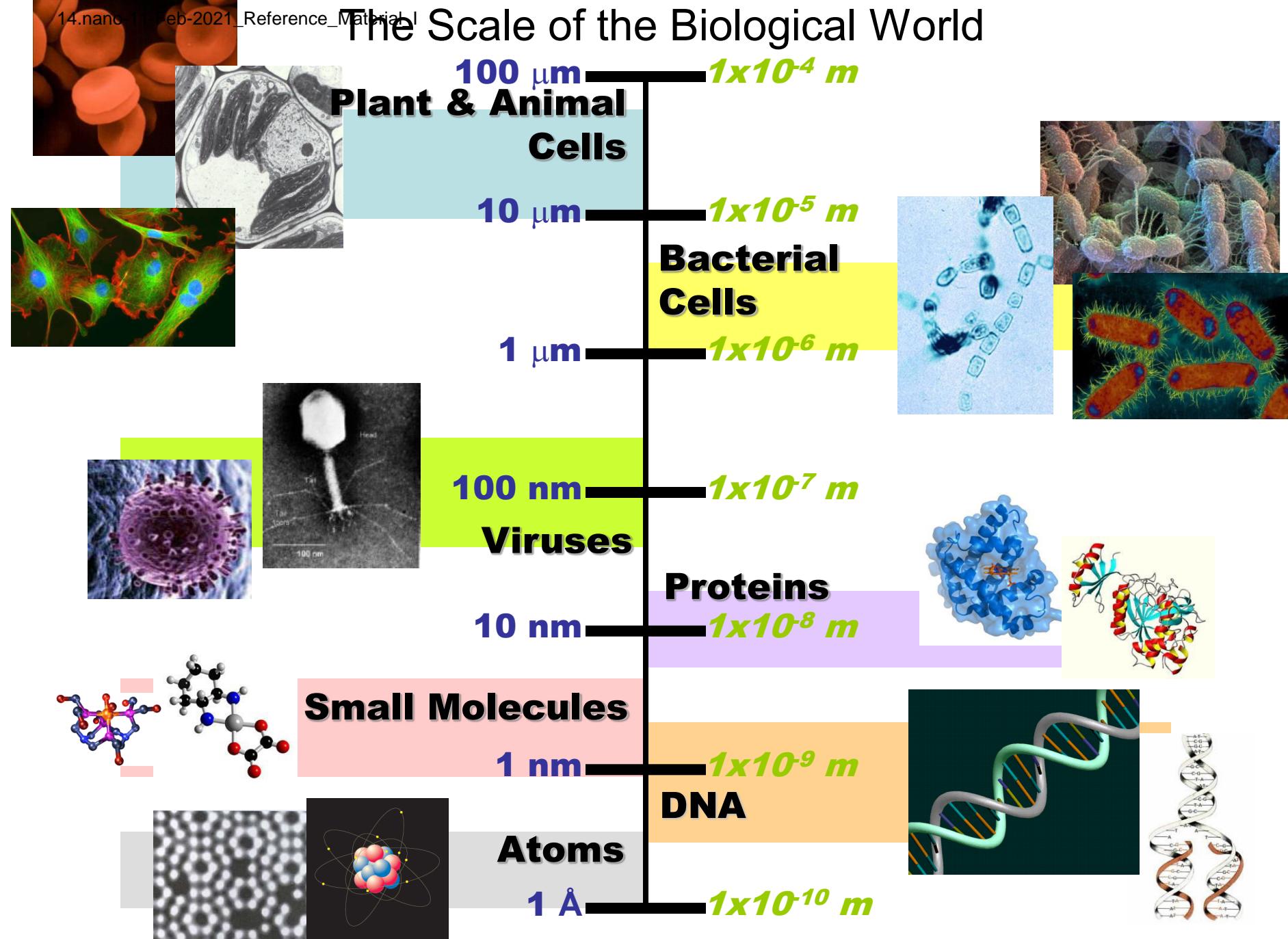
DNA
~2-1/2 nm diameter



Atoms of silicon
spacing 0.078 nm



The Scale of the Biological World



- Examples
 - Carbon Nanotubes
 - Proteins, DNA
 - Single electron transistors
- Not just size reduction but phenomena intrinsic to nanoscale
 - Size confinement
 - Dominance of interfacial phenomena
 - Quantum mechanics
- New behavior at nanoscale is not necessarily predictable from what we know at macroscales.

- Quantum size effects result in unique mechanical, electronic, photonic, and magnetic properties of nanoscale materials
- Chemical reactivity of nanoscale materials greatly different from more macroscopic form, e.g., gold
- Vastly increased surface area per unit mass, e.g., upwards of 1000 m^2 per gram
- New chemical forms of common chemical elements, e.g., fullerenes, nanotubes of carbon, titanium oxide, zinc oxide, other layered compounds

Atoms and molecules are generally less than a nm and we study them in chemistry. Condensed matter physics deals with solids with infinite array of bound atoms. Nanoscience deals with the in-between meso-world

- Quantum chemistry does not apply (although fundamental laws hold) and the systems are not large enough for classical laws of physics
- Size-dependent properties
- Surface to volume ratio
 - A 3 nm iron particle has 50% atoms on the surface
 - A 10 nm particle 20% on the surface
 - A 30 nm particle only 5% on the surface

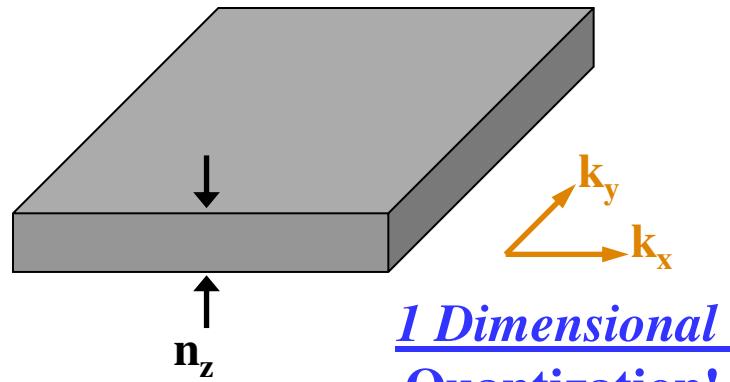
- 1. What novel quantum properties will be enabled by nanostructures (at room temp.)?**
- 2. How different from bulk behavior?**
- 3. What are the surface reconstructions and rearrangements of atoms in nanocrystals?**
- 4. Can carbon nanotubes of specified length and helicity be synthesized as pure species? Heterojunctions in 1-D?**
- 5. What new insights can we gain about polymer, biological...systems from the capability to examine single-molecule properties?**
- 6. How can one use parallel self-assembly techniques to control relative arrangements of nanoscale components according to predesigned sequence?**
- 7. Are there processes leading to economic preparation of nanostructures with control of size, shape... for applications?**

Quantum Confinement in Nanostructures: Overview

Electrons Confined in 1 Direction:

Quantum Wells (thin films):

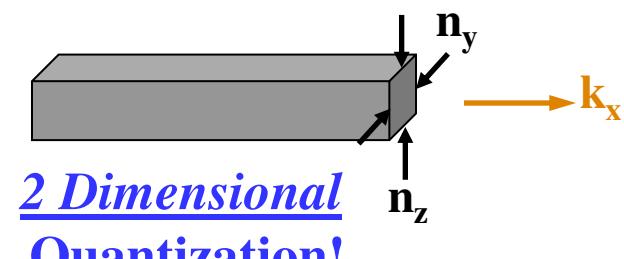
⇒ Electrons can easily move in
2 Dimensions!



Electrons Confined in 2 Directions:

Quantum Wires:

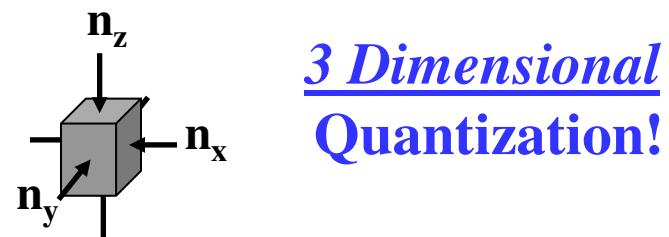
⇒ Electrons can easily move in
1 Dimension!



Electrons Confined in 3 Directions:

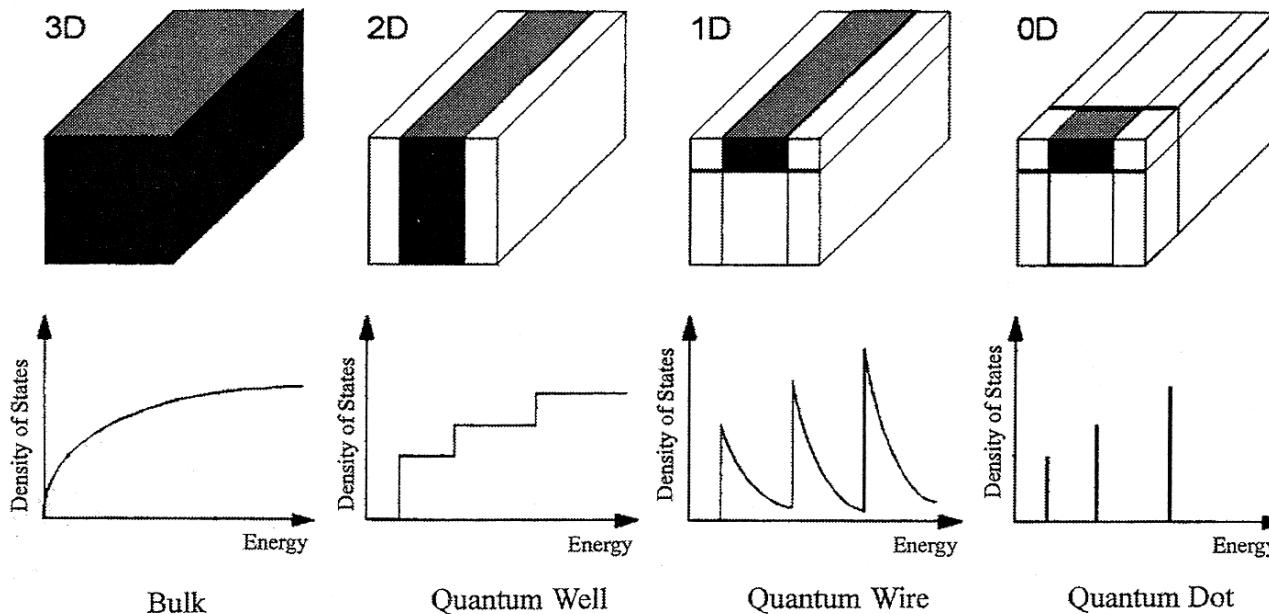
Quantum Dots:

⇒ Electrons can easily move in
0 Dimensions!



Each further confinement direction changes a continuous k component to a discrete component characterized by a quantum number n .

3D → **2D** → **1D** → **0D**

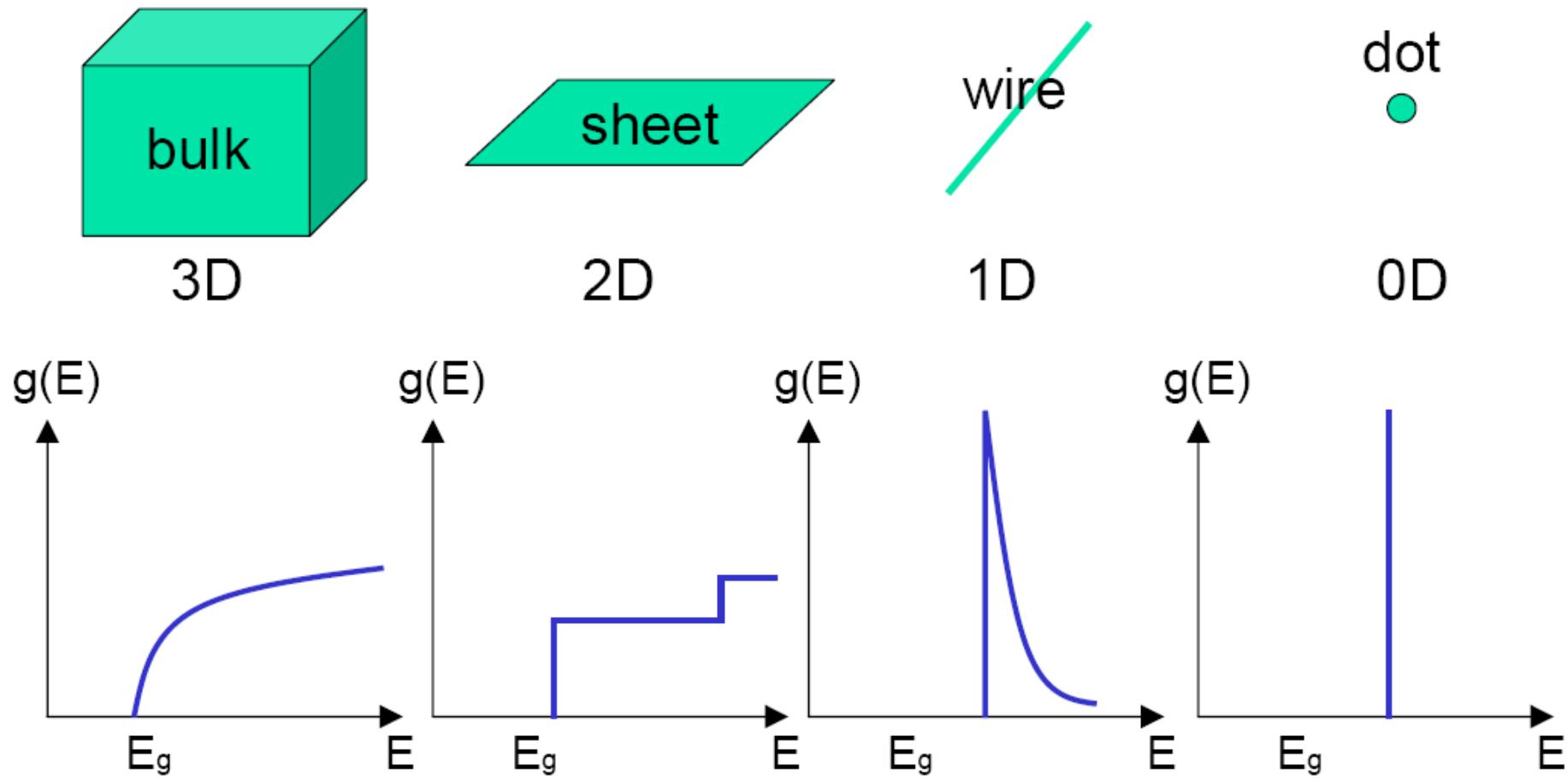


Source: Nanoscale Materials in Chemistry, Wiley, 2001

- If a bulk metal is made thinner and thinner, until the electrons can move only in two dimensions (instead of 3), then it is “2D quantum confinement.”
- Next level is ‘quantum wire’
- Ultimately ‘quantum dot’

Quantum confinement review

$g(E)$ = Density of states



Some Basic Physics

- **Density of states (DoS)**

$$DoS = \frac{dN}{dE} = \frac{dN}{dk} \frac{dk}{dE}$$

in 3D:

$$\begin{aligned} N(k) &= \frac{\text{k space vol}}{\text{vol per state}} \\ &= \frac{4/3 \pi k^3}{(2\pi)^3 / V} \end{aligned}$$

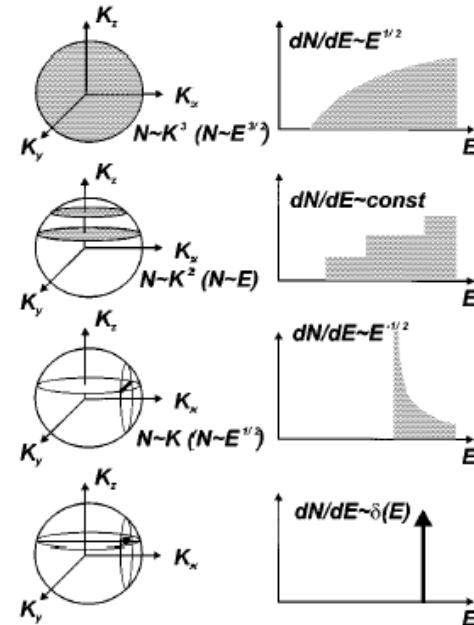
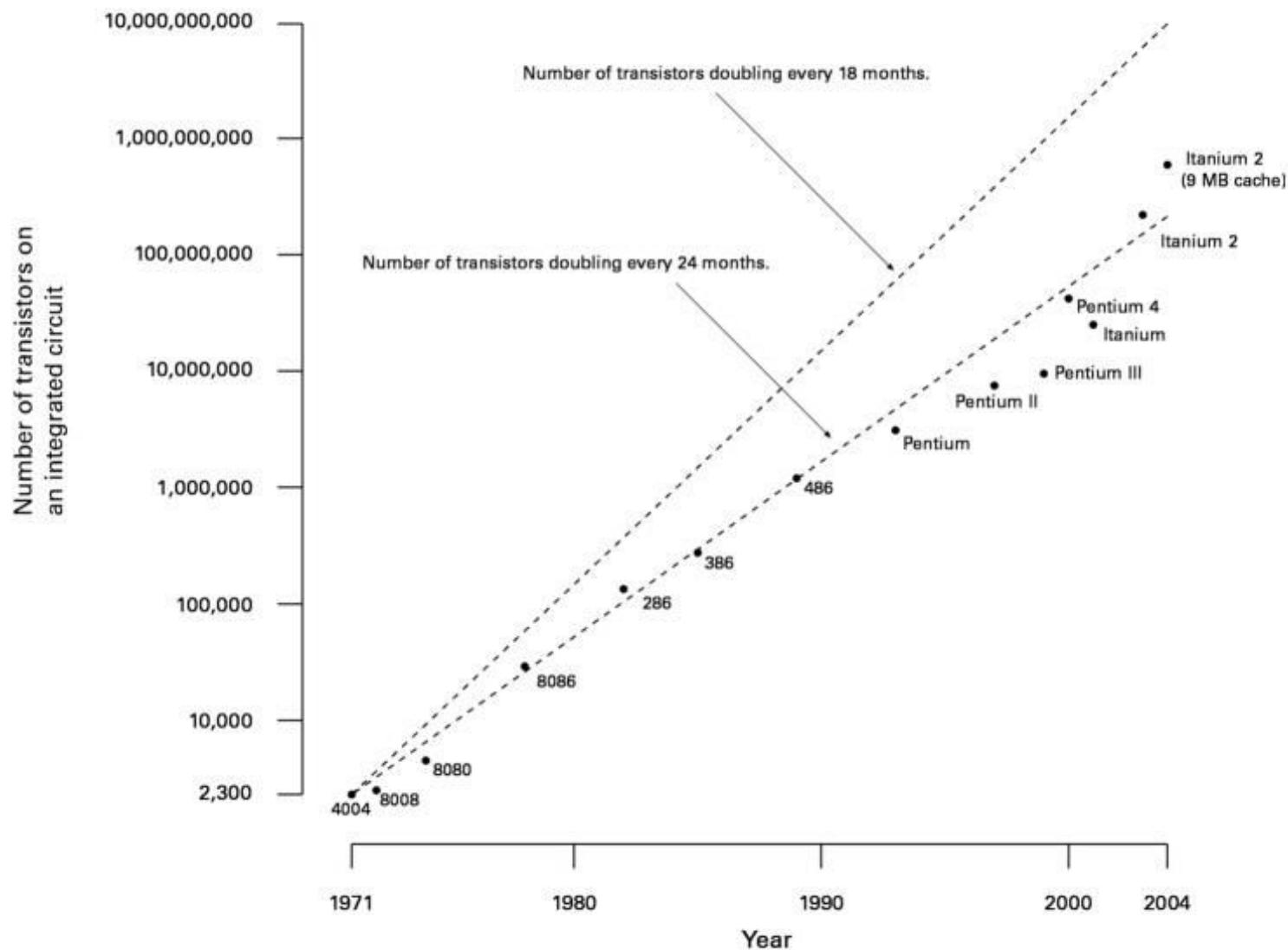


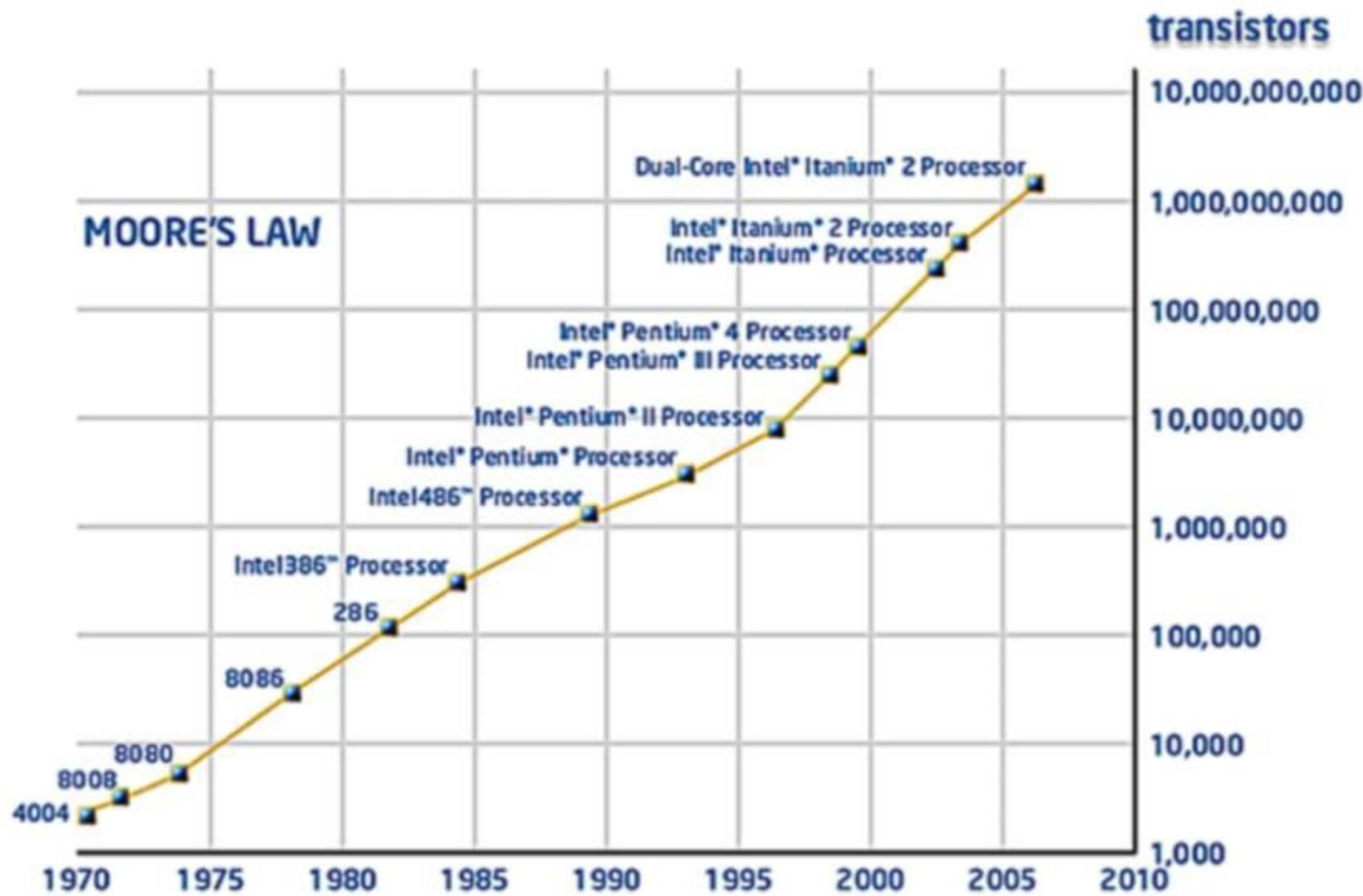
Fig. 1. Density of states for charge carriers in structures with different dimensionalities.

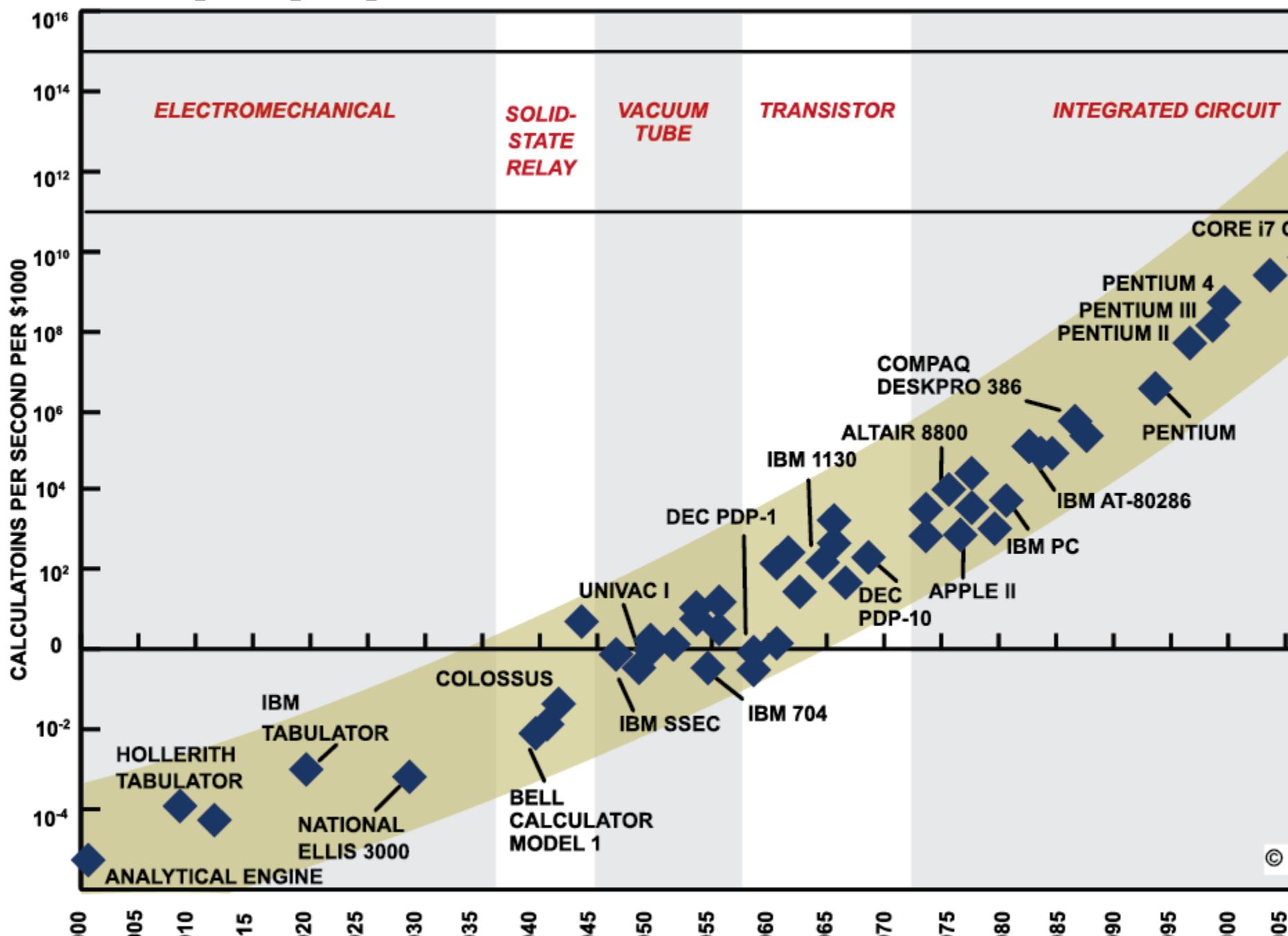
Structure	Degree of Confinement	$\frac{dN}{dE}$
Bulk Material	0D	\sqrt{E}
Quantum Well	1D	1
Quantum Wire	2D	$1/\sqrt{E}$
Quantum Dot	3D	$\delta(E)$

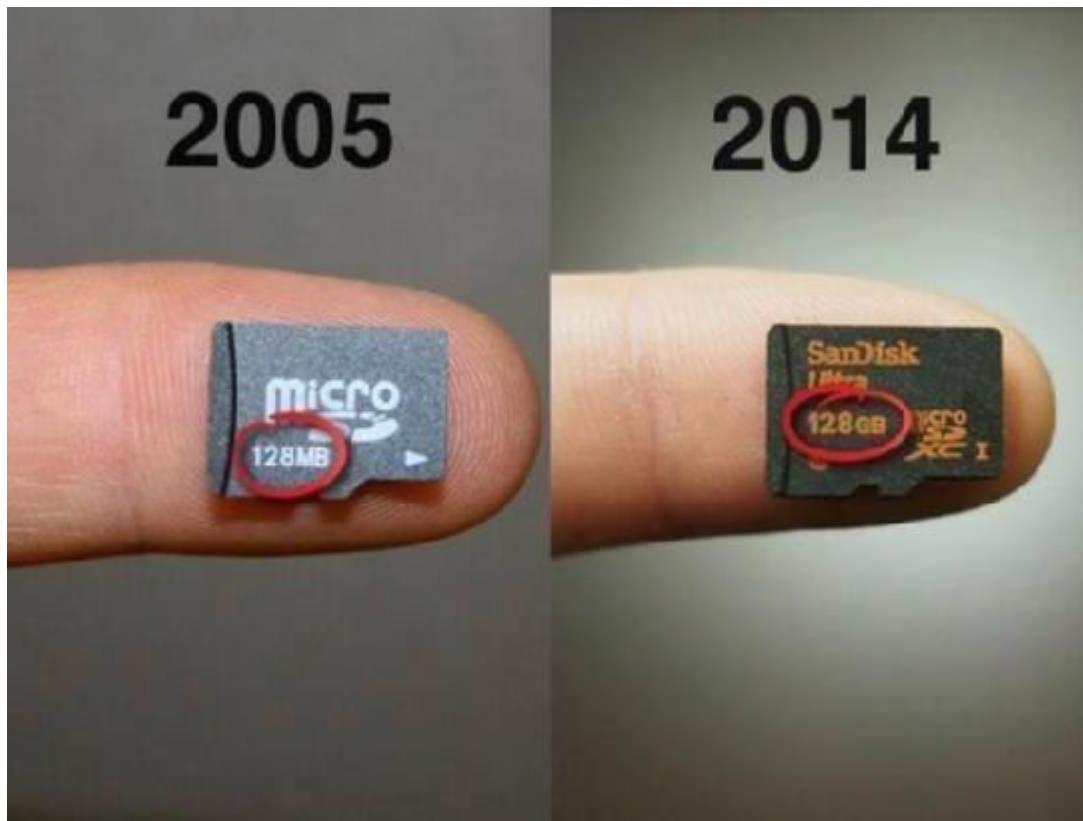
“Because of the recent rapid and radical progress in molecular electronics – where individual atoms and molecules replace lithographically drawn transistors – and related nanoscale technologies, we should be able to meet or exceed the Moore’s Law rate of progress for another 30 years.”

Moore's Law

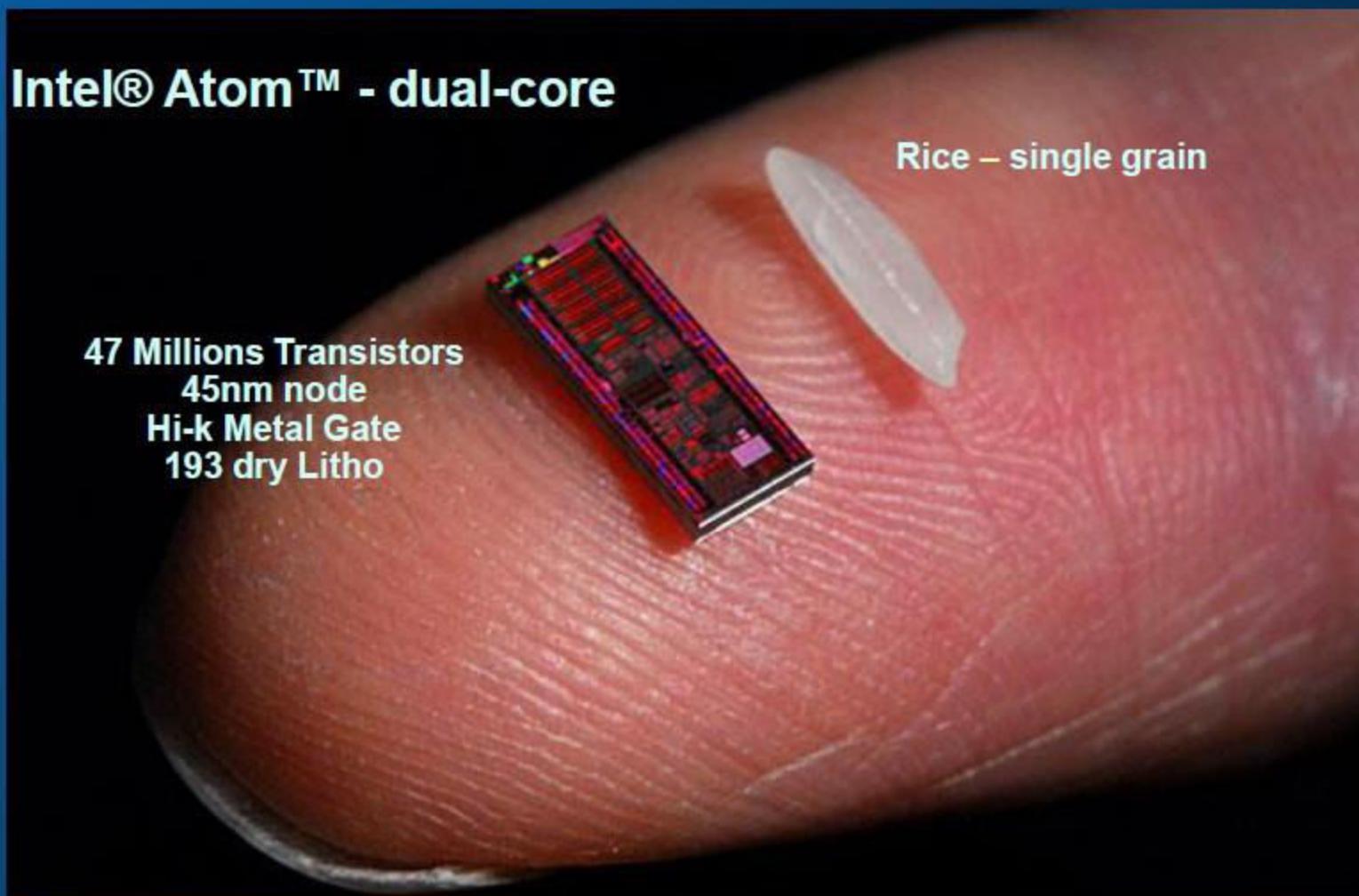




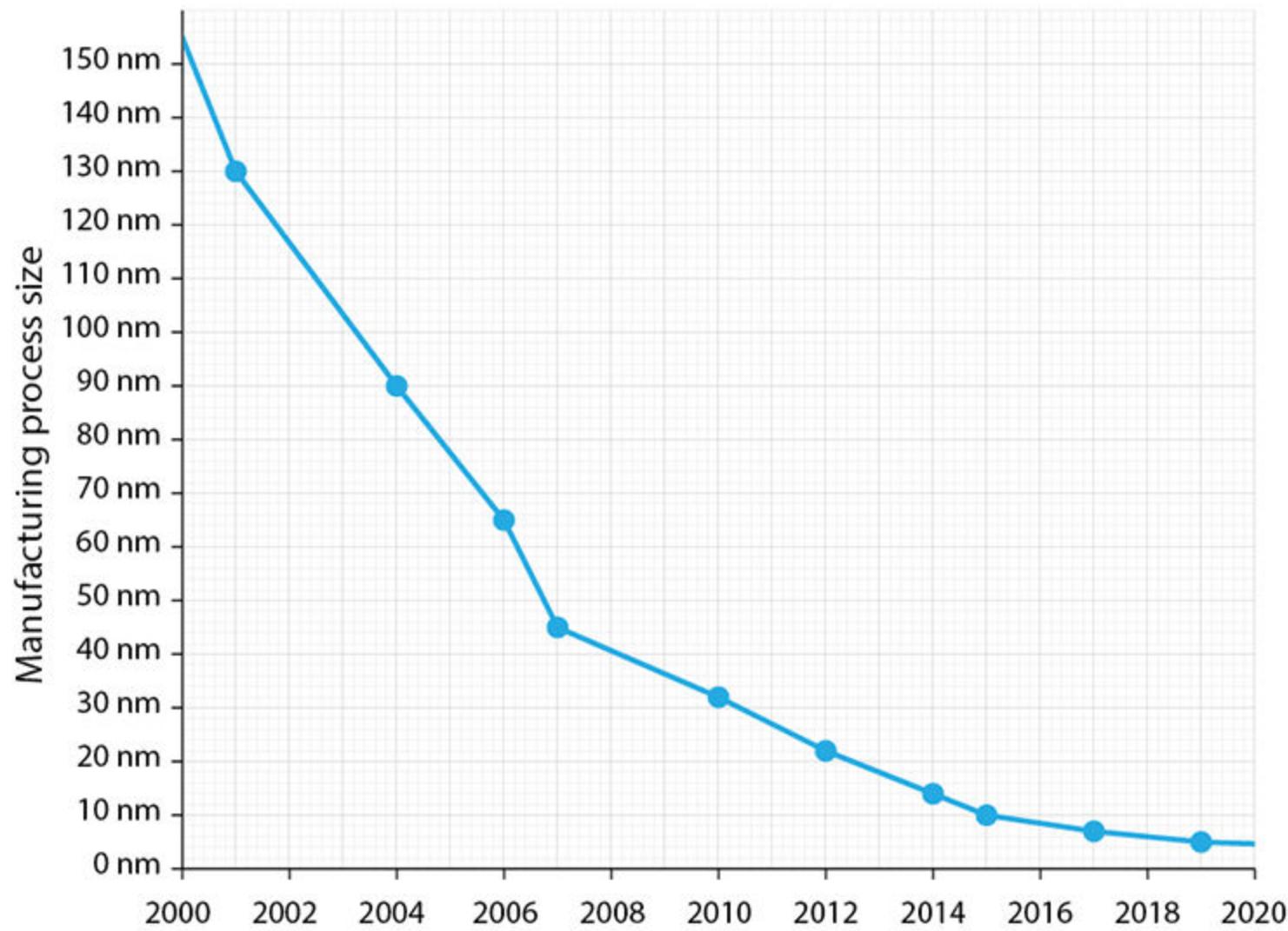




Moore's Law: circa 2008



In 2014, on 14nm technology, the above chip would be 1/8 the size
- Much smaller than the grain of rice!



Electrical Conductivity

- For metals, conductivity is based on their band structure. If the conduction band is only partially occupied by electrons, they can move in all directions without resistance (provided there is a perfect metallic crystal lattice). They are not scattered by the regular building blocks, due to the wave character of the electrons.

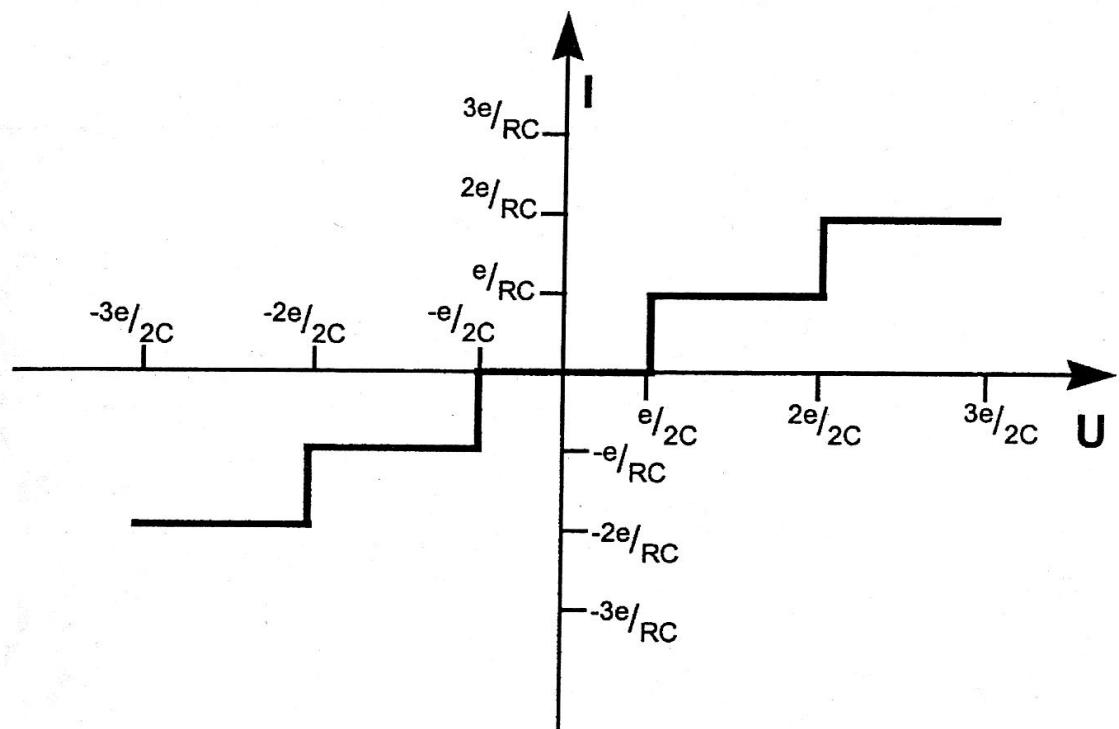
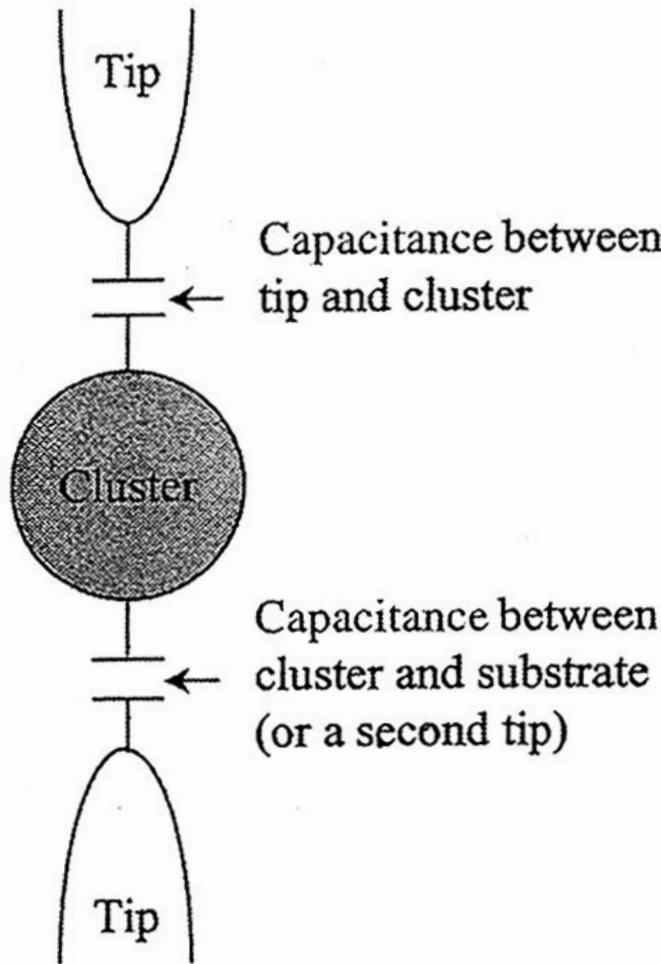
$$\mu = \frac{e\lambda}{4\pi\epsilon_0 m_e v} \quad v = \text{electron speed} \\ \epsilon_0 = \text{dielectric constant in vacuum}$$

τ , mean time between collisions, is λ/v

- For Cu, $v = 1.6 \times 10^6 \text{ m/s}$ at room temp.; $\lambda = 43 \text{ nm}$, $\tau = 2.7 \times 10^{-14} \text{ s}$

- Scattering mechanisms
 - (1) By lattice defects (foreign atoms, vacancies, interstitial positions, grain boundaries, dislocations, stacking disorders)
 - (2) Scattering at thermal vibration of the lattice (phonons)
- Item (1) is more or less independent of temperature while item (2) is independent of lattice defects, but dependent on temperature.
- Electric current  collective motion of electrons; in a bulk metal, Ohm's law: $V = RI$
- Band structure begins to change when metal particles become small. Discrete energy levels begin to dominate, and Ohm's law is no longer valid.

I-V of a Single Nanoparticle



I-V of a Single Nanoparticle

- Consider a single nanoparticle between two electrodes, but cushioned by a capacitance on each side
 - If an electron is transferred to the particle, its coulomb energy  by $E_c = e^2/2c$
 - Thermal motion of the atoms in the particle can initiate a charge & E_c , leading to further electrons tunneling uncontrollably
 - So, $kT \ll e^2/2c$
 - Tunneling current $I = V/R_T$
 - Current begins at coulomb voltage $V_c = \pm e/2c$ which is called coulomb blockade
 - Further electron transfer happens if the coulomb energy of the ‘quantum dot’ is compensated by an external voltage $V_c = \pm ne/2c$ where n is an integer
- Repeated tunneling results in a ‘staircase’ with step height in current, e/Rc
- Possible to charge and discharge a quantum dot in a quantized manner  principle behind some future computers

- Materials reduced to the nanoscale can show different properties compared to what they exhibit on a macroscale, enabling unique applications.
- **For instance, opaque substances become transparent (copper); stable materials turn combustible (aluminum); solids turn into liquids at room temperature (gold); insulators become conductors (silicon).**
- A material such as gold, which is chemically inert at normal scales, can serve as a potent chemical catalyst at nanoscales. Much of the fascination with nanotechnology stems from these quantum and surface phenomena that matter exhibits at the nanoscale.

Chemistry: The Traditional Way

- Canon Ball Chemistry
- Carried out often under extreme conditions
- Irregular, amorphous structures are formed.



Feynman: "There is plenty of room at the bottom"

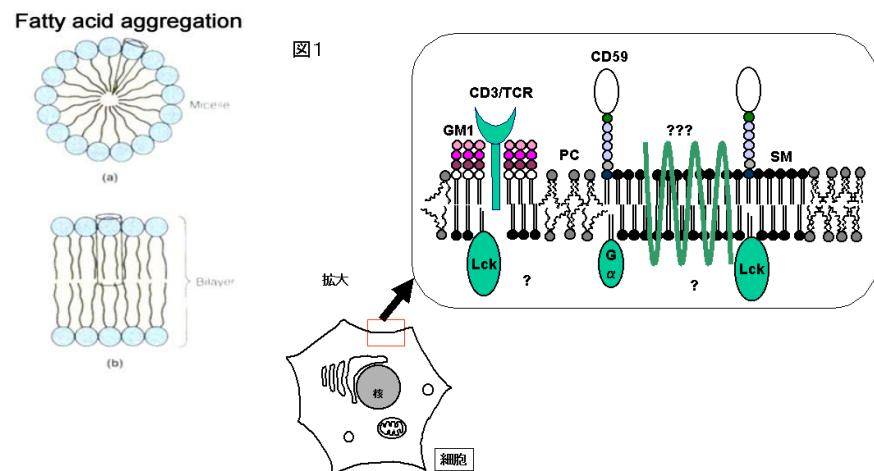
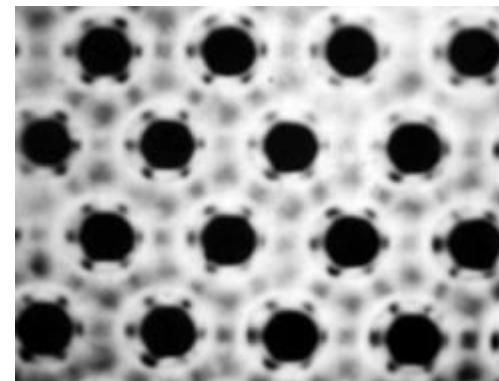
SELF Assembly

SELF-assembly involves the spontaneous and autonomous organization of disorganized interacting components into an organized pattern without direct human or mechanical interference

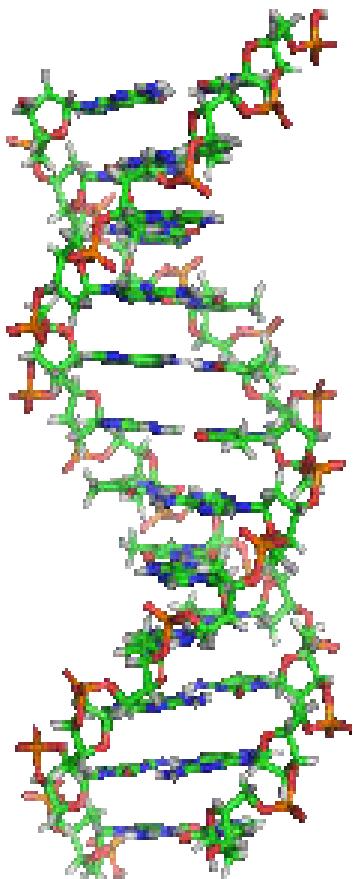
- ❖-common throughout nature and technology.
- ❖-involve components from the molecular (crystals) to the planetary (weather systems) scale and solar systems and many different kinds of interactions.
- ❖-In the natural world, self-assembly occurs over a wide range of size-scales to create structures that display new properties not present in the original components.
- ❖-In any living cell, nanoscale cellular machines spontaneously assemble themselves and drive the processes of life. Self-assembly occurs both in systems at equilibrium - such as the crystallization of proteins or colloids - and in systems far from equilibrium - such as cellular replication of DNA

Nature's Fabrication Technique: Self-assembly

- Self-aggregation of hydrophilic, lipophilic groups
- First layer creates template for growth of second layer
- Ions can be deposited on charged sites
- This kind of self-aggregation leads to ordered, hierarchical structures
- Techniques to study this atomic-scale morphology



DNA



- **Deoxyribonucleic acid**
- DNA is a long polymer of simple units called nucleotides, which are held together by a backbone made of sugars and phosphate groups.

Why DNA?

Although research has shown improvement at creating nano-sized images, it is often a challenge to create desired patterns and details at the nanoscale level.

Nanoscientists are always looking for more accurate and consistent results.

programmed self assembly.

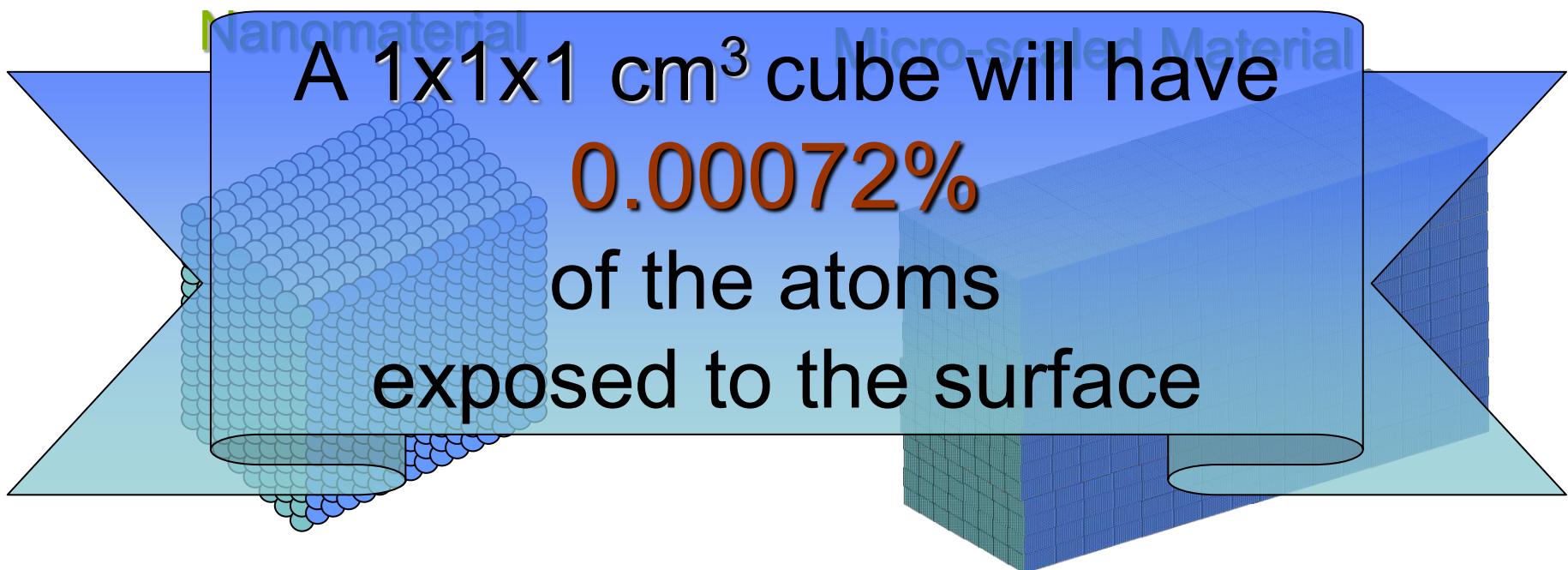
-DNA can be manipulated to create exact copies that are extremely accurate.

-DNA is predictable and programmable.

-DNA also has the ability to store enormous amounts of information

Nanomaterials Have More Atoms on the Surface

Materials of the micro ($1 \times 10^{-6}\text{m}$) and especially nano ($1 \times 10^{-9}\text{m}$) size have more atom exposed on the outside then inside



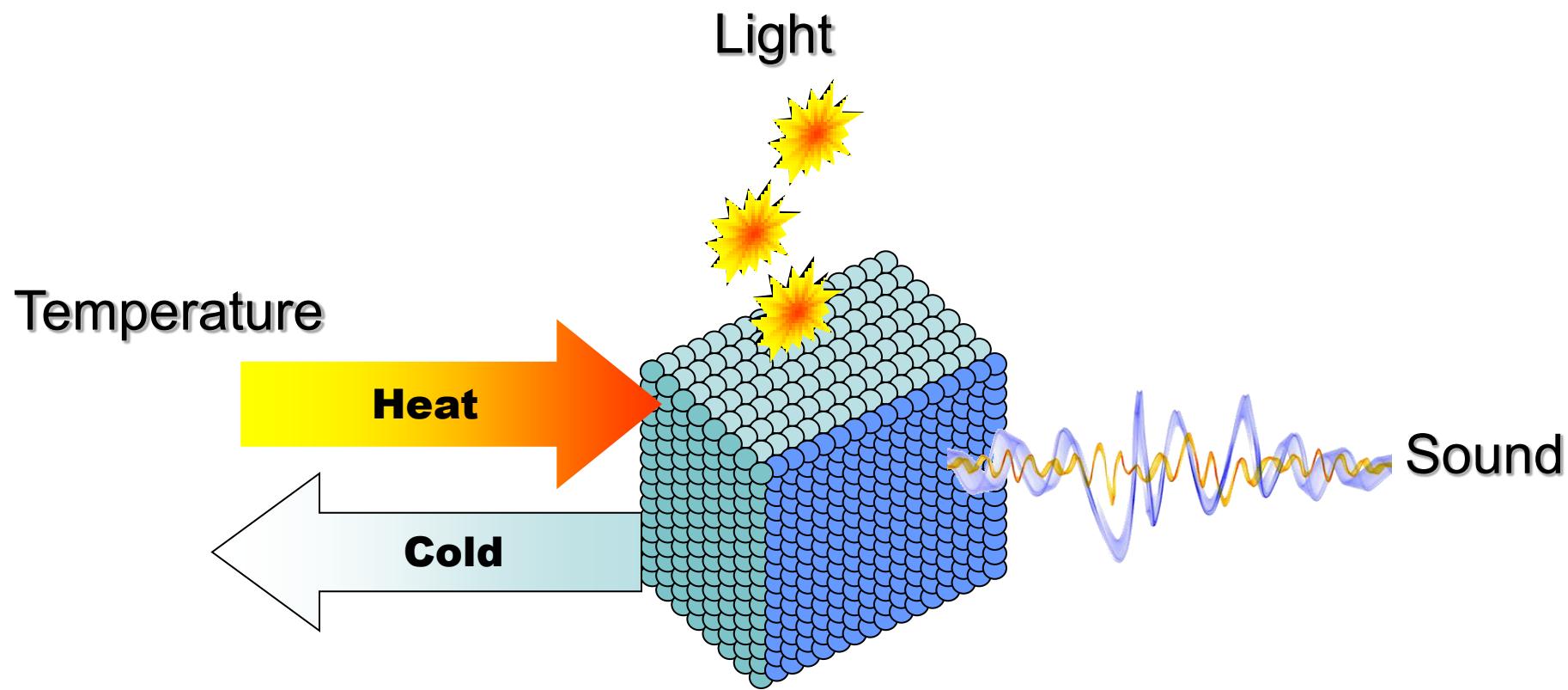
Volume = $18 \times 19 \times 1\text{ nm}^3$ or
15x8x16 atoms = 1920 atoms
total

976 or 51% of the atoms are
at the surface

Volume = $3 \times 3 \times 0.7\text{ }\mu\text{m}^3$ or
~4 million atoms total

976 or 4% of the atoms are at
the surface

Surface Atoms Interact more with the Environment



The forms of energy that affect us in the environment can affect molecules.

Energy comes from the environment to affect molecular nature.
Since more molecules are on the surface, the affect is more pronounced.

Definition:

Carbon Nanotube and Carbon fiber

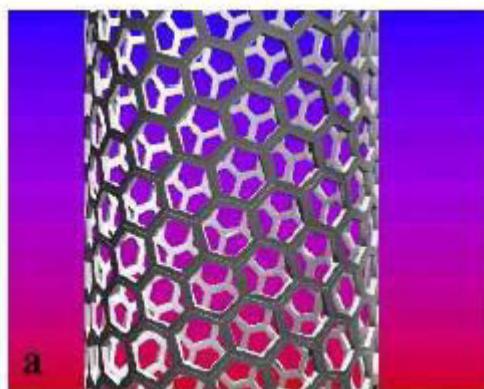
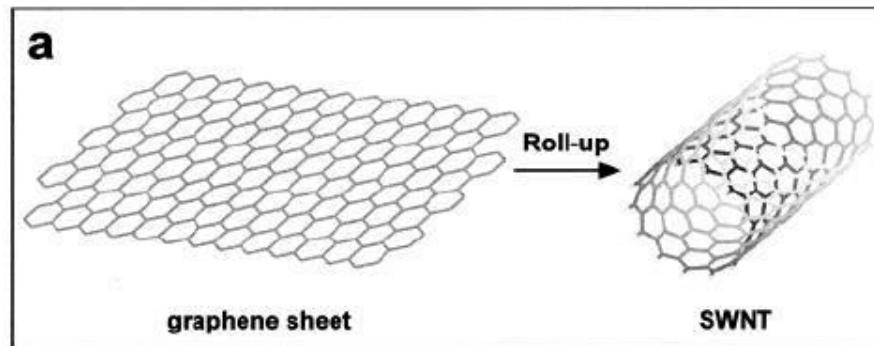
- The history of carbon fiber goes way back...



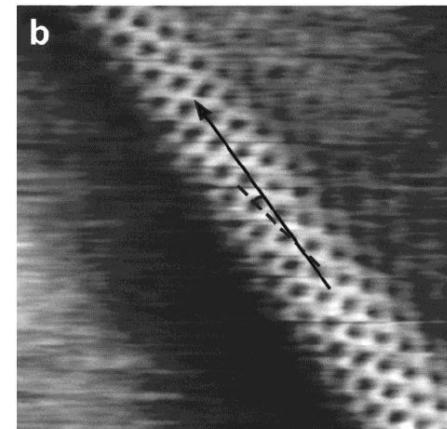
- The history of carbon nanotube starts from 1991

Carbon nanotube

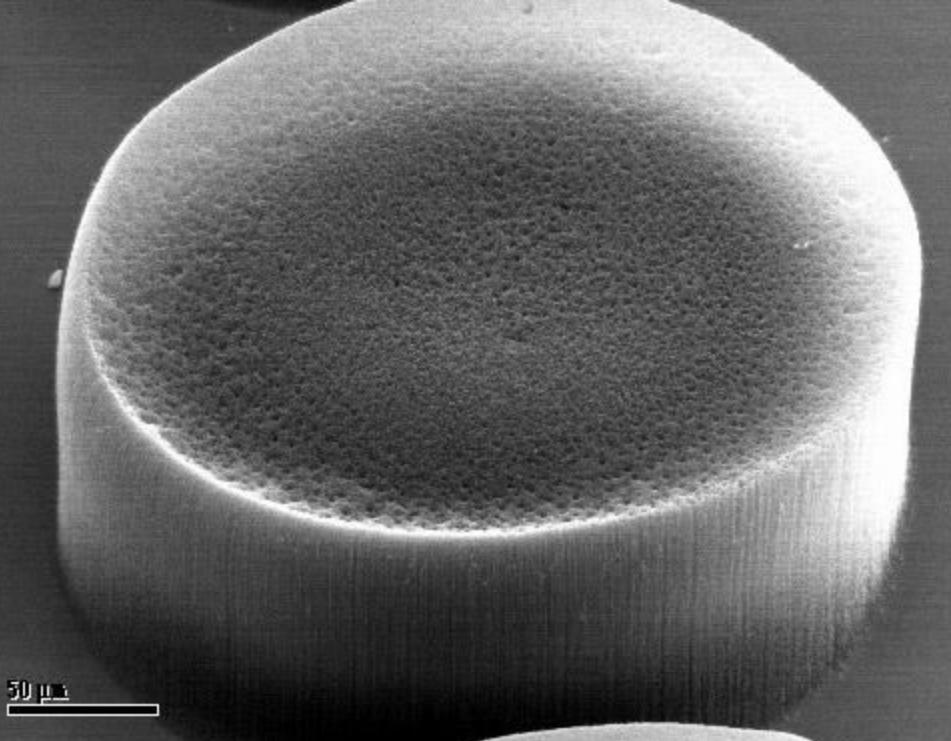
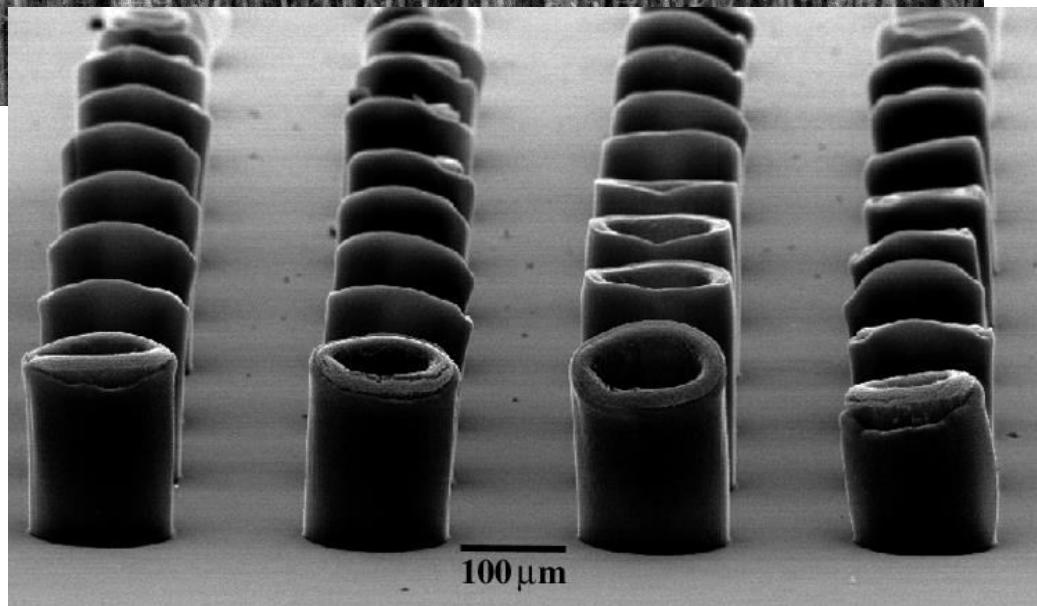
CNT: Rolling-up a graphene sheet to form a tube



Schematic
of a CNT



STM image
of CNT



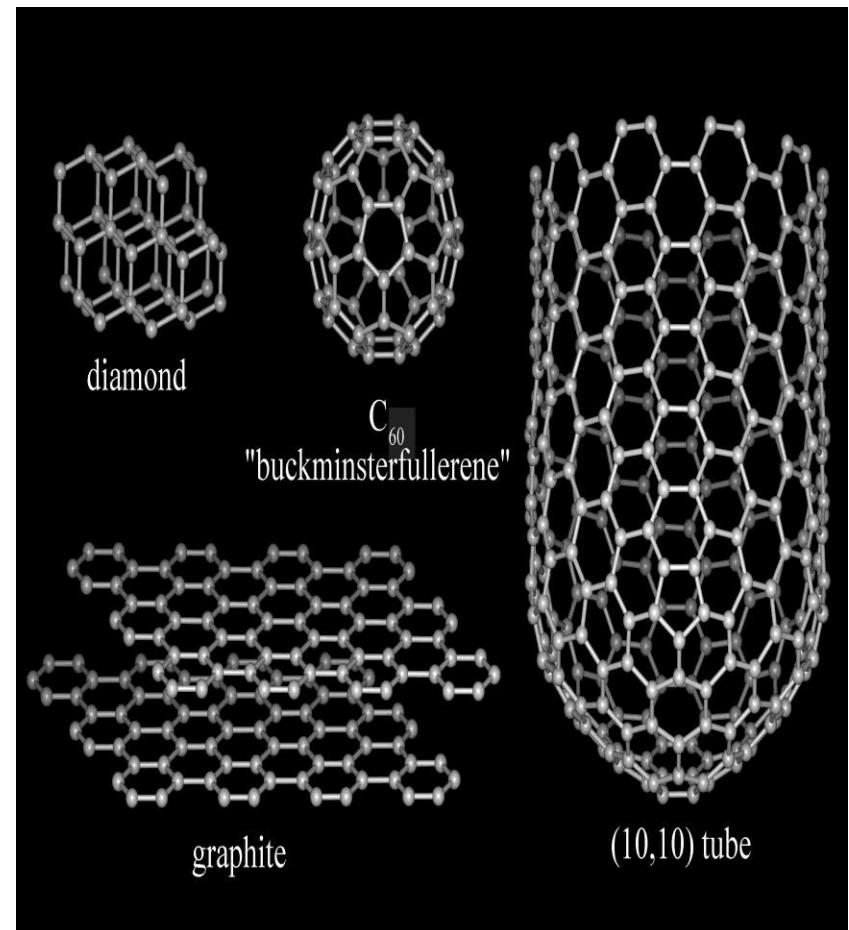
Carbon Nanotubes - SEM Images

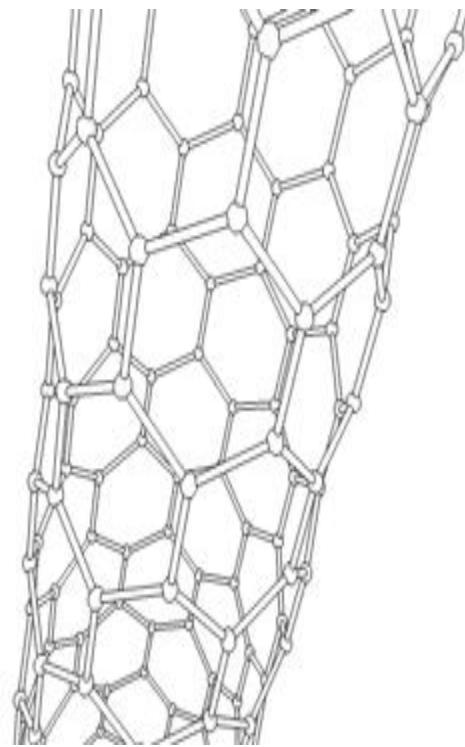
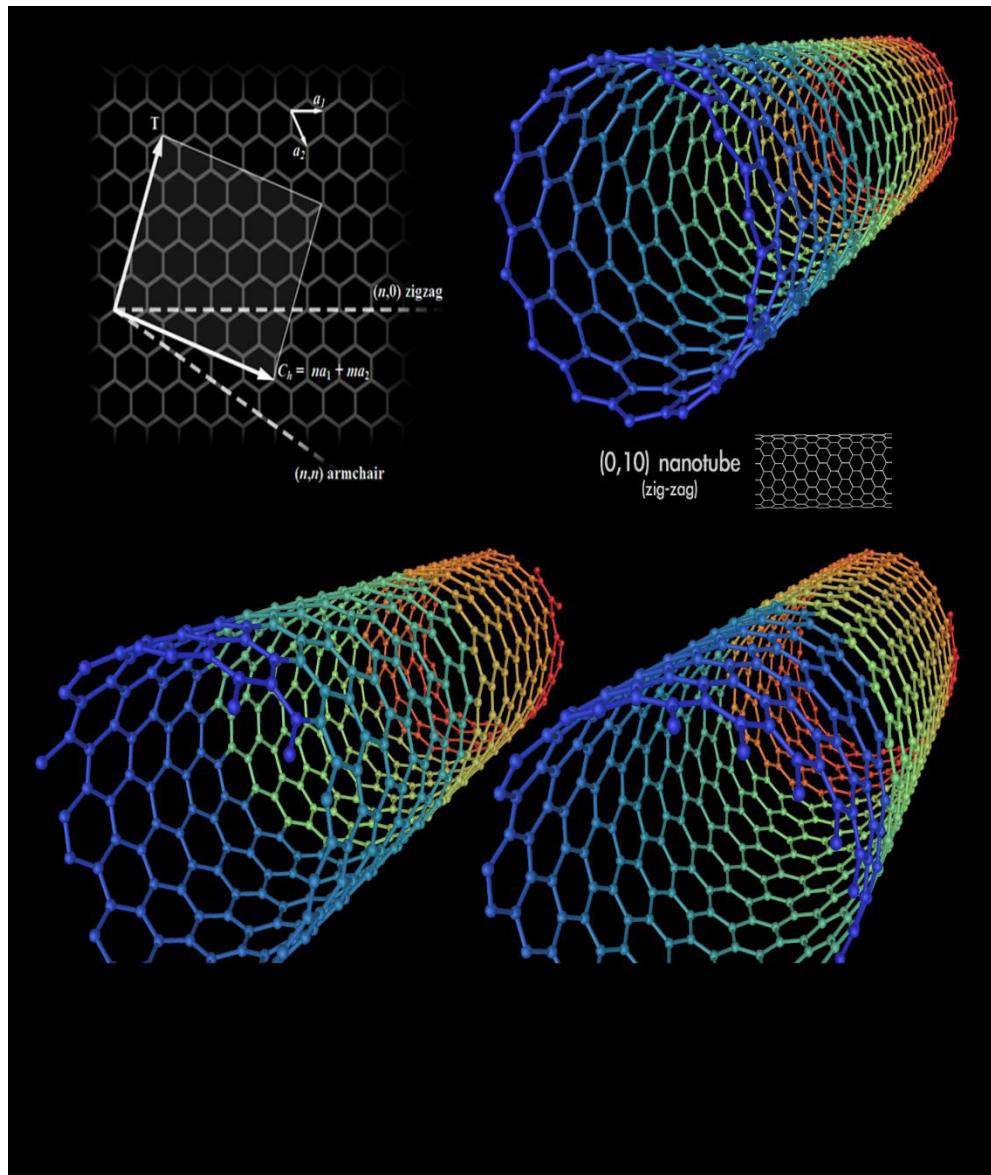
Carbon Nanotubes

- What are they?
- They are single sheet of carbon atoms rolled together. They are very small objects and exhibit many different structures and properties.
- There are 4 different types of carbon nanotubes.
 - Single walled - one atom thick layer of graphite.
 - Multi walled – multiple layers of graphite.
 - Fullerite- solid state manifestation of fullerene
 - Torus – donut shaped.

CNT Background Info

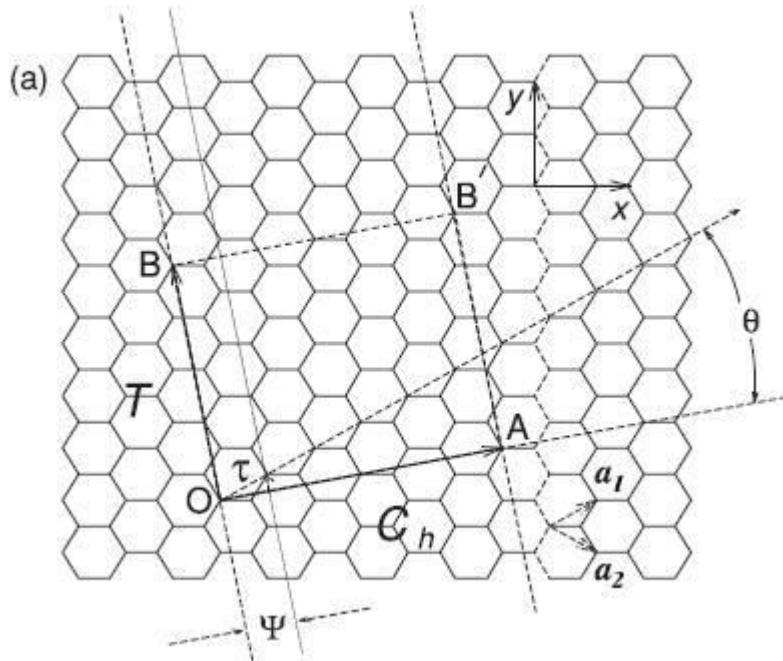
- Discovered by S. Iijima in 1991
- Tubular hexagonal arrays of graphene sheets
- Can be single-walled or multi-walled (~2 nm SWNT diameter)
- Have metallic or semiconducting properties
- Nanoelectronic Applications





Carbon nanotube

Properties depending on how it is rolled up.



a_1, a_2 are the graphene vectors.
OB/AB' overlaps after rolling up.
OA is the rolling up vector.

$$OA = na_1 + ma_2$$

Carbon nanotube properties: Electronic

Electronic band structure is determined by symmetry:

- $n=m$: Metal
- $n-m=3j$ (j non-zero integer): Tiny band-gap semiconductor
- Else: Large band-gap semiconductor.

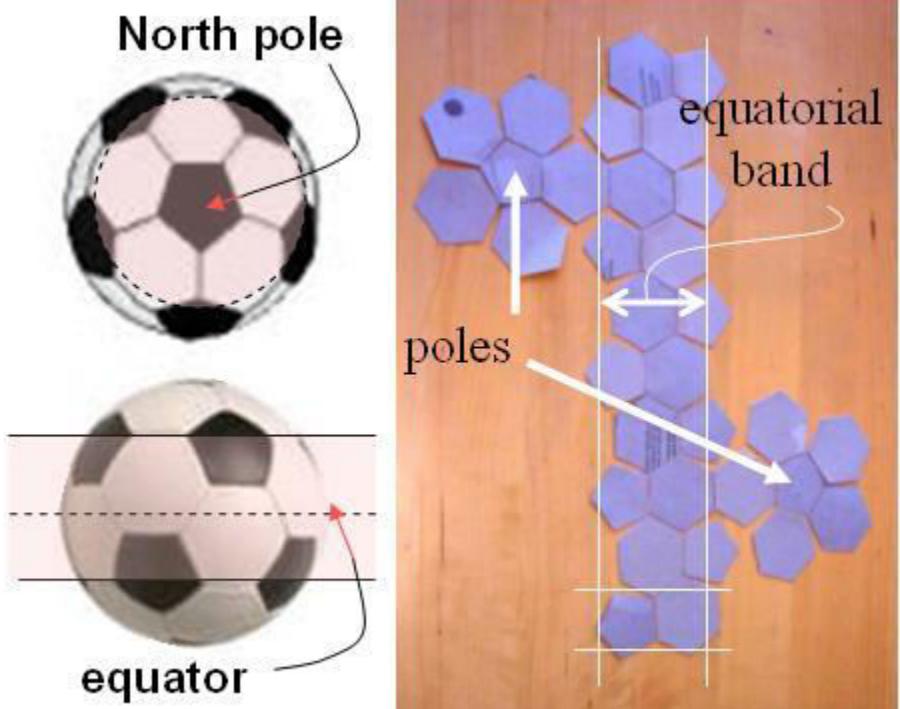
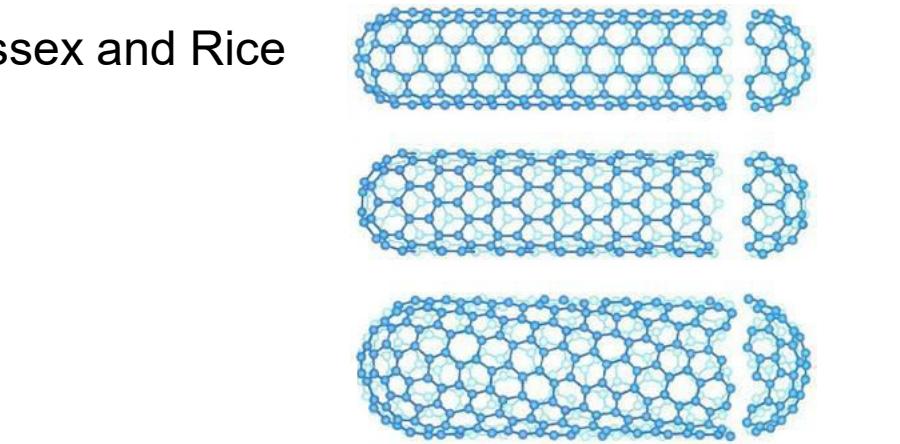
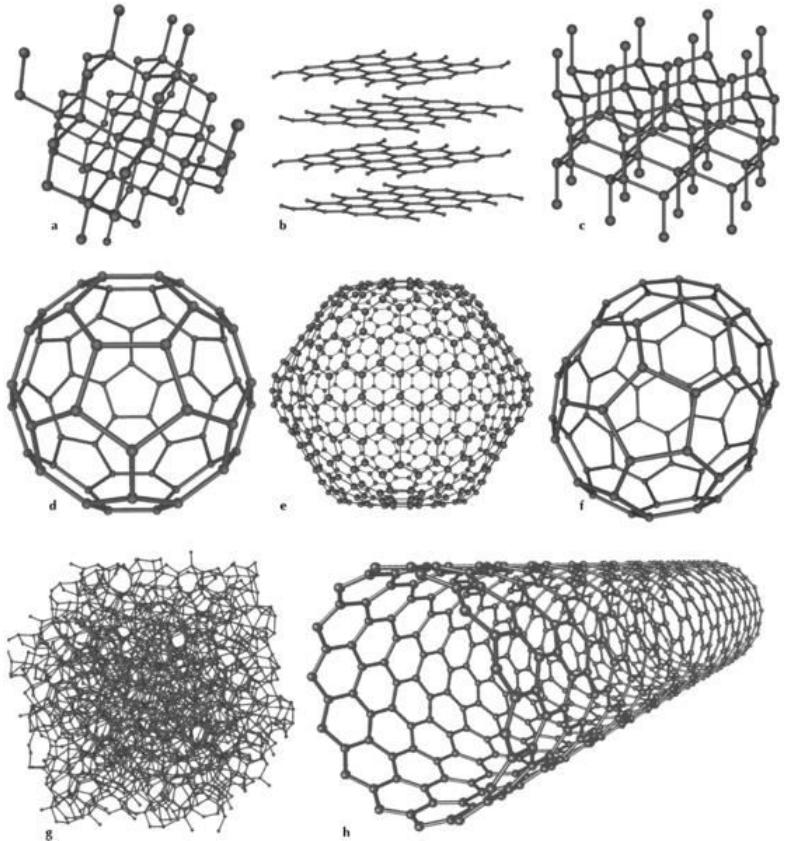
Band-gap is determined by the diameter of the tube:

- For tiny band-gap tube: $E_g \propto 1/R^2$
- For large band-gap tube: $E_g \propto 1/R$

A Stretched Out Buckey Ball Becomes a Nanotube

Fullerenes (aka buckyballs)

- Discovered in 1985 at the University of Sussex and Rice University,



Carbon nanotube properties: Mechanical

- Carbon-carbon bonds are one of the strongest bond in nature
- Carbon nanotube is composed of perfect arrangement of these bonds
- Extremely high Young's modulus

Material	Young's modulus (GPa)
Steel	190-210
SWNT	1,000+
Diamond	1,050-1,200

Applications

■ Electrical

1. Field emission in vacuum electronics
2. Building block for next generation of VLSI
3. Nano lithography

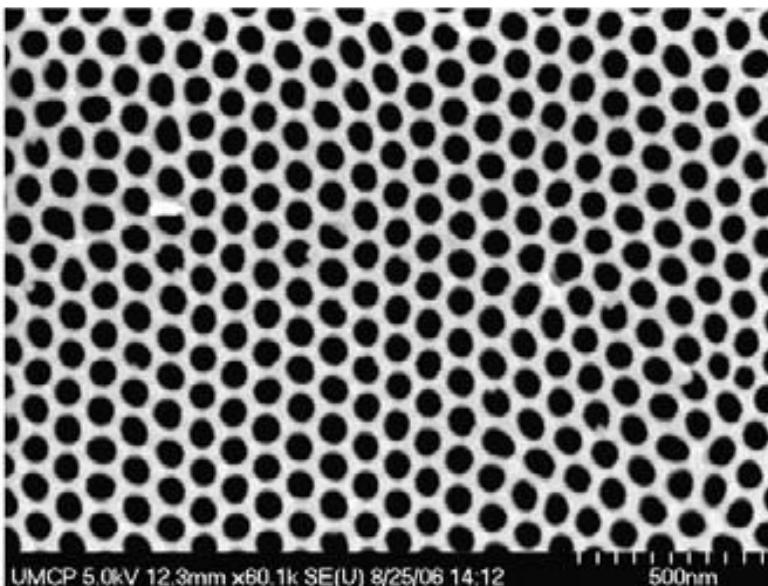
■ Energy storage

1. Lithium batteries
2. Hydrogen storage

■ Biological

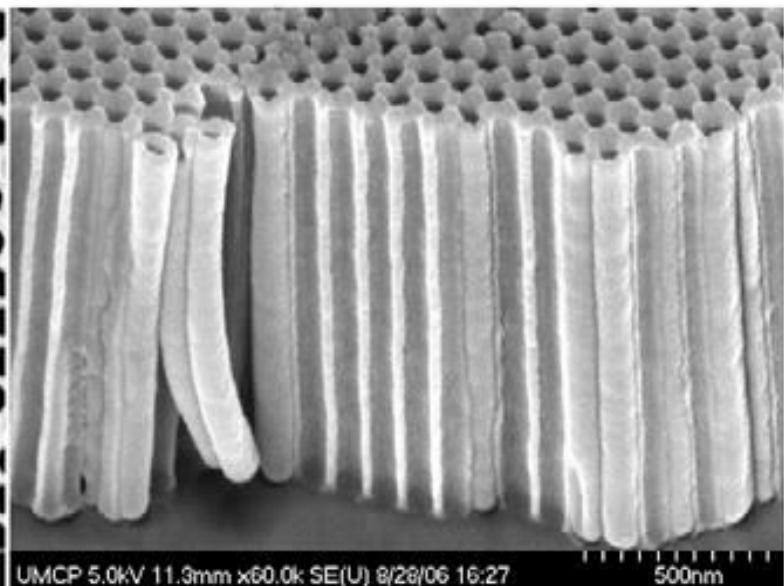
1. Bio-sensors
2. Functional AFM tips
3. DNA sequencing

Nanotubes in practice



UMCP 5.0kV 12.3mm x60.1k SE(U) 8/25/06 14:12

500nm



UMCP 5.0kV 11.3mm x60.0k SE(U) 8/26/06 16:27

500nm

Drug Discovery Today

Field emission scanning electron micrographs (FESEM) of a home-made alumina template (60-nm diameter) after silica “surface sol–gel” template synthesis; top-viewed (left) and cross-sectional viewed image (right). The cross-sectionally viewed image reveals that silica nanotubes were synthesized within the pores of the template.

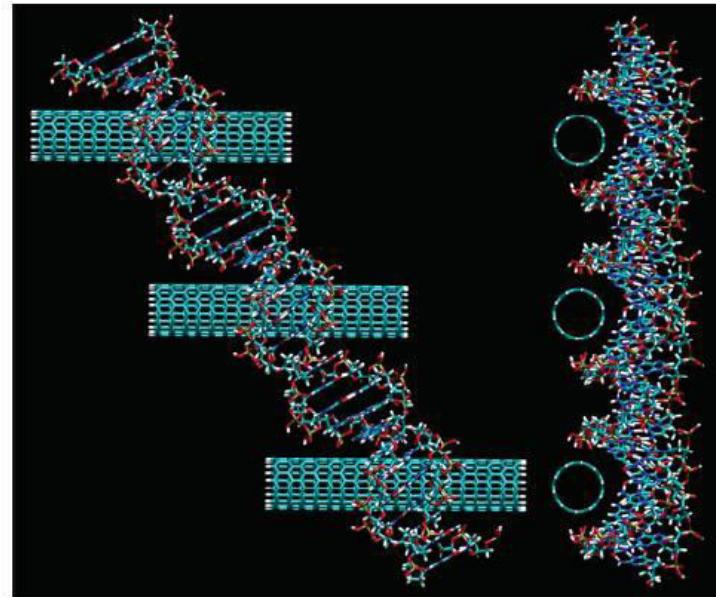
Son, S., Bai, X., Lee, S., 2007. Inorganic Hollow Nanoparticles and Nanotubes in Nanomedicine: Part 1. Drug/Gene delivery applications. *Drug Discovery Today*. Vol 12, No. 15/16, pp 650-656.

Biological applications: Bio-sensing

- Many spherical nano-particles have been fabricated for biological applications.
- Nanotubes offer some advantages relative to nanoparticles by the following aspects:
 1. Larger inner volumes – can be filled with chemical or biological species.
 2. Open mouths of nanotubes make the inner surface accessible.
 3. Distinct inner and outer surface can be modified separately.

Biological applications: DNA sequencing

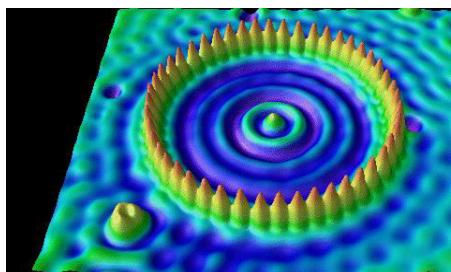
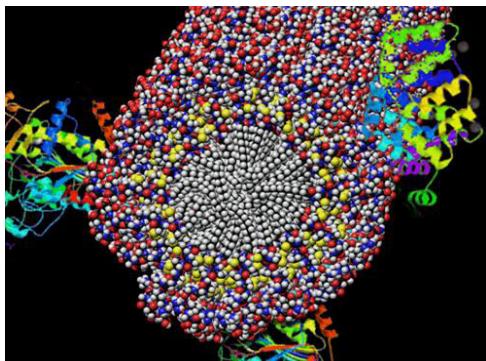
- Nanotube fits into the major groove of the DNA strand
- Apply bias voltage across CNT, different DNA base-pairs give rise to different current signals
- With multiple CNT, it is possible to do parallel fast DNA sequencing



Top view and side view of the assembled CNT-DNA system

Nanotechnology has mechanical applications

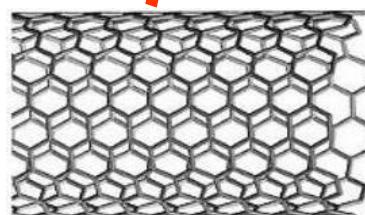
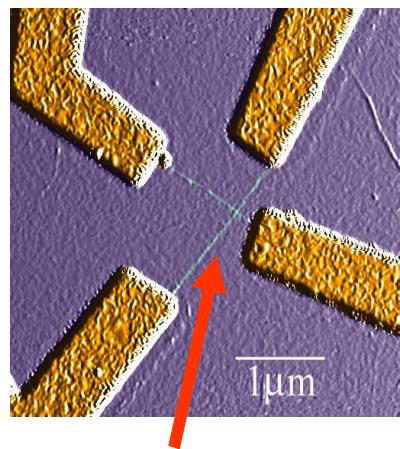
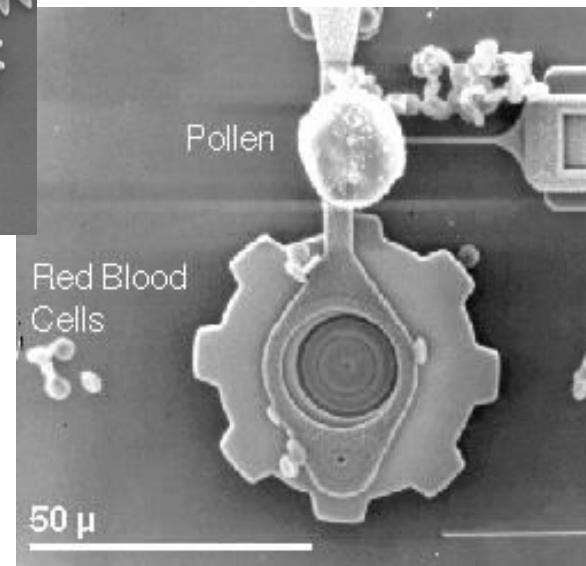
**Self-assembled,
Nature-inspired structure**
Many 10s of nm



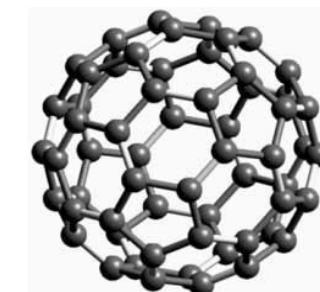
**Quantum corral of 48
iron atoms on copper
surface**
positioned one at a
time with an STM tip
Corral diameter 14 nm



**MicroElectroMechanical
(MEMS) devices**
10 - 100 μm wide

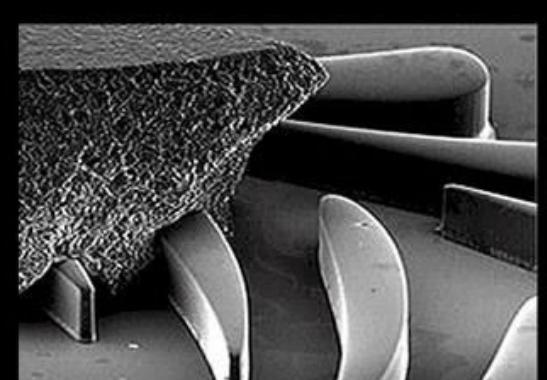
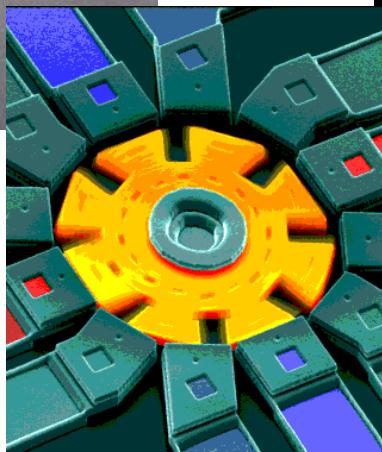
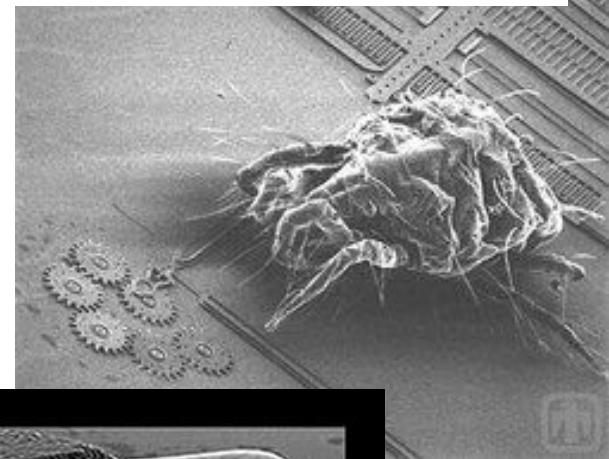
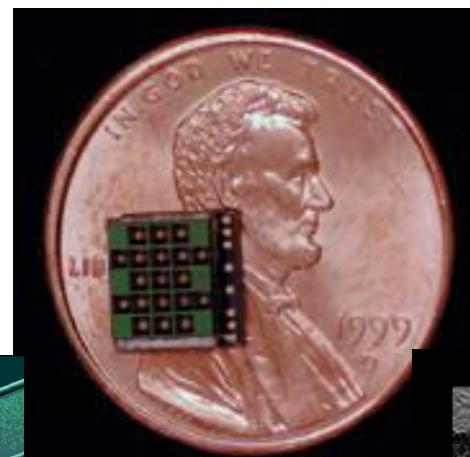
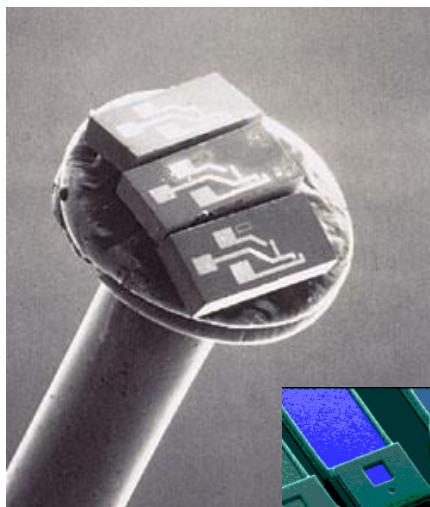


Carbon nanotube
~1.3 nm diameter



Carbon buckyball
~1 nm diameter

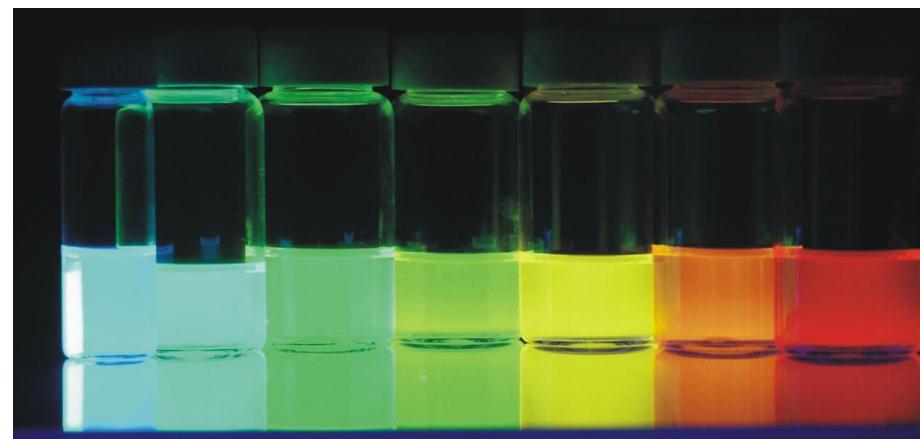
- High proportion of atom on the surface changes characteristics
 - electrostatics (static electricity)
 - wetting



Processing of SiC Microengine Rotor and Blades

Quantum Dot Colors Vary with Size

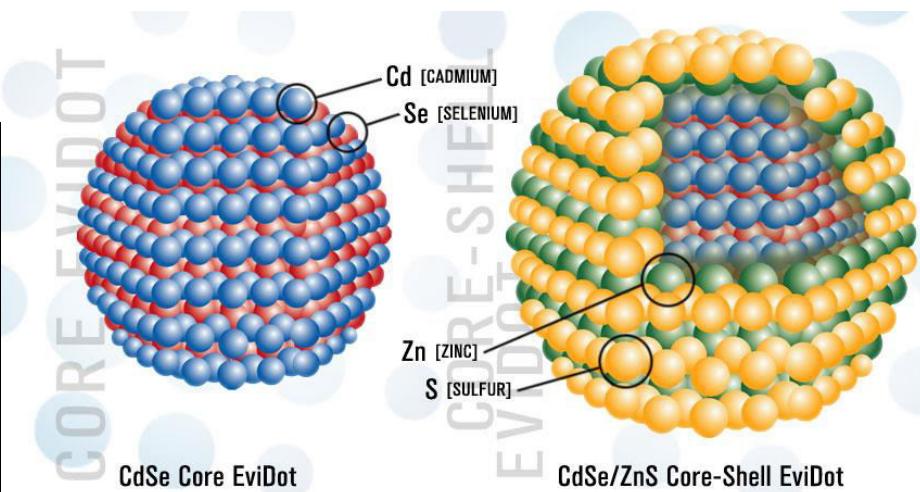
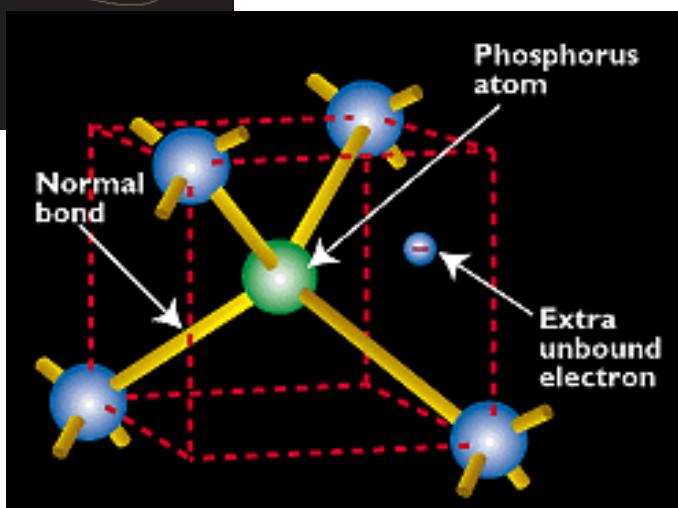
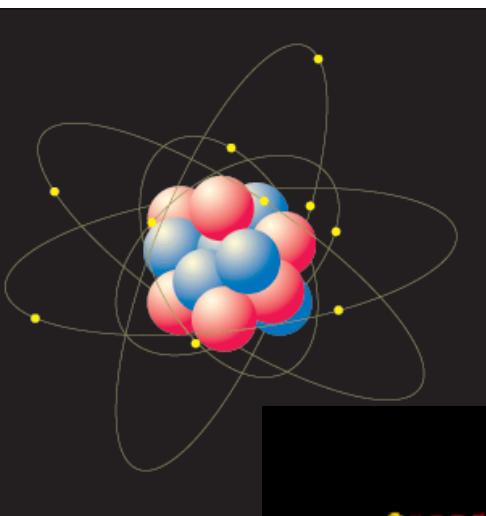
- Semiconductor based material



2.3 → 5.5

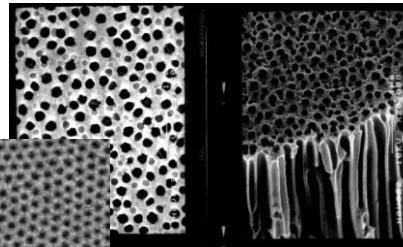
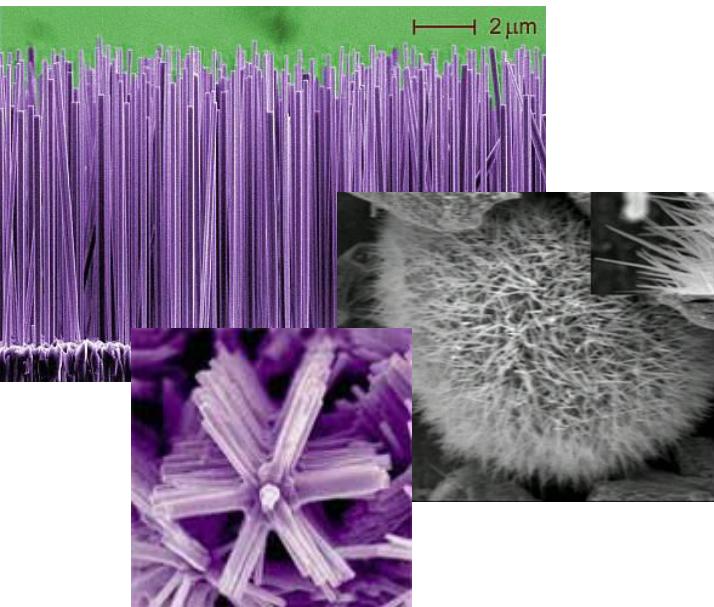
Size (nanometers)

© Copyright 2004, Benoit Dubertret

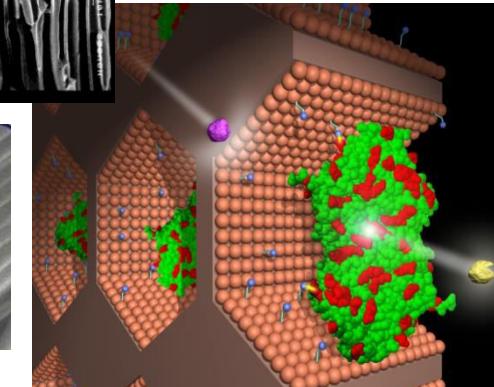


- Any material that has nano-scale features are termed a nanomaterial

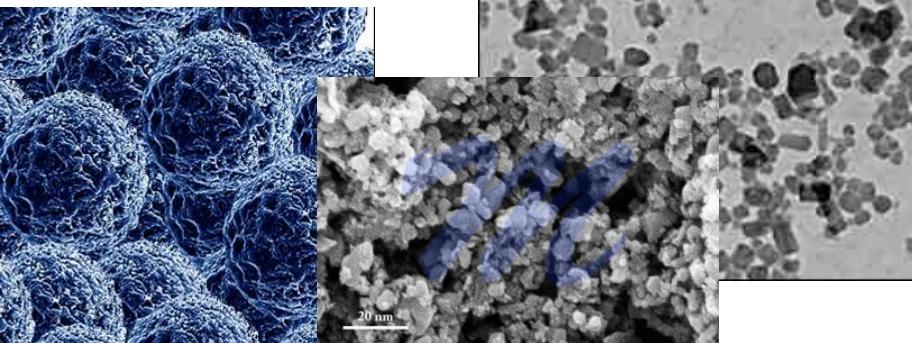
Nanowires



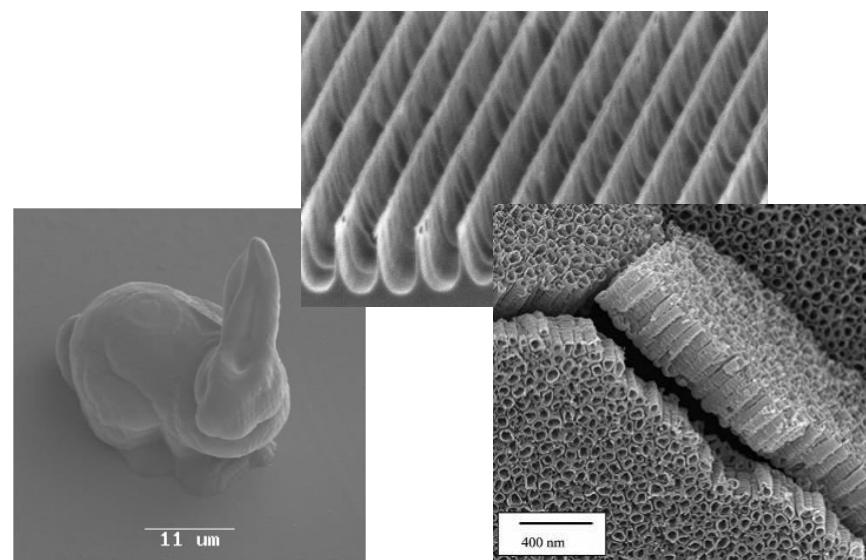
Nanomembranes



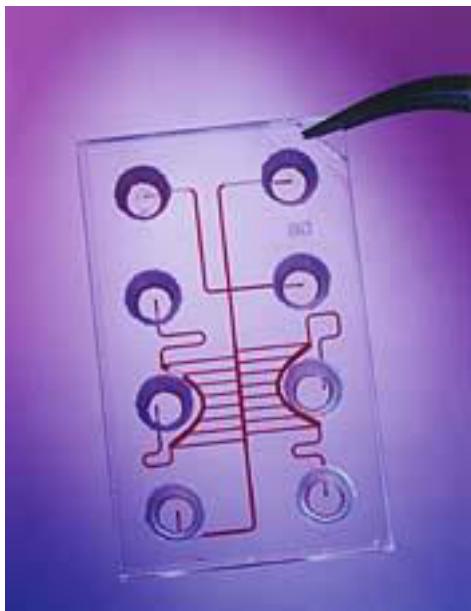
Nanoparticles



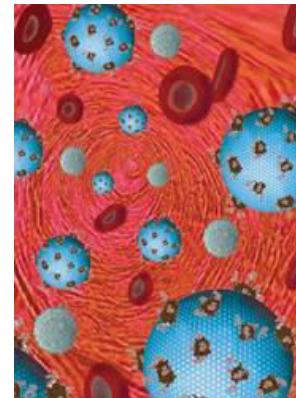
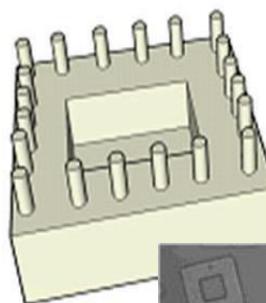
Nano-others



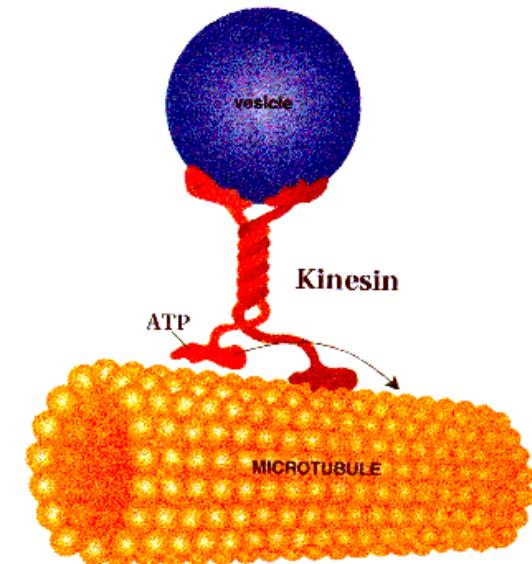
Nanotechnology has biomedical applications



Lab on a Chip
Technology on the micron scale



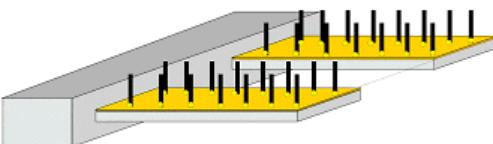
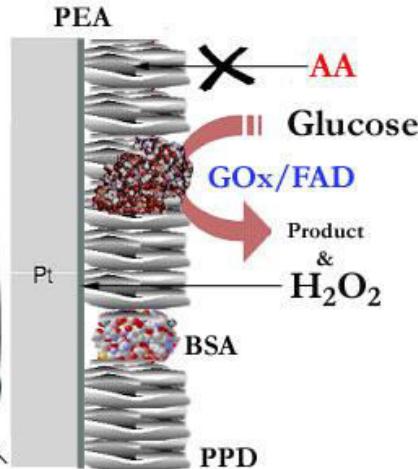
Therapeutic Drug Delivery Devices
10nm-100 μm



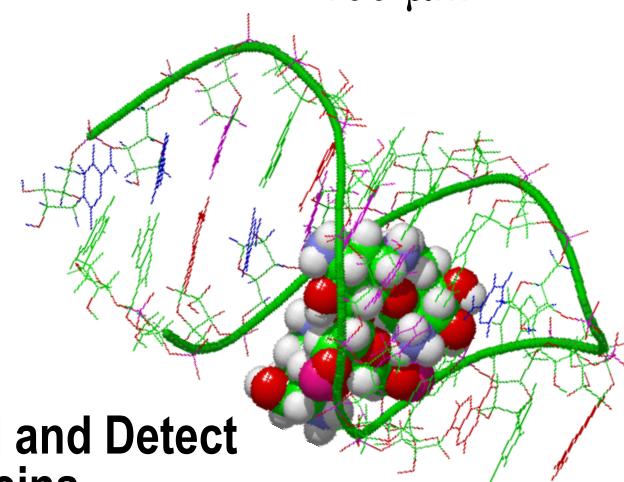
Kinesin walks on Microtubule
 \sim 100 μm



200 nm



Biosensors
Detection from DNA to Proteins
10nm-100 μm



DNA to Bind and Detect Proteins
10nm-100 μm

Sensor

- A sensor is an instrument that responds to a physical stimulus (such as heat, light, sound, pressure, magnetism, or motion)
- It collects and measures data regarding some property of a phenomenon, object, or material
- Sensors are an important part to any measurement and automation application
- The sensor is responsible for converting some type of physical phenomenon into a quantity measurable by a data acquisition (DAQ) system

Why Nanosensors

- Particles that are smaller than the characteristic lengths associated with the specific phenomena often display new chemistry and new physics that lead to new properties that depend on size
- When the size of the structure is decreased, surface to volume ratio increases considerably and the surface phenomena predominate over the chemistry and physics in the bulk
- The reduction in the size of the sensing part and/or the transducer in a sensor is important in order to better miniaturise the devices
- Science of nano materials deals with new phenomena, and new sensor devices are being built that take advantage of these phenomena
- Sensitivity can increase due to better conduction properties, the limits of detection can be lower, very small quantities of samples can be analysed, direct detection is possible without using labels, and some reagents can be eliminated.

Definition of Nanosensors

- Nanosensor: an extremely small device capable of detecting and responding to physical stimuli with dimensions on the order of one billionth of a meter
- Physical Stimuli: biological and chemical substances, displacement, motion, force, mass, acoustic, thermal, and electromagnetic

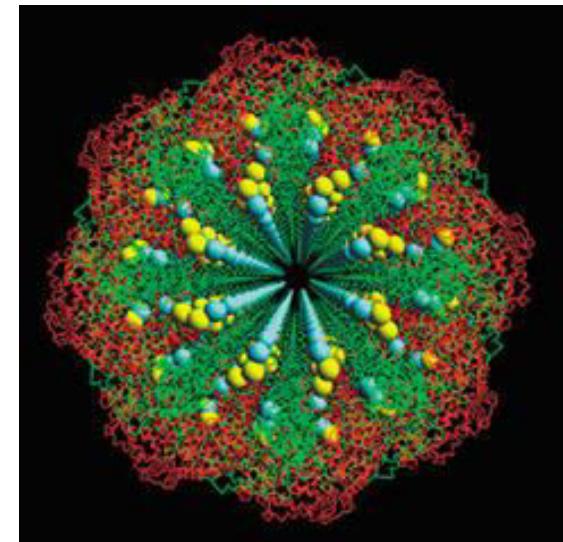


http://www.rpi.edu/locker/25/001225/public_html/nanowebprojects/laurasmith/nanosensors.ppt

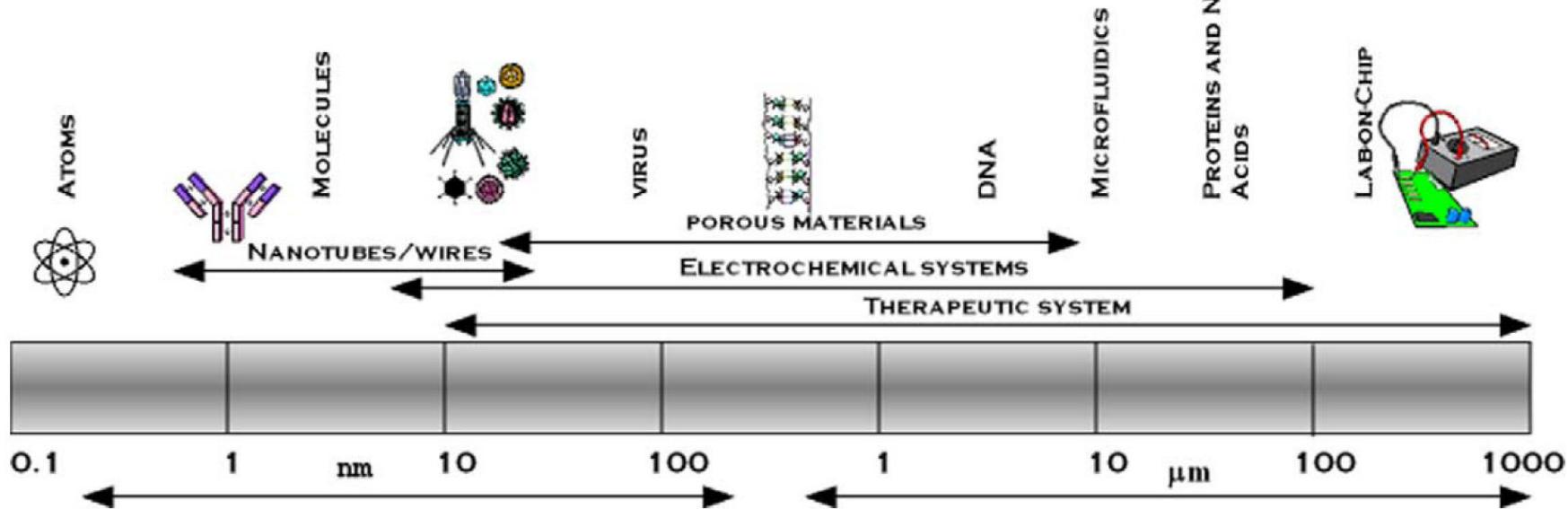
Definition of Nanosensors (cont.)

Current nanosensors device:

- Nanostructured materials - e.g. porous silicon
- Nanoparticles
- Nanoprobes
- Nanowire nanosensors
- Nanosystems
 - Cantilevers, NEMS, mostly theoretical



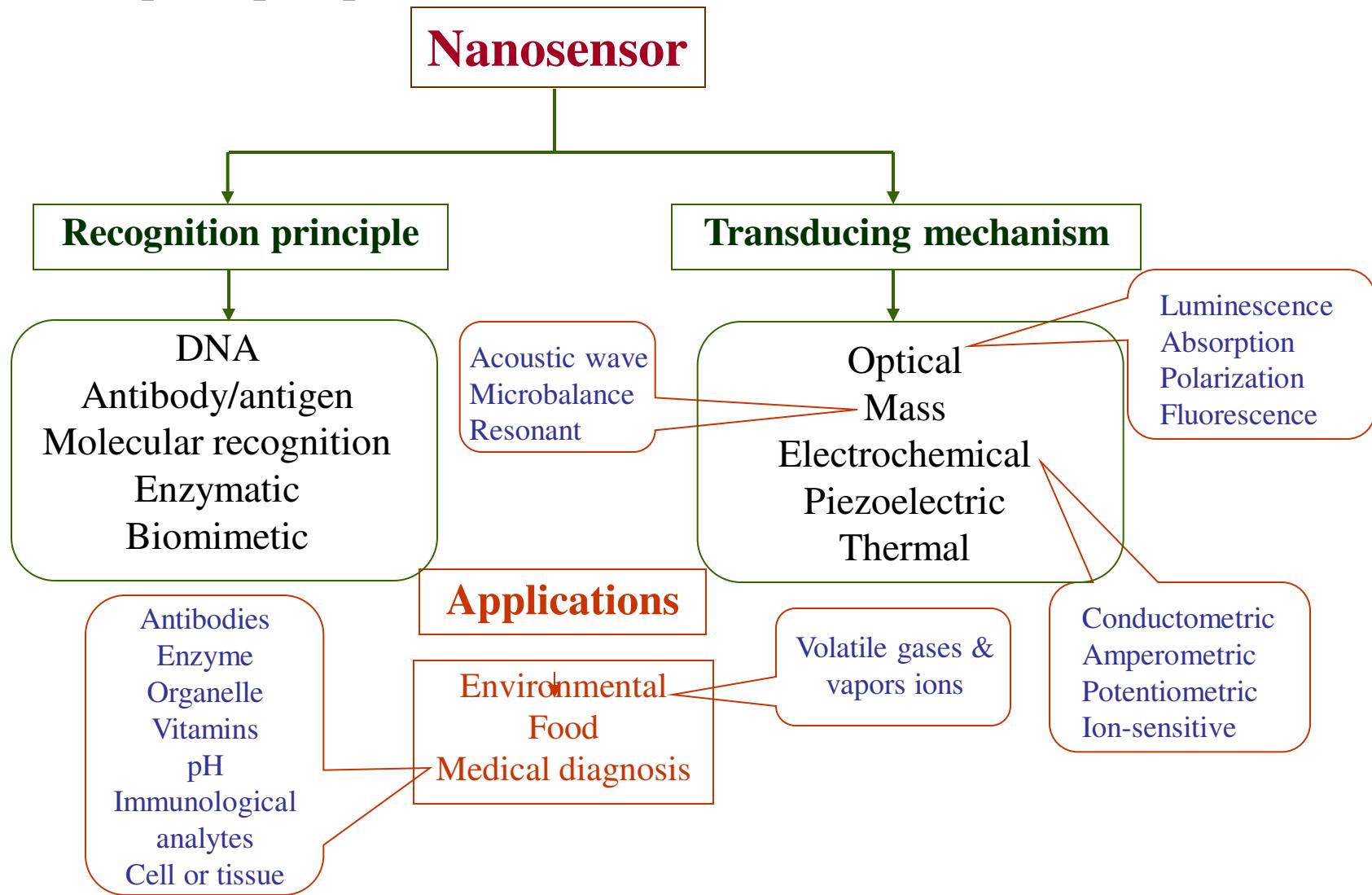
Size and compatibility



Nano sensors deliver real-time information about the antibodies to antigens, cell receptors to their glands, and DNA and RNA to nucleic acid with a complimentary sequence

Sensitivity of the conventional biosensors is in the range between 10^3 and 10^4 colony forming units (CFU)/ml. The dimensional compatibility of nanostructured materials renders nanotechnology as an obvious choice derived from its ability to detect ~ 1 CFU/ml sensitivity

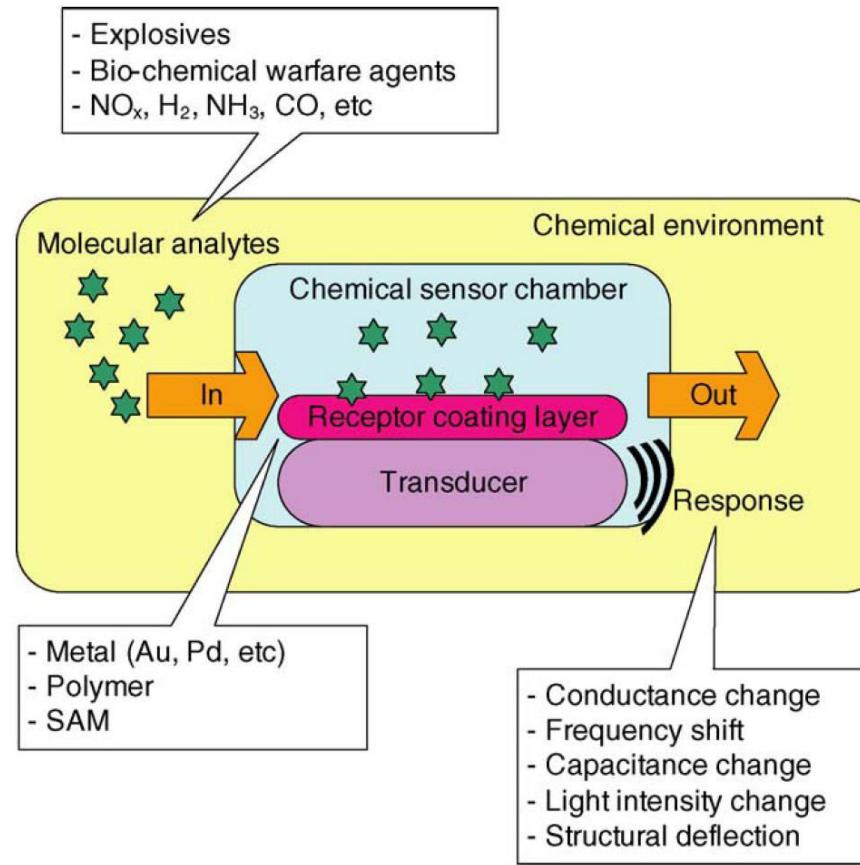
Reduced detection time than conventional methods



Scheme 1. Representation of recognition process and application of Nanosensor

Electronic Nose

General structure of a chemical sensor



Three figures of merit related to gas sensing technology:

- Reversibility
- Sensitivity
- Selectivity

Exceptional properties of carbon nanotubes

- ✓ CNT have a high length-to-radius ratio, which allows for greater control over the unidirectional properties of the materials produced
- ✓ they can behave as metallic, semiconducting or insulating material depending on their diameter, their chirality, and any functionalisation or doping
- ✓ they have a high degree of mechanical strength. In fact they have a greater mechanical strength and flexibility than carbon fibres
- ✓ their properties can be altered by encapsulating metals inside them to make electrical or magnetic nanocables or even gases, thus making them suitable for storing hydrogen or separating gases

Metal oxide nano-crystals for sensing

- Metal oxides possess a broad range of electronic, chemical, and physical properties that are often highly sensitive to changes in their chemical environment.
- The sensing properties of semiconductor metal oxide (nano-belts, nano-wires or nano-ribbons) assures improved selectivity and stability due to there crystallinity
- Their peculiar characteristics and size effects make them interesting both for fundamental studies and for potential nano-device applications, leading to a third generation of metal oxide gas sensors

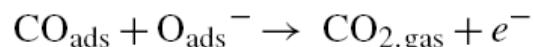
Working principle of metal oxide gas sensors

1. Conductimetric metal oxide gas sensors rely on changes of electrical conductivity due to the interaction with the surrounding atmosphere
2. When a metal oxide is semiconducting, the charge transfer process induced by surface reactions determines its resistance

Sensing mechanism in metal oxide gas sensors is related to ionosorption of species over their surfaces

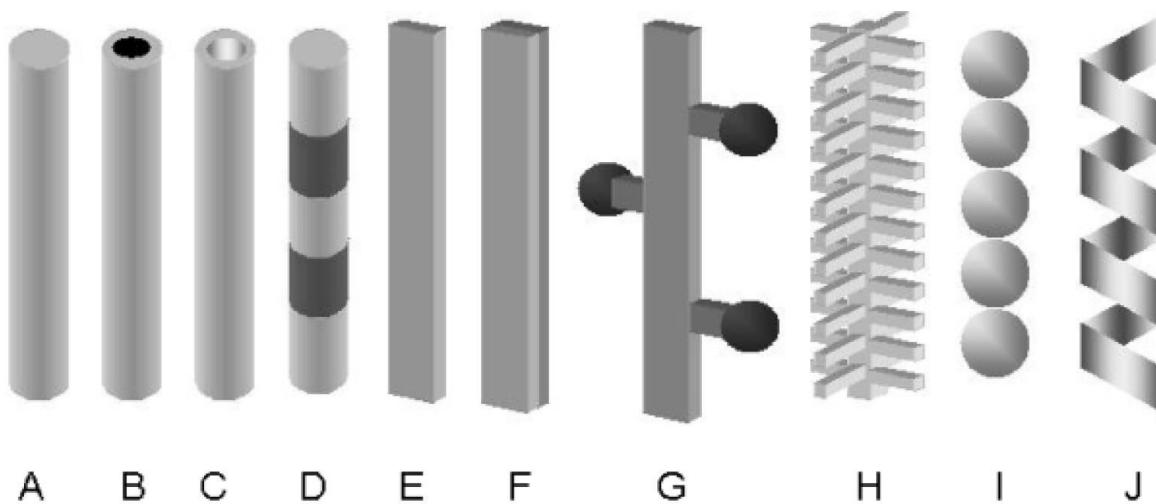
Ionosorbed species when operating in ambient air are oxygen and water

For some reducing gases, gas detection is related to the reactions between the species to be detected and ionosorbed surface oxygen

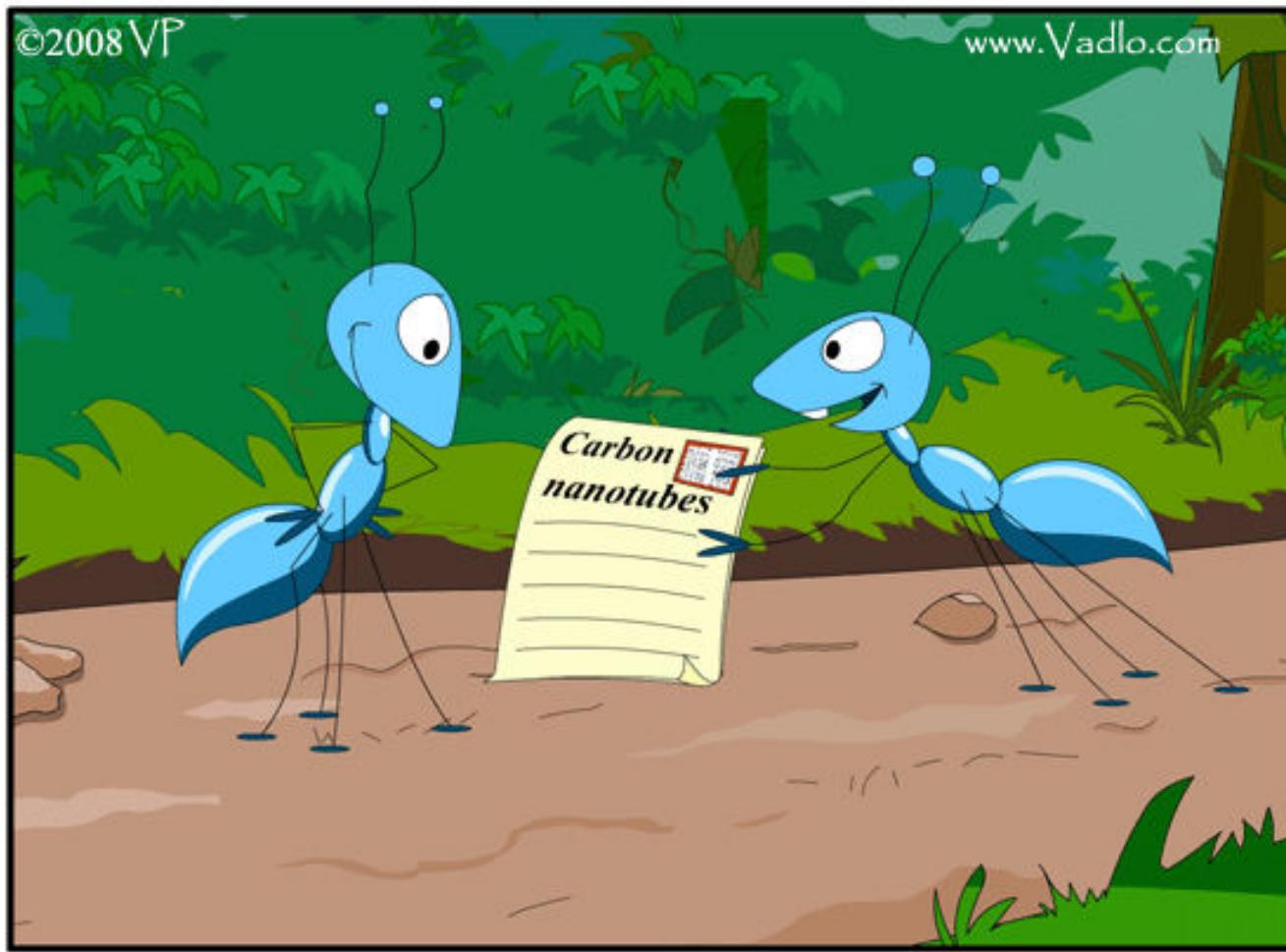


These consume ionosorbed oxygen and in turn change the electrical conductance of metal oxide

A schematic summary of the kinds of quasi-one-dimensional metaloxide nanostructures



(A) nanowires and nanorods; (B) core-shell structures with metallic inner core, semiconductor, or metal-oxide; (C) nanotubules/nanopipes and hollow nanorods; (D) heterostructures; (E) nanobelts/nanoribbons; (F) nanotapes, (G) dendrites, (H) hierarchical nanostructures; (I) nanosphere assembly; (J) nanosprings.



Finally, we can drink Coke with a straw.

1

Lasers: Fundamentals, Types, and Operations

Subhash Chandra Singh, Haibo Zeng, Chunlei Guo, and Weiping Cai

The acronym LASER, constructed from Light Amplification by Stimulated Emission of Radiation, has become so common and popular in every day life that it is now referred to as *laser*. Fundamental theories of lasers, their historical development from milliwatts to petawatts in terms of power, operation principles, beam characteristics, and applications of laser have been the subject of several books [1–5]. Introduction of lasers, types of laser systems and their operating principles, methods of generating extreme ultraviolet/vacuum ultraviolet (EUV/VUV) laser lights, properties of laser radiation, and modification in basic structure of lasers are the main sections of this chapter.

1.1

Introduction of Lasers

1.1.1

Historical Development

The first theoretical foundation of LASER and MASER was given by Einstein in 1917 using Plank's law of radiation that was based on probability coefficients (Einstein coefficients) for absorption and spontaneous and stimulated emission of electromagnetic radiation. *Theodore Maiman* was the first to demonstrate the earliest practical laser in 1960 after the reports by several scientists, including the first theoretical description of *R.W. Ladenburg* on stimulated emission and negative absorption in 1928 and its experimental demonstration by *W.C. Lamb* and *R.C. Rutherford* in 1947 and the proposal of *Alfred Kastler* on optical pumping in 1950 and its demonstration by *Brossel, Kastler, and Winter* two years later. *Maiman's* first laser was based on optical pumping of synthetic ruby crystal using a flash lamp that generated pulsed red laser radiation at 694 nm. Iranian scientists *Javan* and *Bennett* made the first gas laser using a mixture of He and Ne gases in the ratio of 1 : 10 in the 1960. *R. N. Hall* demonstrated the first diode laser made of gallium arsenide (GaAs) in 1962, which emitted radiation at 850 nm, and later in the same year *Nick Holonyak* developed the first semiconductor visible-light-emitting laser.

1.1.2

Basic Construction and Principle of Lasing

Basically, every laser system essentially has an active/gain medium, placed between a pair of optically parallel and highly reflecting mirrors with one of them partially transmitting, and an energy source to pump active medium. The gain media may be solid, liquid, or gas and have the property to amplify the amplitude of the light wave passing through it by stimulated emission, while pumping may be electrical or optical. The gain medium used to place between pair of mirrors in such a way that light oscillating between mirrors passes every time through the gain medium and after attaining considerable amplification emits through the transmitting mirror.

Let us consider an active medium of atoms having only two energy levels: excited level E_2 and ground level E_1 . If atoms in the ground state, E_1 , are excited to the upper state, E_2 , by means of any pumping mechanism (optical, electrical discharge, passing current, or electron bombardment), then just after few nanoseconds of their excitation, atoms return to the ground state emitting photons of energy $h\nu = E_2 - E_1$. According to Einstein's 1917 theory, emission process may occur in two different ways, either it may induced by photon or it may occur spontaneously. The former case is termed as *stimulated emission*, while the latter is known as *spontaneous emission*. Photons emitted by stimulated emission have the same frequency, phase, and state of polarization as the stimulating photon; therefore they add to the wave of stimulating photon on a constructive basis, thereby increasing its amplitude to make lasing. At thermal equilibrium, the probability of stimulated emission is much lower than that of spontaneous emission ($1 : 10^{33}$), therefore most of the conventional light sources are incoherent, and only lasing is possible in the conditions other than the thermal equilibrium.

1.1.3

Einstein Relations and Gain Coefficient

Consider an assembly of N_1 and N_2 atoms per unit volume with energies E_1 and $E_2 (E_2 > E_1)$ is irradiated with photons of density $\rho_\nu = N h\nu$, where $[N]$ is the number of photons of frequency ν per unit volume. Then the stimulated absorption and stimulated emission rates may be written as $N_1 \rho_\nu B_{12}$ and $N_2 \rho_\nu B_{21}$ respectively, where B_{12} and B_{21} are constants for up and downward transitions, respectively, between a given pair of energy levels. Rate of spontaneous transition depends on the average lifetime, τ_{21} , of atoms in the excited state and is given by $N_2 A_{21}$, where A_{21} is a constant. Constants B_{12} , B_{21} , and A_{21} are known as *Einstein coefficients*. Employing the condition of thermal equilibrium in the ensemble, Boltzmann statistics of atomic distribution, and Planck's law of blackbody radiation, it is easy to find out $B_{12} = B_{21}$, $A_{21} = B_{21}(8\pi h\nu^3/c^3)$, known as *Einstein relations*, and ratio, $R = \exp(h\nu/kT) - 1$, of spontaneous and stimulated emissions rates. For example, if we have to generate light of 632.8 nm ($\nu = 4.74 \times 10^{14}$ Hz) wavelength at room temperature from the system of He–Ne, the ratio of spontaneous and stimulated emission will be almost 5×10^{26} , which shows that for getting strong lasing one

has to think apart from the thermal equilibrium. For shorter wavelength, laser, ratio of spontaneous to stimulated emission is larger, ensuring that it is more difficult to produce UV light using the principle of stimulated emission compared to the IR. Producing intense laser beam or amplification of light through stimulated emission requires higher rate of stimulated emission than spontaneous emission and self-absorption, which is only possible for $N_2 > N_1$ (as $B_{12} = B_{21}$) even though $E_2 > E_1$ (opposite to the Boltzmann statistics). It means that one will have to create the condition of *population inversion* by going beyond the thermal equilibrium to increase the process of stimulated emission for getting intense laser light.

If a collimated beam of monochromatic light having initial intensity I_0 passes through the mentioned active medium, after traveling length x , intensity of the beam is given by $I(x) = I_0 e^{-\alpha x}$, where α is the absorption coefficient of the medium, which is proportional to the difference of N_1 and N_2 . In the case of thermal equilibrium $N_1 \gg N_2$ the irradiance of the beam will decrease with the length of propagation through the medium. However, in the case of population inversion, $(N_2 > N_1) - \alpha$, will be positive and the irradiance of the beam will increase exponentially as $I(x) = I_0 e^{kx}$, where k is the gain coefficient of the medium and may be given by $k = (nN_d h\nu_{21} B_{21})/c$, where N_d is $N_2 - N_1$, c is speed of light, and n is refractive index of the medium.

1.1.4

Multilevel Systems for Attaining Condition of Population Inversion

Considering the case of two energy level system under optical pumping, we have already discussed that $B_{12} = B_{21}$, which means that even with very strong pumping, population distribution in upper and lower levels can only be made equal. Therefore, optical as well as any other pumping method needs either three or four level systems to attain population inversion. A three level system (Figure 1.1a) irradiated by intense light of frequency ν_{02} causes pumping of large number of atoms from lowest energy level E_0 to the upper energy level E_2 . Nonradiative decay of atoms from E_2 to E_1 establishes population inversion between E_1 and E_0 (i.e., $N_1 > N_0$), which is practically possible if and only if atoms stay for longer time in the state E_1 (metastable state, i.e., have a long lifetime) and the transition from E_2 to E_1 is rapid. If these conditions are satisfied, population inversion will be achieved between E_0 and E_1 , which makes amplification of photons of energy $E_1 - E_0$ by stimulated emission. Larger width of the E_2 energy level could make possible absorption of a wider range of wavelengths to make pumping more effective, which causes increase in the rate of stimulated emission. The three level system needs very high pumping power because lower level involved in the lasing is the ground state of atom; therefore more than half of the total number of atoms have to be pumped to the state E_1 before achieving population inversion and in each of the cycle, energy used to do this is wasted. The pumping power can be greatly reduced if the lower level involved in the lasing is not ground state, which requires at least a four level system (Figure 1.1b). Pumping transfers atoms from ground state to E_3 , from where they decay rapidly into the metastable state E_2 to make N_2 larger than

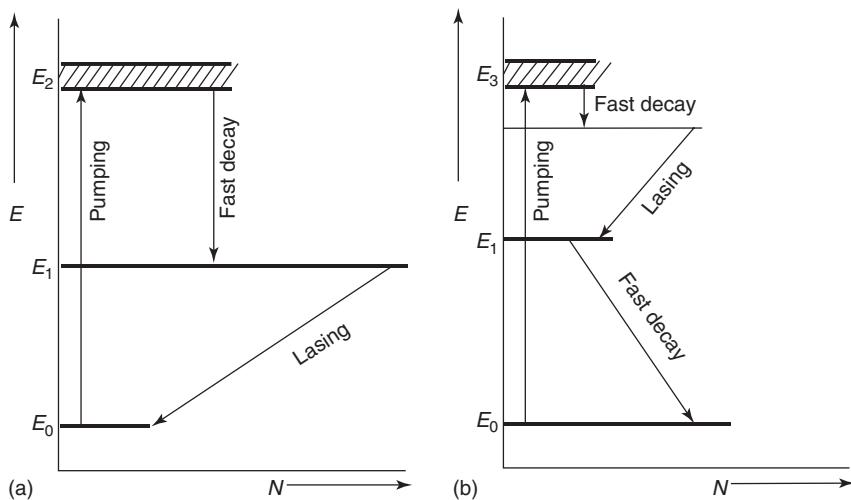


Figure 1.1 Energy level diagram for (a) three- and (b) four level laser systems.

N_1 to achieve the condition of population inversion between E_2 and E_1 at moderate pumping.

1.1.5

Threshold Gain Coefficient for Lasing

Laser beam undergoes multiple oscillations (through active medium) between pair of mirrors to achieve considerable gain before it leaves the cavity through partially reflecting mirror. Laser oscillation can only sustain in the active medium if it attains at least unit gain after a round-trip between mirrors and maintains it overcoming various losses inside the cavity. If we incorporate these losses, the effective gain coefficient reduces to $k - \Upsilon$, where Υ is the loss coefficient of the medium. If round-trip gain G were less than unity, the laser oscillation would die out, while it would grow if the G value were larger than unity. Let us consider that the laser beam of intensity I_0 passes through the active medium, homogeneously filled in the length L between the space of two mirrors M_1 and M_2 with reflectivities R_1 and R_2 , respectively. The beam of intensity I_0 initiates from the surface of M_1 and attains intensity I_1 ($I_1 = I_0 \exp(k - \Upsilon)L$) after traveling a length L to reach at the surface of M_2 . After reflection from M_2 and traveling back to M_1 , the light intensity becomes I_2 ($I'_1 = I_1 R_2$ due to reflection and $I_2 = I'_1 \exp(k - \Upsilon)L$), which finally becomes I'_2 after reflection from M_1 to complete a round-trip ($I'_2 = I_2 R_1 = I_0 R_1 R_2 \exp(2k - 2\Upsilon)L$). Waves starting from the surface of mirror M_1 and those that have completed one or more round trips are in the same phase. Now, the gain $G(I'_2/I_0)$ attained in a round-trip should be at least unity to sustain the laser oscillation inside the cavity, therefore $R_1 R_2 \exp(2k - 2\Upsilon)L = 1$ is the threshold condition, which gives a value of $\Upsilon + (2L)^{-1} \ln(R_1 R_2)^{-1}$ for threshold gain (k_{th}) coefficient.

1.1.6

Optical Resonator

An optical resonator is an arrangement of optical components, which allows a beam of light to circulate in a closed path so that it retraces its own path multiple times, in order to increase the effective length of the media with the aim of large light amplification analogous to the positive feedback in electronic amplifiers. Combination of optical resonator with active medium is known as *optical oscillator*. A set of two parallel and optically flat mirrors, with one highly reflecting $M_1(R \approx 100\%)$ and another partially transmitting $M_2(R > 95\%)$, makes a simple optical oscillator as shown in Figure 1.2. Some of the pumped atoms in the excited states undergo spontaneous emission generating seed photons, which pass through the active medium and get amplified through stimulated emission. Most of the energy gets reflected from both the mirrors, passes through the active medium, and continues to get amplified until steady state level of oscillation is reached. After attaining this stage, amplification of wave amplitude within the cavity dies away and extra energy produced by stimulated emission exits as laser output from the window M_2 . The gain coefficient inside the cavity should be greater than the threshold gain coefficient (k_{th}) in order to start and maintain laser oscillation inside the cavity. Owing to the diffraction effects, it is practically difficult to maintain a perfectly collimated beam with the combination of two parallel plane mirrors, which causes significant amount of diffraction losses. Such losses could be reduced by using a combination of concave mirrors and other optics in different optical arrangements. The optical configurations, which are able to retain the light wave inside the cavity after several transversals, are known as *stable resonators*. Some of the stable resonators are shown in the Figure 1.3. Laser oscillators with different geometries have their own benefits and losses. For example, in an oscillator having assembly of two parallel mirrors, it is difficult to align them in a strictly parallel manner. A slight deviation from the parallel geometry of the laser beam causes its walk away from the cavity axis after few reflections. However, it is beneficial in the sense that a large fraction of the active medium (mode volume) is pumped in this geometry. Confocal resonators are very simple to align, although lesser fraction of the active medium is being pumped.

Every laser resonator is characterized by a quantity Q termed as *quality factor*, which is defined by $Q = (2\pi \times \text{energy stored}) / (\text{energy dissipated per cycle})$. The Q value of laser cavities lies in the range of $\sim 10^5 - 10^6$. Significance of higher Q value lies in the sense of capacity to store larger energy. In terms of line width $\Delta\nu$,

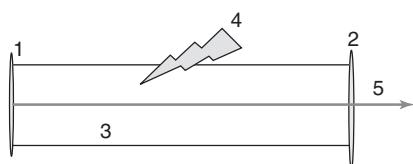


Figure 1.2 Basic geometry of laser cavity: (1) 100% and (2) 95–98% reflecting mirrors, (3) active medium, (4) pumping source, and (5) laser output.

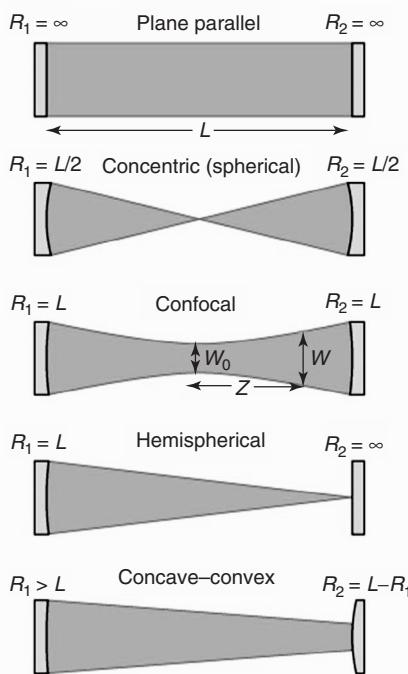


Figure 1.3 Different geometries of stable optical resonators.

and frequency ν , the quality factor can be defined as $Q = \nu / \Delta\nu$. A higher Q value associates with lower relative line width and vice versa.

A resonator that cannot maintain laser beam parallel to its axis is termed as *unstable resonator*. Such resonators suffer from high losses, but can make efficient use of the mode volume and have easy way of adjustment for the output coupling of the laser. Figure 1.4 illustrates an unstable resonator having active medium between the mirrors. Output power of the laser and inner diameter of the annular-shaped beam can be easily adjusted by varying the distance between the two reflecting

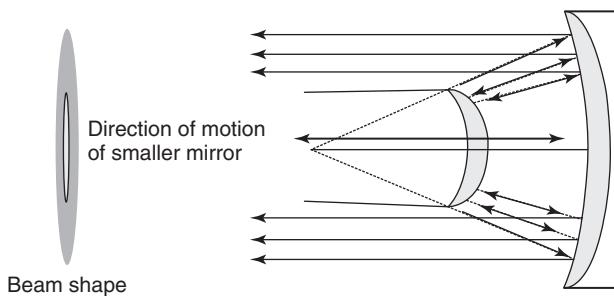


Figure 1.4 A sketch of unstable optical resonator with annular beam shape.

mirrors. Resonators having low irradiation volume or unstable cavities require active media with a large gain coefficients, such as CO₂ gas.

1.1.7 Laser Modes

The output of laser beam actually consists of a number of closely spaced spectral lines of different frequencies in a broad frequency range. The discrete spectral components are termed as *laser modes*, and coverage range is the line width of the atomic transition responsible for the laser output. Laser modes are categorized into axial and transverse modes.

- 1) **Axial modes:** Let $d\phi = 2\pi/\lambda^*(2L)$ be the phase change in the laser wave after a round-trip in the cavity. In order to sustain laser oscillation inside the cavity, the phase change should be an integral multiple of 2π , that is, $2\pi/\lambda^*(2L) = 2p\pi$. In terms of frequency, this expression transforms to $\nu = pc/2L$; therefore separation between two adjacent p and $p + 1$ modes is given by $\Delta\nu = c/2L$ (Figure 1.5). In the particular case of Nd: YAG (neodymium-ion-doped yttrium aluminum garnet) laser, $\lambda = 1064$ nm and $L = 25$ cm, $p = 2L/\lambda \approx 47 \times 10^4$ axial mode exists inside the laser cavity. If line width of the laser at 1064 nm is about $\Delta\nu = 1$ GHz, then only $\Delta\nu/\Delta\nu \approx 1$ axial mode oscillates in the cavity, while others die out. The axial modes are constructed by the light waves moving exactly parallel to the cavity axis. Light incident on a mirror and that reflected from that mirror construct a standing wave similar to a string bounded at both the ends. All the axial modes are due to the propagation of plane waves along the line joining centers of two reflecting mirrors.
- 2) **Transverse modes:** Unlike the plane waves propagating along the axis of the cavity in axial modes, there are some other waves traveling out of the axis that are not able to repeat their own path termed as transverse electromagnetic

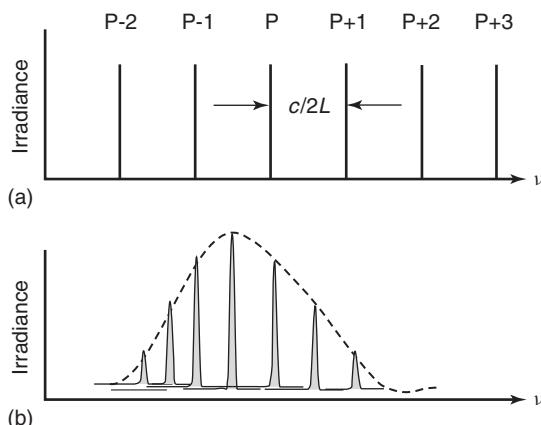


Figure 1.5 Axial laser modes (a) a simple illustration and (b) inside the laser line width, which shows that the mode at the center of the line has maximum intensity.

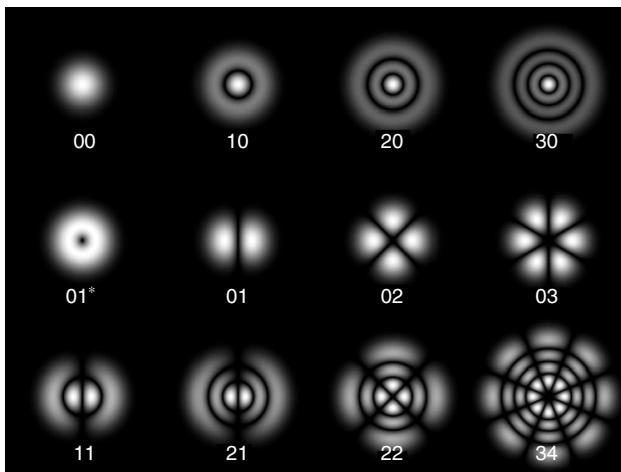


Figure 1.6 Various TEM modes of the laser.

(TEM) modes. These modes can be practically seen in the form of pattern when the laser beam falls on any surface. These modes are assigned by two integers p and q in the form of TEM_{pq} , where p and q are the number of minima in the horizontal and vertical directions, respectively, in the pattern of the laser beam. TEM_{00} means that there is no minima in the beam spot, and this is known as *uniphase mode*. On the contrary, TEM_{01} shows that there is no minima in the horizontal scanning and one minima in vertical. Laser beam spots on the screen with several TEM modes are displayed in Figure 1.6.

1.2

Types of Laser and Their Operations

Depending on the nature of the active media, lasers are classified into three main categories, namely, solid, liquid, and gas. Scientists and researchers have investigated a wide variety of laser materials as active media in each category since 1958, when lasing action was observed in ruby crystal. It is inconvenient to discuss all lasers having these materials as active media. Here, representative active medium for each of the categories and their operating principle with energy level diagram is discussed.

1.2.1

Solid Laser

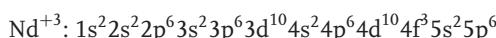
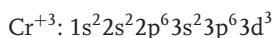
1.2.1.1 Doped Insulator Laser

Solid state lasers have active media obtained by embedding transition metals (Ti^{+3} , Cr^{+3} , V^{+2} , Co^{+2} , Ni^{+2} , Fe^{+2} , etc.), rare earth ions (Ce^{+3} , Pr^{+3} , Nd^{+3} , Pm^{+3} , Sm^{+2} , $\text{Eu}^{+2,+3}$, Tb^{+3} , Dy^{+3} , Ho^{+3} , Er^{+3} , Yb^{+3} , etc.), and actinides

such as U^{+3} into insulating host lattices. Energy levels of active ions are only responsible for lasing actions, while physical properties such as thermal conductivity and thermal expansivity of the host material are important in determining the efficiency of the laser operation. Arrangement of host atoms around the doped ion modifies its energy levels. Different lasing wavelength in the active media is obtained by doping of different host materials with same active ion. $Y_3Al_5O_{12}$, $YAlO_3$, $Y_3Ga_5O_{12}$, $Y_3Fe_5O_{12}$, $YLiF_4$, Y_2SiO_5 , $Y_3Sc_2Al_3O_{12}$, $Y_3Sc_2Ga_3O_{12}$, $Ti:Al_2O_3$, $MgAl_2O_4$ (spinel), $CaY_4[SiO_4]_3O$, $CaWO_4$ (Scheelite), $Cr:Al_2O_3$, NdP_5O_4 , $NdAl_3[BO_3]_4$, $LiNdP_4O_{12}$, $Nd:LaMgAl_{11}O_{19}$, $LaMgAl_{11}O_{19}$, $LiCaAlF_6$, $La_3Ga_5SiO_4$, $Gd_3Sc_2Al_3O_{12}$, $Gd_3Ga_5O_{12}$, $Na_3Ga_2Li_3F_{12}$, Mg_2SiO_4 (Forsterite), CaF_2 , Al_2BeO_4 (Alexandrite), and so on, are some of the important hosts. Active atom replaces an atom in the host crystal lattice. Nd:YAG is one of the best lasing material and is representative of solid state lasing materials.

1.2.1.1.1 Dopant Energy Levels in the Host Matrices

Transition metal and rare earth ions have partially filled and unfilled 3d and 4f subshells, respectively. For example, the electronic configurations of trivalent Cr and Nd ions are as follows:



There are unshielded partially filled d electrons in the transition metal ions, while partially filled 4f electrons of the rare earth ions are shielded by 5p and 5s sub shells. Owing to the electronic shielding of inner subshells in rare earth ions, crystal field effect on the energy levels of transition metal ions are pronounced as compared to that on energy levels of rare earth ions. When one of these ions is doped into a host lattice, three main types of interactions occur: (i) columbic interaction between electrons in the unfilled shell, (ii) the crystal field, and (iii) spin-orbit interactions. The columbic interaction between electrons causes splitting of energy levels of a single electron configuration into several levels denoted by pair of values of L and S (L and S are vector sum of angular, l , and spin, s , momenta, respectively, of electrons). Crystal field splitting dominates for transition metal, while spin-orbit interaction is the major contributor for rare earth ions in the modification of energy level of isolated host atom. The energy level diagram for $Cr^{+3}:Al_2O_3$ (ruby) and $Nd^{+3}:YAG$ are displayed in Figure 1.7.

1.2.1.1.2 Pumping Techniques in Solid State Lasers

Pumping of electrons from the ground state to the excited state to achieve population inversion condition is an essential requirement for lasing. Optical pumping is the best and most efficient pumping method for solid state active media due to their broad optical absorption bands. A significant fraction of incident optical energy can be easily used for the pumping of ground state electrons using pulsed as well as continuous light sources. Excess light energy raises temperature of the laser materials; therefore pulsed light sources are more suitable for dissipation of heat

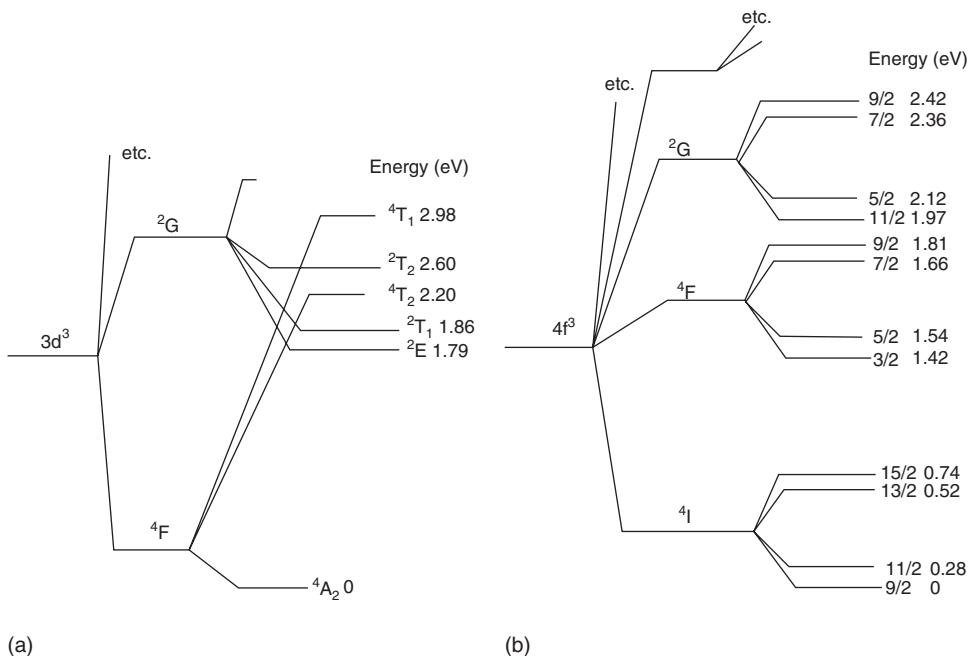


Figure 1.7 Energy level diagrams for doped insulator lasers: (a) ruby and (b) Nd:YAG lasers.

through circulating water jackets. Low-pressure quartz/glass-sealed krypton/xenon lamps are mostly used for pulsed pumping light sources, while tungsten halogen lamps and high-pressure mercury discharge lamps are utilized for continuous optical pumping. An inductive, capacitive, and resistive (LCR) circuit and trigger unit as shown in Figure 1.8 is basically used for operating the flashtube. The detail circuit diagram of the power supply is presented in Ref. [3]. High-voltage pulse of the trigger coil ionizes some gas in the tube and makes it conductive. This causes rapid discharge of the capacitor through the tube and generation of intense optical radiation for almost few milliseconds. A small inductor in the series protects damage of the tube due to high capacitor discharge current. Light source and active medium should be arranged in such a way that maximum pumping radiation falls on the active medium. Active media in solid state lasers are cylindrical and rod shaped with few millimeters diameter and few centimeter lengths. Several arrangements of cylindrical flash lamp and rod-shaped active media are used for optical pumping to get laser radiation. The flash lamp and active medium assembly are placed inside gold-plated reflectors of circular or elliptical cross section. In the first practical operating laser, ruby rod was pumped by helical flash lamp inside the cylindrical reflecting cavity. Such arrangement has significant uniformity of irradiation inside the rod but exhibits poor optical coupling. Side-by-side arrangement of flash lamp and laser rod inside the cylindrical gold-plated reflector or wrapping both together with a metal foil are simpler approaches having good optical coupling

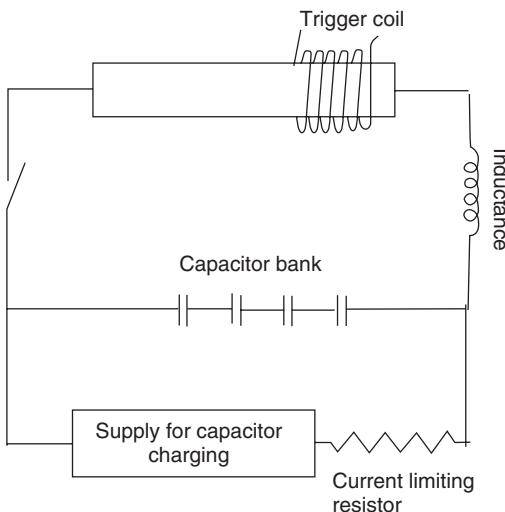


Figure 1.8 Trigger unit and LCR circuit diagram for solid state lasers.

but poor uniformity of irradiation. An elliptical reflector having flash lamp at one focus and laser rod at the other focus is the most popular and best way of optical pumping in the solid state lasers. Light radiation leaving from the first focus gets focused close to the axis of laser rod placed at the second focus to make uniform energy distribution. Combination of a number of elliptical reflectors having laser rod at the common foci and several flash lamps at the other foci is used for better optical pumping with more uniform energy distribution. Various geometries for the arrangement of laser rod and flash lamps are illustrated in Figure 1.9. Nd:YAG laser is widely used in the processing of materials and various characterizations. Here we discuss energy level diagram and operating principles of Nd:YAG lasers.

1.2.1.1.3 Nd:YAG Laser Construction and Operation

The schematic diagram of Nd:YAG laser head as shown in Figure 1.10, consists of oscillator section, rear mirror, quarter-wave plate, Pockel cells, polarizer, pump chambers, injection seeder, output coupler, D-Lok monitor, fold mirrors, amplifier section, harmonic generator (HG), temperature controller, dichroic mirrors, and Beam Lock pointing sensor. It may have single or multipump chambers, and each chamber consists of single or multiple flash lamps depending on the power of laser. The laser head end panel contains coolant, output connector, coolant input connector, neutral/ground connector, control cable connector, high-voltage connector, Q-switch input connector, and nitrogen purge input connector. The HGs have potassium di-hydrogen phosphate (KDP) and beta barium borate (BBO) crystals for frequency doubling and tripling, respectively. It can be operated in long pulse and Q-switch modes. Long pulse mode has light pulses of almost $200\text{ }\mu\text{s}$ duration and separated from each other by $2\text{--}4\text{ }\mu\text{s}$. The total energy of the pulse

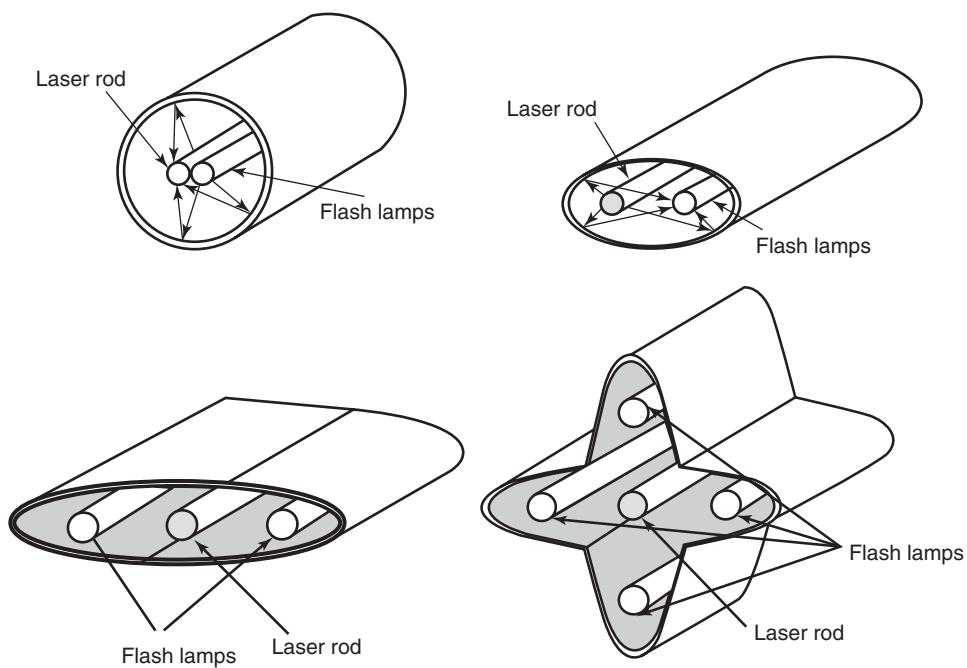


Figure 1.9 Different geometries for the arrangement of flash lamp and laser rod in solid states lasers.

train is similar to that of a single Q-switched pulse. During Q-switched operation, the pulse width is less than 10 ns and the peak optical power is tens of megawatts.

The properties of Nd:YAG are the most widely studied and best understood of all solid state laser media. Its energy level diagram, optical arrangements for Q-switching and stable and unstable resonators are depicted in Figure 1.11. The active medium is triply ionized neodymium, which is optically pumped by a flash lamp whose output matches principle absorption bands in the red and near infrared (NIR). Excited electrons quickly drop to the $F_{3/2}$ level, the upper level of the lasing transition, where they remain for a relatively longer time ($\sim 230 \mu\text{s}$). The strongest transition is $F_{3/2} \rightarrow I_{11/2}$, emitting a photon in NIR region (1064 nm). Electrons in the $I_{11/2}$ state quickly relax to the ground state, which makes its population low. Therefore, it is easy to build up a population inversion for this pair of states with high emission cross section and low lasing threshold at room temperature. There are also some other competing transitions at 1319, 1338, and 946 nm from the same upper state, but having lower gain and a higher threshold than the 1064 nm wavelength. In normal operation, these factors and wavelength-selective optics limit oscillation to 1064 nm. A laser comprising just an active medium and resonator will emit a pulse of laser light each time the flash lamp fires. However, the pulse duration will be long, about the same as the flash lamp and its peak power will be low. When a Q-switch is added to the resonator to shorten the pulse, output peak power is raised dramatically. Owing to the long lifetime of $F_{3/2}$, a large

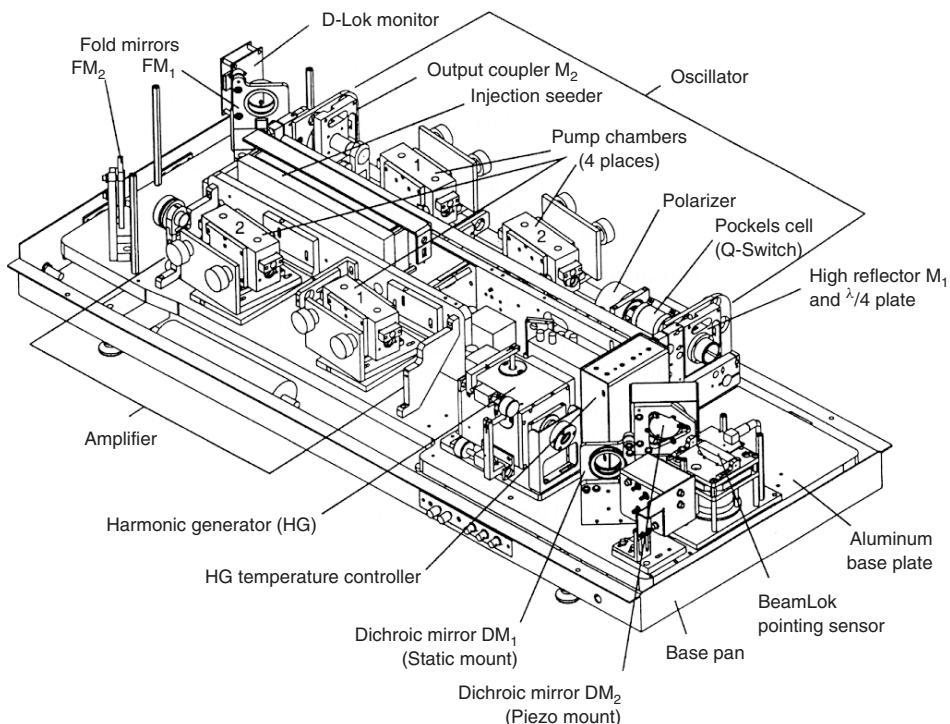


Figure 1.10 Assembly of various components in the head of an Nd:YAG laser system with four pump chambers.

population of excited neodymium ions can build up in the YAG rod in a way similar to which a capacitor stores electrical energy. When oscillation is prevented for some time to build up high level of population inversion by electro-optical Q-switching and after that if the stored energy gets quickly released, the laser will emit a short pulse of high-intensity radiation.

1.2.1.2 Semiconductor Laser

Semiconductor lasers also known as *quantum well lasers* are smallest, cheapest, can be produced in mass, and are easily scalable. They are basically p-n junction diode, which produces light of certain wavelength by recombination of charge carrier when forward biased, very similar to the light-emitting diodes (LEDs). LEDs possess spontaneous emission, while laser diodes emit radiation by stimulated emission. Operational current should be higher than the threshold value in order to attain the condition of population inversion. The active medium in a semiconductor diode laser is in the form of junction region of 2 two-dimensional layers. No external mirror is required for optical feedback in order to sustain laser oscillation. The reflectivity due to the refractive index differences between two layers or total internal reflection to the active media is sufficient for this purpose. The diodes end faces are cleaved, and parallelism of reflecting surfaces is assured. Junction

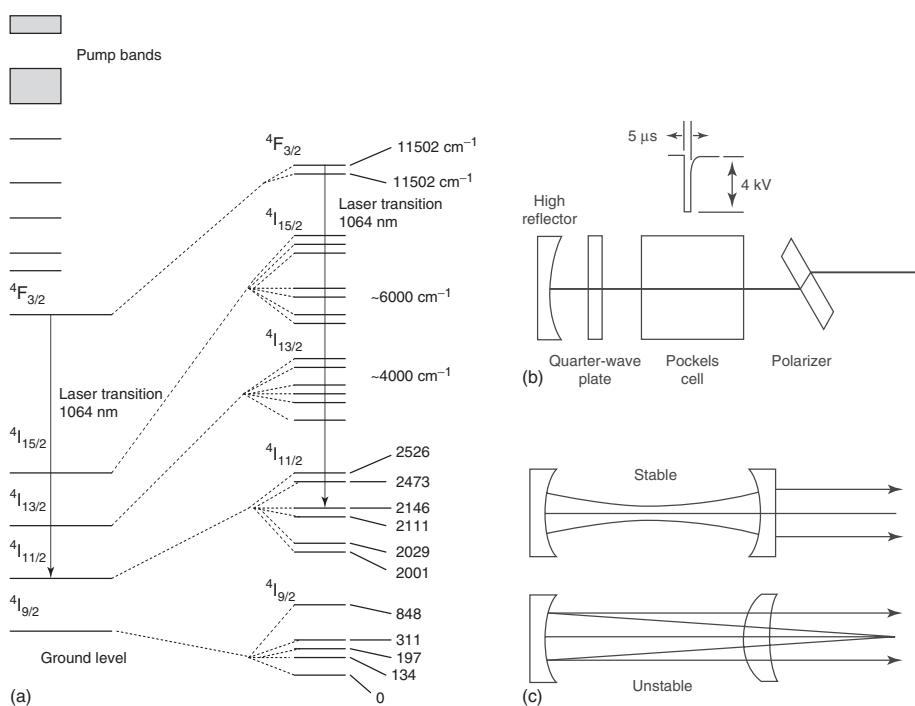


Figure 1.11 (a) Energy level diagram for the transition of Nd:YAG laser (b) The Q-switch comprises a polarizer, a quarter-wave plate, high quality reflector, and pockels cell, and (c) stable and unstable resonator configurations.

made from a single type of semiconductor material is known as *homojunction*, while that obtained from two different semiconductors is termed as *heterojunction*. Semiconductors of p and n type with high carrier density are brought together for constructing p-n junction with very thin ($\approx 1\text{ }\mu\text{m}$) depletion layer. Figure 1.12 illustrates GaAs homojunction semiconductor diode laser. Lasing occurs in the confined narrow region, and optical feedback is done by reflections between cleaved end faces. For GaAs $n = 3.6$, therefore reflectivity R from the material-air interface is $R = (n - 1)^2/(n + 1)^2 = 0.32$, which is small but sufficient for lasing.

When the operating current is small, the population inversion built compensates losses in the system and no lasing action is done. Increase of the current above a critical value named as *threshold current* commences lasing action, and the intensity of laser radiation increases rapidly with further increase in the operating current. Semiconductor lasers have large divergence compared to any other laser systems, which is due to their small cross section of active region. Actual dimension (d) of the active medium is of the order of light wavelength (λ), which causes diffraction and hence divergence by an angle of $\theta \approx \lambda/d$. Homojunction semiconductor lasers have some disadvantages over heterojunction lasers. Both of the laser systems should have confinement of injected electrons and emitted light in the junction region in order to initiate efficient stimulated emission process. In

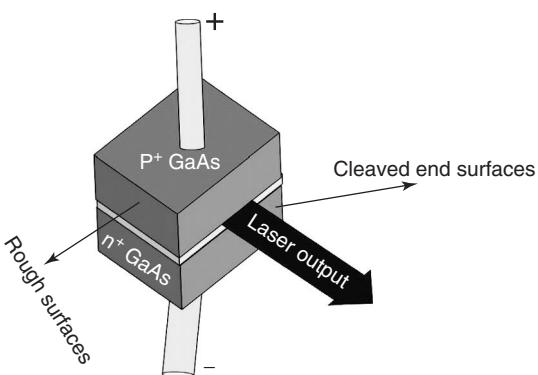


Figure 1.12 Basic geometry of semiconductor laser system.

the homojunction laser, confinement of light is the consequence of the presence of hole and electrons close to the junction. Homojunction lasers operate under such confinement mechanism but have high threshold current density and low efficiency. Electrons have to travel different distances before they recombine with the holes. In contrast, heterojunction lasers exhibit much higher lasing efficiency and low threshold current density compared to their homojunction counterparts. Another difficulty with homojunction laser is to prevent the radiation from spreading out sideways from the gain region, which causes loss instead of gain. Therefore they can only be used in the pulsed mode.

Heterojunction lasers are constructed by sandwiching a thin layer of GaAs between two layers of ternary semiconductor compound $\text{Ga}_{1-x}\text{Al}_x\text{As}$ with comparatively lower refractive indices and higher band gap energy. Lower refractive indices of surrounding layers causes confinement of laser radiation inside the active medium by the mechanism of total internal reflection, which makes laser oscillation to sustain in the medium. Higher band gap energy of the surrounding media creates potential barrier to prevent charge carriers to diffuse from the junction region, that is, provides a way for the confinement of charge carriers in the junction region, which enhances the condition of population inversion and hence stimulated emission. The electrical circuit for pumping the semiconductor diode lasers is similar to the doped insulator lasers.

1.2.2

Gas Laser

Gas lasers are widely available in almost all power (milliwatts to megawatts) and wavelengths (UV-IR) and can be operated in pulsed and continuous modes. Based on the nature of active media, there are three types of gas lasers viz atomic, ionic, and molecular. Most of the gas lasers are pumped by electrical discharge. Electrons in the discharge tube are accelerated by electric field between the electrodes. These accelerated electrons collide with atoms, ions, or molecules in the active media and

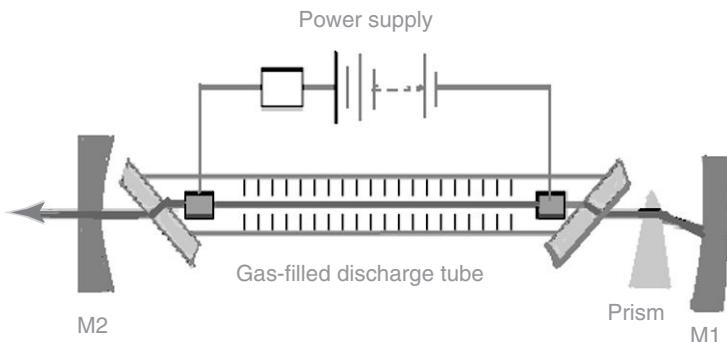


Figure 1.13 Construction of gas laser system (argon ion laser with prism-based wavelength tuning).

induce transition to higher energy levels to achieve the condition of population inversion and stimulated emission. An example of gas laser system is shown in Figure 1.12.

1.2.2.1 Atomic Gas Laser; He:Ne Laser

He–Ne laser is the simplest and representative of atomic gas lasers. The active medium is a 10 : 1 mixture of He and Ne gases filled in a narrow tube of few millimeter diameters and 0.1–1 m long at a pressure of about 10 Torr. Discharge tube and circuit are very similar as shown in Figure 1.13. A resistor box is used in series with power supply in order to limit the discharge current because tube resistance falls too low once discharge is initiated. Energy levels of Ne atom are directly involved in the laser transitions, and He atoms provide an efficient excitation mechanism to the Ne atoms. Helium atoms from their ground state 1^1S , are pumped to the excited atomic states 2^1S and 2^3S by impact with accelerated electrons in the discharge tube. Neon atoms have 3s and 2s atomic states, which are closer to the 2^1S and 2^3S states of the helium atoms, respectively. Collision between excited helium atoms in the 2^1S and 2^3S states and neon atoms in the ground states reinforce transfer of energy from helium to neon atoms. Helium atoms in the 2^1S and 2^3S states excite neon atoms from ground state to the 3s and 2s states, respectively, and return to the ground state. Excited states 3s and 2s of Ne atom have longer life times as compared to its lower (3p and 2p states), therefore they serve as metastable states and are used in achieving the condition of population inversion between s and p states. Transitions $3s \rightarrow 3p$, $3s \rightarrow 2p$, and $2s \rightarrow 2p$ of neon atoms are consequences of lasing at $3.39\text{ }\mu\text{m}$, 632.8 nm , and $1.15\text{ }\mu\text{m}$ wavelengths, respectively. Lifetimes of 3p and 2p atomic states are shorter; therefore Ne atoms from these states rapidly decay to the 1s state by nonradiative transitions. Neon atoms in the 1s state go to the ground state after losing energy through collision with the wall of the tube. The energy level diagram of the He–Ne laser is displayed in Figure 1.14. Another important atomic laser is copper vapor laser, but it is beyond the scope of this book.

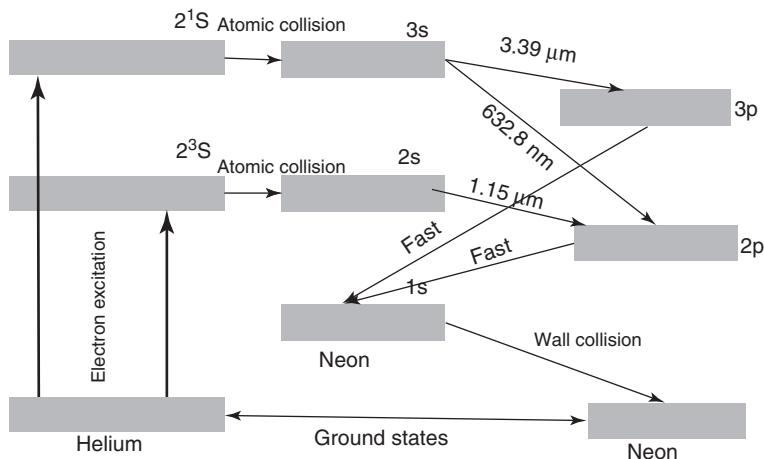


Figure 1.14 Energy level diagram for He-Ne laser system.

1.2.2.2 Ion Laser: Argon Ion Laser

1.2.2.2.1 Physical Construction

Argon ion laser is one of the widely used ion gas lasers, which typically generates several watts power of a green or blue output beam with high beam quality. The core component of an argon ion laser is an argon-filled tube made of ceramics, for example, beryllium oxide, in which an intense electrical discharge between two hollow electrodes generates a plasma with a high density of argon (Ar^+) ions. A solenoid around the tube (not shown in Figure 1.13) is used for generating a magnetic field, which increases output power of the beam by magnetic confinement of the plasma near the tube axis.

A typical device, containing a tube with a length of the order of 1 m, can generate 2.5–5 W of output power of laser beam in the green spectral region at 514.5 nm, using several tens of kilowatts of electric power. The dissipated heat is removed with a chilled water flow around the tube. The laser can be switched to other wavelengths such as 457.9 nm (blue), 488.0 nm (blue-green), or 351 nm (ultraviolet) by rotating the intracavity prism. The highest output power is achieved on the standard 514.5 nm line. Without an intracavity prism, argon ion lasers have a tendency for multiline operation with simultaneous output at various wavelengths.

1.2.2.2.2 Working of Ar Ion Laser

The argon ion laser is a four level laser, which facilitates to achieve population inversion and low threshold for lasing. The neutral argon atoms filled between two hollow electrodes inside the plasma tube (Figure 1.13) are pumped to the $4p$ energy level by two steps of collisions with electrons in the plasma. The first step ionizes atoms to make ions in the $3p$ (E_1) state, and the second one excites these ions from the ground state E_1 either directly to the $4p^4$ levels (E_3) or to the $4p^2$ levels (E_4), from which it cascades almost immediately to the $4p^2$ (E_3). The $4p$ ions

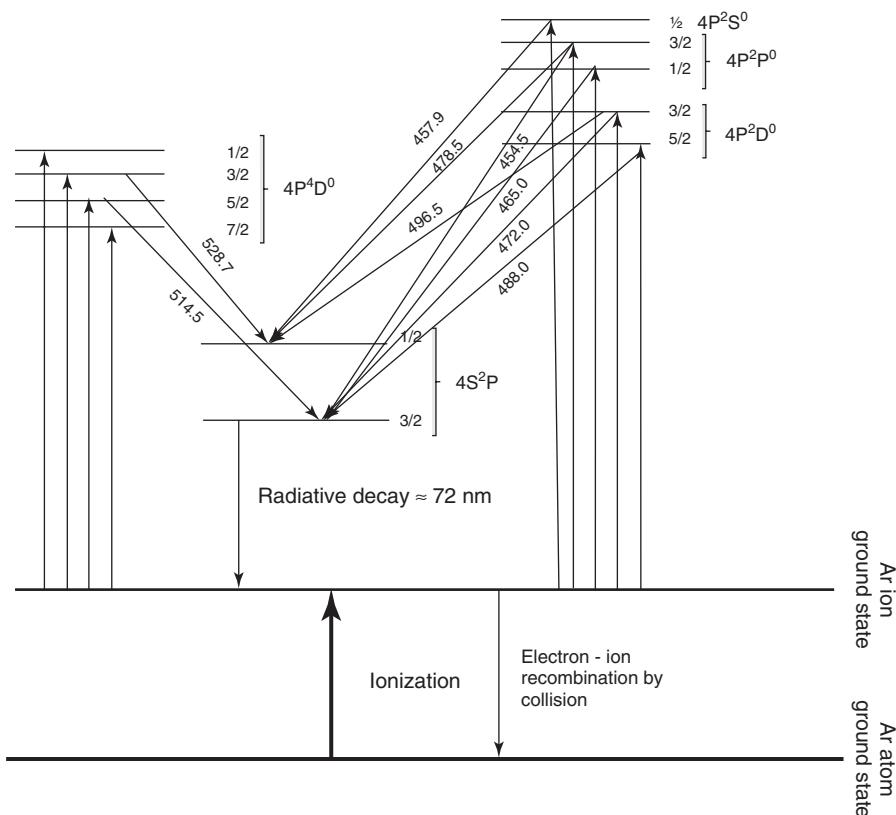


Figure 1.15 Energy level diagram for argon ion laser system.

eventually decay to $4s$ levels (E_2), either spontaneously or when stimulated to do so by a photon of appropriate energy. The wavelength of the photon depends on the specific energy levels involved and lies in between 400 and 600 nm. The ion decays spontaneously from $4s$ to the ground state, emitting a deep ultraviolet photon of about 72 nm. There are many competing emission bands as shown in Figure 1.15. These can be preferentially selected using a prism in front of one of the end mirrors. This prism selects a specific wavelength to send it back through the cavity to stimulate identical emissions, which stimulates more and more emissions and make regenerative process. This facilitates laser to operate at a single wavelength. Removal of the prism allows for broadband operation, that is, several wavelengths are kept rather than keeping only a particular wavelength. The mirrors reflect a number of lines within a maximum range of about 70 nm. Energy level diagram showing various transitions of Ar ion laser is illustrated in Figure 1.15.

1.2.2.3 Molecular Laser

Unlike isolated atoms and ions in atomic and ionic lasers, molecules have wide energy bands instead of discrete energy levels. They have electronic, vibrational,

and rotational energy levels. Each electronic energy level has a large number of vibrational levels assigned as V , and each vibrational level has a number of rotational levels assigned as J . Energy separation between electronic energy levels lies in the UV and visible spectral ranges, while those of vibrational–rotational (separations between two rotational levels of the same vibrational level or a rotational level of one vibrational level to a rotational level from other lower vibrational level) levels, in the NIR and far-IR regions. Therefore, most of the molecular lasers operate in the NIR or far-IR regions.

1.2.2.3.1 Carbon Dioxide (CO_2) Laser

Carbon dioxide is the most efficient molecular gas laser material that exhibits for a high power and high efficiency gas laser at infrared wavelength. It offers maximum industrial applications including cutting, drilling, welding, and so on. It is widely used in the laser pyrolysis method of nanomaterials processing. Carbon dioxide is a symmetric molecule ($\text{O}=\text{C}=\text{O}$) having three (i) symmetric stretching [$i00$], (ii) bending [$0j0$], and (iii) antisymmetric stretching [$00k$] modes of vibrations (inset of Figure 1.16), where i , j , and k are integers. For example, energy level [002] of molecules represents that it is in the pure asymmetric stretching mode with 2 units of energy. Very similar to the role of He in He–Ne laser, N₂ is used as intermediately in CO₂ lasers. The first, $V = 1$, vibrational level of N₂ molecule lies close to the (001) vibrational level of CO₂ molecules. The energy difference between vibrational levels of N₂ and CO₂ in CO₂ laser is much smaller (0.3 eV) as compared to the difference between the energy levels of He and Ne (20 eV) in He–Ne laser; therefore comparatively larger number of electrons in the discharge tube of CO₂ laser having energies higher than 0.3 eV are present. In addition to

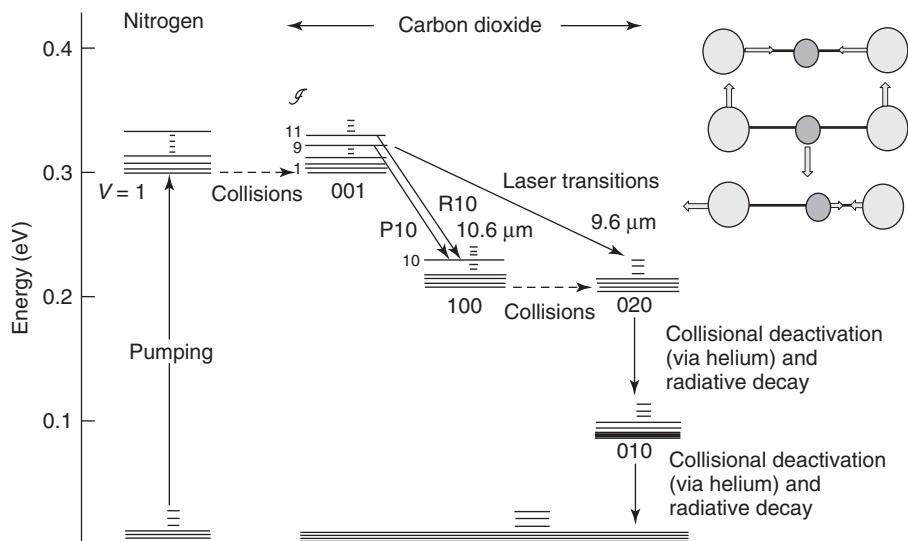


Figure 1.16 (a) Absorption, emission curves, and (b) energy level diagram of dye laser system.

this, $V = 1$ state of N_2 is metastable, which provides longer time for the collision between excited N_2 molecules and the ground state CO_2 molecules to excite them to (001) state. These two favorable conditions make it easy to attain high level of population inversion between 001 and 100, and 020 vibrational states of CO_2 .

Transitions between 001 initial level to 100 and 020 final vibrational states make stimulated emissions of several IR radiations between 9.2 and 10.6 μm wavelengths. Helium gas is also mixed in the gas mixture in order to increase efficiency of lasing. Helium helps in transporting waste heats to the tube wall and de-exciting (100) and (020) energy levels by collision process. The amounts of N_2 , CO_2 , and He in CO_2 laser depends on the type and application of system, but usually, the amount of nitrogen and CO_2 molecules are comparable, while helium concentration is higher than either. Low pressure (~ 10 Torr) is generally used for CW lasers, while quite higher pressure is used for high-energy and pulsed laser applications. Depending on power level and beam quality of CO_2 lasers, various internal structures are being used, and they are called *sealed tube laser*, *gas flow laser*, transversally excited atmospheric (TEA) laser, and *gas dynamic laser*. Detailed discussions of these are beyond the scope of the book, although interested readers may consult Ref. [4].

In the far-IR region of 10 μm wavelength, the usual optical material has large absorbance and therefore cannot be used as windows and reflecting mirrors in the cavity. Materials such as Ge, GaAs, ZnS, ZnSe, and some alkali halides having transparency in the IR region are used.

1.2.2.3.2 Nitrogen Laser

Lasing transition in N_2 laser takes place between two electronic energy levels, therefore this laser operates in the ultraviolet region at 337 nm wavelength. Here, upper electronic level has a shorter lifetime compared to the lower one, hence CW operation cannot be achieved, but pulsed operation with narrow pulse width is possible. The pulse width is narrow because as soon as lasing starts, population of the lower state increases, while that at upper state decreases and rapidly a state at which no lasing is possible is rapidly achieved. Such a laser system is known as *self-terminating*.

1.2.2.3.3 Excimer Lasers

Excimers are molecules such as ArF , KrF , XeCl , and so on, that have repulsive or dissociating ground states and are stable in their first excited state. Usually, there are less number of molecules in the ground state; therefore direct pumping from ground state is not possible. Molecules directly form in the first excited electronic state by the combination of energetic halide and rare gas ions. Condition of population inversion can be easily achieved because the number of molecules in the ground state is too low as compared to that in the excited state. Lasing action is done by transition from bound excited electronic state to the dissociative ground state. Population in the ground state always remains low because molecules dissociate into atoms at this point. Usually a mixture of halide such as F_2 and rare gas such as Ar is filled into the discharge tube. Electrons in the discharge tube dissociate and ionize halide molecules and create negative halide ions. Positive

Ar^+ and negative F^- ions react to produce ArF^* molecules in the first excited bound state, followed by their transition to the repulsive ground state to commence lasing action. Various excimer lasers are developed in the wavelength range of 120–500 nm with 20–15% efficiency and up to 1 J peak and 200 W average powers. These lasers are widely used in materials processing and characterizations as well as for the pumping of dye lasers.

1.2.3

Liquid Laser

Liquids are more homogeneous as compared to solids and have larger density of active atoms as compared to the gasses. In addition to these, they do not offer any fabrication difficulties, offer simple circulation ways for transportation of heat from cavity, and can be easily replaced. Organic dyes such DCM (4-dicyanomethylene-2-methyl-6-p-dimethylaminostyryl-4H-pyran), rhodamine, styryl, LDS, coumarin, stilbene, and so on, dissolved in appropriate solvents act as gain media. When the solution of dye molecules is optically excited by a wavelength of radiation with good absorption coefficient, it emits radiation of longer wavelength, known as *fluorescence*. The energy difference between absorbed and emitted photons is mostly used by nonradiative transitions and creates heat in the system. The broader fluorescence band in dye/liquid lasers facilitates them with the unique feature of wavelength tuning. Organic dye lasers, as tunable and coherent light sources, are becoming increasingly important in spectroscopy, holography, and in biomedical applications. A recent important application of dye lasers involves isotope separation. Here, the laser is used to selectively excite one of several isotopes, thereby inducing the desired isotope to undergo a chemical reaction more readily. The dye molecules have singlet (S_0 , S_1 , and S_2) and triplet (T_1 and T_2) group of states with fine energy levels in each of them (Figure 1.17). Singlet

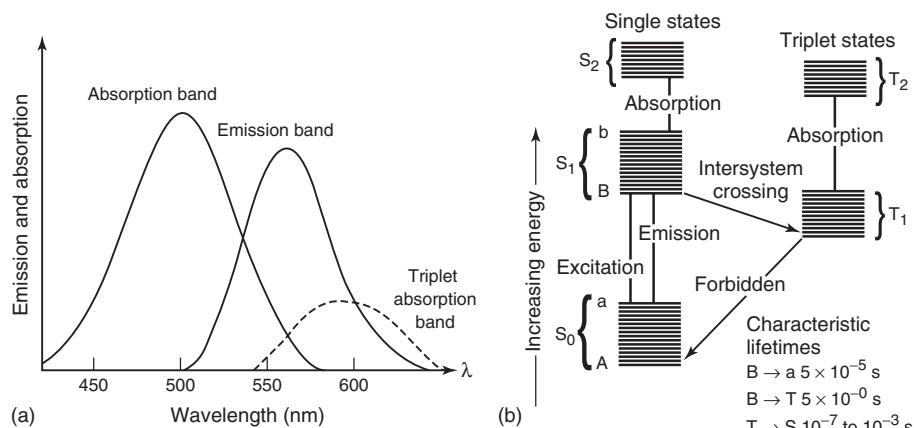


Figure 1.17 Energy level diagram for carbon dioxide (molecular) laser system. Inset shows three mode of vibration of CO_2 molecule.

and triplet states correspond to the zero and unit values of total spin momentum of electrons, respectively. According to selection rules for transitions in quantum mechanics, singlet–triplet and triplet–singlet transitions are quite less probable as compared to the transitions between two singlet or two triplet states. Optical pumping of dye molecules initially at the bottom of S_0 state transfers them to the top of S_1 state. Collisional relaxation of these molecules takes them to the bottom of S_1 state, from where they transit to the top of S_0 state with stimulated emission of radiation. Most of the states in the complex systems are usually neither pure singlet nor pure triplet. Singlet states have small contribution of triplet and vice versa. In the case of most of the dye molecules, unfortunately, T_1 state lies just below the S_1 , therefore few molecules transit from S_1 to T_1 by losing some energy through nonradiative transitions. Difference between T_1 and T_2 states is almost same as the wavelength of lasing transition, therefore emitted lasing radiation gets absorbed, which reduces laser gain and may cease the laser action. Therefore, some of the dye lasers operate in the pulsed mode with the pulse duration shorter than the time required to attain a significant population in the state T_1 . Some of the dyes also absorb laser radiation corresponding to the transition from S_1 to upper singlet transitions. Therefore, one should select the dye molecule so that energy differences between these states do not lie between the ranges of laser radiation.

1.3

Methods of Producing EUV/VUV, X-Ray Laser Beams

EUV and VUV coherent light sources, that is, EUV/VUV lasers, are in high demand in order to continue the validity of Moore's law in future, in the high-density data writing on disks, materials synthesis and characterizations, and spectroscopic investigations. For example, in photolithography for making the micro/nanopatterns on microelectronic chips, the width of the pattern is proportional to the wavelength of laser light used. Therefore, we would reach a limit for further miniaturization of electronic devices, which causes failure of the well-known Moore's law if shorter wavelength lasers would not be developed. Similarly, for data writing on the optical discs, if we have shorter wavelength lasers, larger amount of data can be written on the same size of disc. In addition to these, such sources are the future of 3D high-density data writing, microscopy at the atomic scale, crystallography, and medical sciences. Following are the methods for developing short wavelength lasers.

1.3.1

Free Electron Lasers (FEL)

In contrast to the other laser sources, free electron lasers (FELs) have an active medium that consists of a beam of free electrons, propagating at relativistic velocities in a spatially periodic magnet (undulator). Here, electrons experience the Lorentz force, execute transverse oscillations, and emit synchrotron radiation in the forward direction (Figure 1.18). We know that an accelerated charged particle

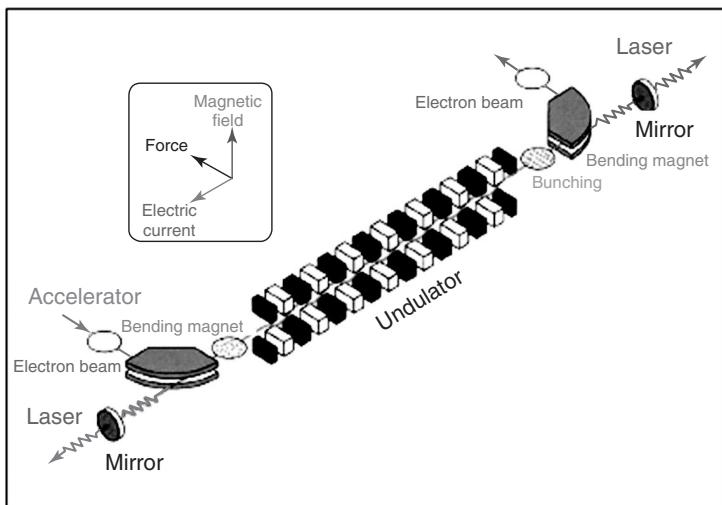


Figure 1.18 Construction geometry of free electron laser system.

moving with relativistic velocity emits radiation, which may be considered as spontaneous. The emitted spontaneous radiation interacts with the electron beam in order to stimulate laser radiation. Since electric field of light is perpendicular to the direction of its propagation, electrons cannot interact with photons unless they also have a velocity component parallel to the light electric field. Hence, electron beam is allowed to travel through wiggler magnet, which generates a spatially periodic magnetic field along the direction of propagation of electron beam. After passing through such field, electrons undergo transverse oscillations as shown in Figure 1.18. In the present case, wavelength of emitted radiation is given by the expression $\lambda = \lambda_u/2\gamma^2$ where λ_u is the undulator length and γ is the ratio of total electron energy to its rest energy. Similar to other lasers, emitted radiation oscillates between two mirrors of the laser cavity and interacts with the electron beam to enhance power of the laser radiation before leaving the cavity as laser output. Since mirrors have very low reflectivity in the X-ray region, therefore there is severe problem in achieving intense X-ray laser beam. In another mechanism known as self-amplified stimulated emission (SASE), radiation of the spontaneous emission from relativistic electron interacts with the same electron to stimulate it for stimulated emission. Following are the salient features of FELs.

- 1) **Tunability:** FELs generate coherent, high-power radiation that is widely tunable, currently ranging in wavelength from microwaves, through terahertz radiation and infrared, to the visible spectrum, to soft X-rays. They have the widest frequency range of any laser type. Wavelength of FELs can be easily tuned by varying electron energy and period and amplitude of magnetic field. More shorter wavelength can be produced by harmonic generation.

- 2) **Pulse duration:** Based on the electron beam time structure pulse duration ranges from CW to ultrashort pulsed regime (fraction of picoseconds)
- 3) **Coherence:** It is transverse and longitudinal for oscillators and coherent seed amplifiers, while only transverse for SASE
- 4) **Brilliance:** Depending on the status of the art of the electron beam technology, the FEL brilliance can be larger, in some spectral regions (in particular in VUV-X), by many orders of magnitude than the brilliance of the existing sources (lasers and synchrotron radiation).

We do not go into much details, but interested readers are referred to Refs. [6–11].

1.3.2

X-Ray Lasers

X-ray lasers provide coherent beam of electromagnetic radiation with the wavelength range from ~ 30 nm down to ~ 0.01 nm. The first X-ray laser beam was initiated in 1980 by underground nuclear explosion at Nevada test site, while the first laboratory demonstration of X-ray laser was made in 1984 in the form of Nova laser. Most of the experimental demonstration of light amplification in this spectral range has come from the high-density plasma produced by interaction of high-energy laser with the solid target. Like the above case of FEL, unavailability of good-quality of cavity mirrors has utilized the mechanism of SASE in X-ray lasers. In such case intensity of the laser beam depends on the amplifier length. Gain of the active medium depends on the mass and temperature of the ions and multiplicity and population of the upper/lower levels. Most of the X-ray lasers utilize transitions among L, M, N, and so on, shells of highly ionized atoms. The short life times of the excited states (picoseconds) require very large pumping rates in order to achieve and maintain the condition of population inversion. Laser produced plasmas (LPPs) have high electron temperatures (~ 0.1 – 1.0 keV) and densities ($\sim 10^{18}$ – 21 cm $^{-3}$), which is required for higher degree of ionization and excitation in X-ray lasers. Therefore LPPs are used as primary medium for X-ray lasers. In addition to these, LPPs have uniform density and temperature and thus can provide good media for amplification and propagation of X-rays. Of the various processes suggested, two main (i) collisional excitation and (ii) electron recombination are responsible for attaining the population inversion condition. In case of collisional excitation upper state of the lasing has lower probability of decay with dipole radiation, compared to the lower state, which creates the condition of population inversion between them. In contrast, the rate of population at any state by three-body recombination process is proportional to the forth power of the principle quantum number. The combination of preferential population of upper levels and fast radiative decay of lower states leads to an inversion amongt the $N = 2, 3, 4, \dots$ states [12].

1.3.3

EUV/VUV Lasers through Higher Harmonic Generation

Higher harmonic generation (HHG) is a nonlinear optical process used for the generation of shorter wavelength laser light from the interaction of high-intensity longer wavelength lasers source with nonlinear optical medium. The obtained new frequencies are integral multiples ($n\omega$) of the fundamental (ω) frequency of original laser light. This phenomenon was first observed in 1961 by Franken *et al.* [13] with ruby laser and quartz as nonlinear medium. The first HHG result was found in 1988, which shows that intensity of the spectra decreases with the increase of order, reaches a plateau, where the intensity remains constant for several orders, and finally ends abruptly at a position called *high harmonic cutoff*. They are portable sources of EUV/soft X-rays, synchronized with the fundamental laser and operated at the same frequencies with much shorter pulse width. These are more spatially coherent compared to X-ray lasers and cheaper than FELs. The harmonic cutoff increases linearly with increasing laser intensity up to the saturation intensity I_{sat} where harmonic generation stops. The saturation intensity can be increased by changing the atomic intensities of lighter noble gases, but these have lower conversion efficiency. HHG strongly depends on the driving laser field; therefore the produced harmonics have similar spatial and temporal coherence. Owing to the phase matching and ionization conditions required for HHG, the produced new pulses are with shorter pulse duration compared to the driving laser. Mostly, harmonics are produced in very short time frame, when phase matching condition is satisfied. Instead of shorter temporal window, they emit colinearly with the driving laser pulse and have very tight angular confinement. Gaseous media and LPP on the solid surfaces are two sources used as nonlinear active media for the generation of harmonics.

The shortest wavelength producible with the harmonic generation is given by cutoff of the plateau, which is given by the maximum energy gained by ionized electron from the light electric field. The energy of cutoff is given by $E_{\text{max}} = I_p + 3.17U_p$, where U_p is the ponderomotive potential from the laser field and I_p is the ionization potential. It is assumed that electron is initially produced by ionization into the continuum with zero initial velocity and is accelerated by the laser electric field. After half period of the laser electric field, direction of motion of electron is reversed and it is accelerated back to the parent nuclei. After attaining high kinetic energy of the order of hundreds of electron volts (depending on the intensity of laser field), when electron enters into the parent nuclei, it emits Bremsstrahlung-like radiation during a recombination process with the atoms as it returns back to the ground state. The three-step model of ionization, acceleration of electron, and its recombination with parent nuclei by the emission of EUV photons is shown in Figure 1.19.

Wide wavelength ranges of laser systems starting from hundreds of micrometers (molecular liquid lasers) to X-ray regions (FELs, X-ray lasers, and HHG lasers) with CW and pulsed lasers having various pulse width from milliseconds to attoseconds and repletion rates from single pulse to the megahertz are available nowadays.

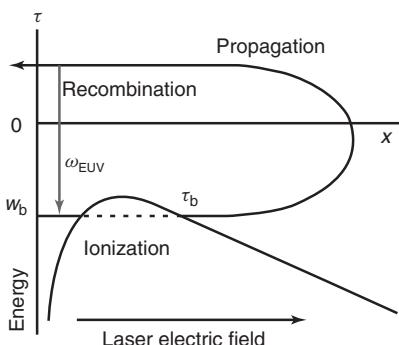


Figure 1.19 Three-step model of ionization and acceleration of electron and its recombination with parent nuclei by the emission of EUV photons.

Figure 1.20 represents the spectrum of available laser system with wavelength range and active media.

1.4 Properties of Laser Radiation

Light produced from the lasers have several valuable characteristics not shown by light obtained from other conventional light sources, which make them suitable for a variety of scientific and technological applications. Their monochromaticity, directionality, laser line width, brightness, and coherence of laser light make them highly important for various materials processing and characterization applications. These properties are discussed separately in the following subsections.

1.4.1 Monochromaticity

Theoretically, waves of light with single frequency ν of vibration or single wavelength λ is termed as *single color* or *monochromatic* light source. Practically, no source of light including laser is ideally monochromatic. *Monochromaticity* is a relative term. One source of light may be more monochromatic than others. Quantitatively, degree of monochromaticity is characterized by the spread in frequency of a line by $\Delta\nu$, line width of the light source, or corresponding spread in wavelength $\Delta\lambda$. For small value of $\Delta\lambda$, frequency spreading, $\Delta\nu$, is given as $\Delta\nu = -(c/\lambda^2)\Delta\lambda$ and $\Delta\lambda = (c/\nu^2)$. The most important property of laser is its spectacular monochromaticity. Based on the type of laser media, solid, liquid, or gas and molecular, atomic, or ions, and the type of excitation, produced laser line consists of color bands that range from broad (as dye laser $\Delta\lambda \sim 200$) to narrow (for gas discharge lines, $\Delta\lambda \sim 0.01$ nm). Utilizing suitable filters one can get the monochromaticity as good as a single line of lasing transition. But such a single line also contains a set of closely spaced lines of discrete frequencies, known as *laser*

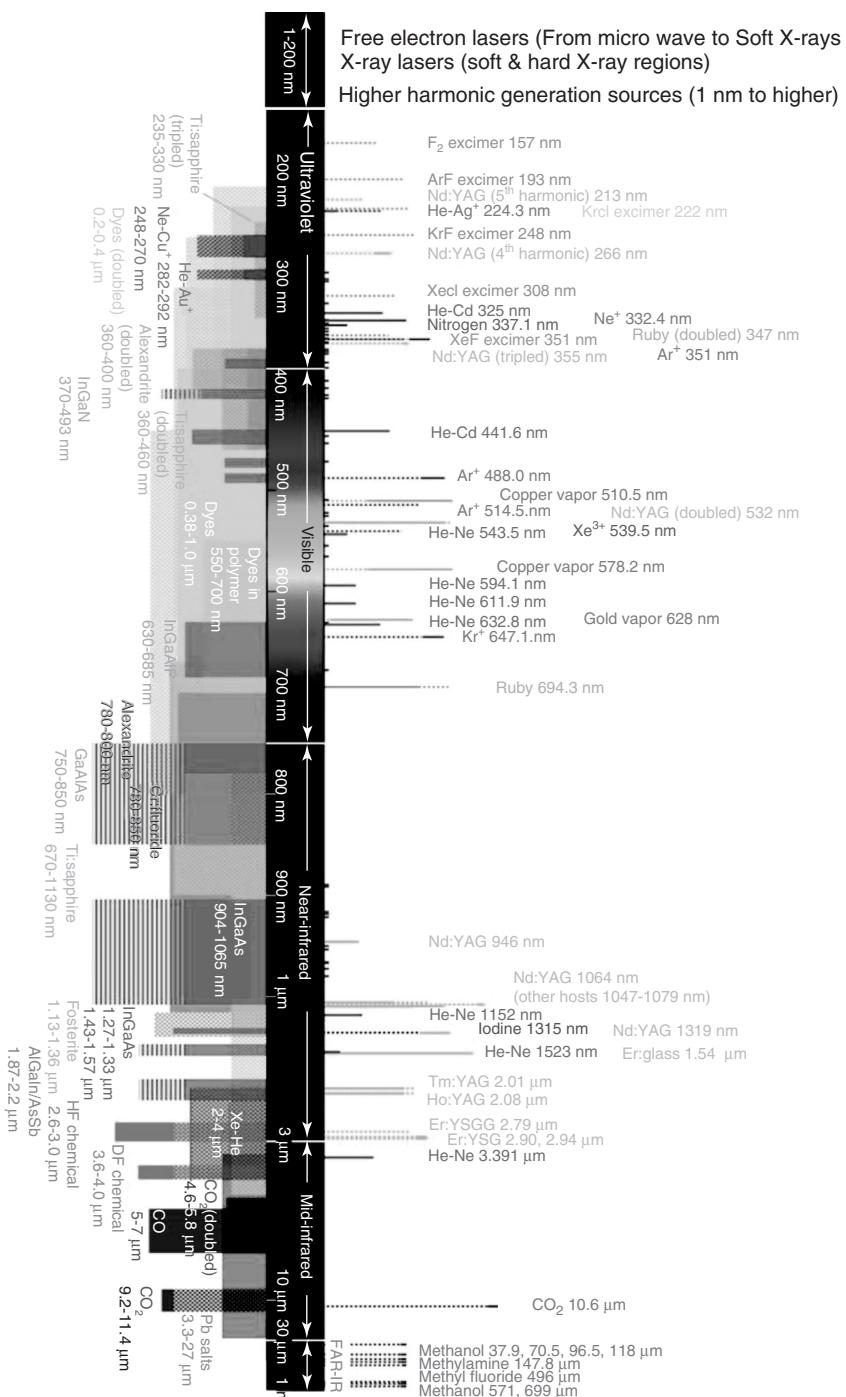


Figure 1.20 Spectrum of available laser systems, their wavelength range and active media.

modes (Figure 1.5). Suppressing all other modes excluding the central intense mode using mode locking one can increase monochromaticity of laser line. Suppressing of modes is possible by increasing the separation ($\Delta\nu_{\text{sep}} = c/2L$) between two modes, which can be done by reducing the cavity length. When the axial mode separation approaches the line width, $\Delta\nu$, of the lasing transition it is possible that only single mode oscillates. Now, the line width of the laser is equal to that of the single longitudinal mode, which is too narrow. The width of the laser line is directly related to the quality factor Q of the cavity and is given by $Q = \nu/\Delta\nu$. The quality factor Q actually defines ($Q = 2\pi \times$ energy stored in the resonator at resonance/energy dissipated per cycle) the ratio of energy stored in the cavity at resonance condition and energy dissipated per cycle. High degree of monochromaticity of laser line is required in the diagnosis of closely spaced rotational levels of molecules using selective excitation of that level. The laser line would be absolutely monochromatic if it is oscillating with single frequency, that is, width of the laser line is zero ($\Delta\nu = 0$). The single-mode laser has the highest degree of monochromaticity, but it has also not achieved the ideal monochromaticity condition.

1.4.2

Directionality

One of the most striking properties of laser is its directionality, that is, its output is in the form of an almost parallel beam. Owing to its directional nature it can carry energy and data to very long distances for remote diagnosis and communication purposes. In contrast, conventional light sources emit radiation isotropically; therefore, very small amount of energy can be collected using lens. Beam of an ideal laser is perfectly parallel, and its diameter at the exit window should be same to that after traveling very long distances, although in reality, it is impossible to achieve. Deviation in the parallelism of practical laser beam from the ideal is not due to any fault in the laser design, but due to diffraction from the edges of mirrors and windows. From the theory of diffraction, we know that circular aperture has angle of diffraction given by $\theta = \sin^{-1}(1.22\lambda/D)$. The spreading of the beam does depend on the physical nature of aperture and on the type of transverse mode oscillating inside the cavity. In the particular case of TEM_{00} mode oscillating inside the nearly confocal cavity, the value of minimum diameter at the center of cavity is given by $w_0 = (\lambda r/2\pi)^{1/2}$, where r is the radius of curvature of cavity mirrors (Figure 1.3). The beam diameter, w , varies with distance, z , from the point of minimum diameter and is given by $w = w_0(2z/r)$. The beam radius at any point inside and outside the cavity is determined by the distance from the cavity axis where intensity reduces $1/e$ times of its maximum value.

1.4.3

Coherence

Coherence is one of the striking properties of the lasers, over other conventional sources, which makes them useful for several scientific and technological

applications. The basic meaning of coherence is that all the waves in the laser beam remain spatially and temporally in the same phase. Photons generated through stimulated emission are in phase with the stimulating photons. For an ideal laser system, electric field of light waves at every point in the cross section of beam follows the same trend with time. Such a beam is called *spatially coherent*. The length of the beam up to which this statement is true is called *coherence length* (L_C) of the beam. Another type of coherence of the laser beam is temporal coherence, which defines uniformity in the rate of change in the phase of laser light wave at any point on the beam. The length of time frame up to which rate of phase change at any point on the laser beam remains constant is known as *coherence time* (t_C). Let P_1 and P_2 be two points on the same laser beam at time t and $t + \tau$. The correlation function between phases at these two points is termed as *mutual coherence function*, which is a complex number with magnitude between 0 and 1 (0 for completely incoherent beam and 1 for ideally coherent). *Coherence time* (t_C) is also defined as the time taken by the atoms/molecules in active medium to emit a light wave train of length L_C . These two coherences are thus related by $t_C = L_C/c$. The coherence time of the laser beam is almost inverse ($t_C \approx 1/\Delta\nu$) of the width, $\Delta\nu$, of the laser transition. The lasers operating in the single mode (well-stabilized lasers) have narrow line width, therefore they exhibit higher coherence time and coherence length compared to those operating in multimode. Spatial and temporal coherences of continuous laser beams are much higher compared to those of pulsed laser systems because temporal coherence in the pulse lasers are limited by the presence of spikes within the pulse or fluctuation in the frequency of emission.

1.4.4 **Brightness**

Lasers are more intense and brighter sources compared to other conventional sources such as the sun. A 1 mW He–Ne laser, which is a highly directional low divergence laser source, is brighter than the sun, which is emitting radiation isotropically. Brightness is defined as power emitted per unit area per unit solid angle. In the particular case of 1 mW He–Ne laser with 3.2×10^{-5} rad beam divergence and 0.2 mm spot diameter at exit window the solid angle ($\pi(3.2 \times 10^{-5})^2$) is 3.2×10^{-9} sr and spot area ($\pi(2 \times 10^{-4})^2$) at the exit window is $1.3 \times 10^{-7} \text{ m}^2$. Thus the brightness of the beam is given by $((1 \times 10^{-3})/(1.3 \times 10^{-7} \text{ m}^2 \times 3.2 \times 10^{-9} \text{ sr})) = 2.4 \times 10^{12} \text{ W/m}^{-2} \text{ sr}^{-1}$, which is almost 10^6 times brighter than the sun ($1.3 \times 10^6 \text{ W/m}^{-2} \text{ sr}^{-1}$).

1.4.5 **Focusing of Laser Beam**

In practice, every laser system has some angle of divergence, which increases the spot size of laser beam and reduces its brightness. If a convergent lens of suitable focal length is inserted in the path of the laser beam, it focuses laser energy into

small spot area at focal point. If w_L is the radius of the beam and f is the focal length of convergent lens, then radius of the spot at focal point is given as $r_s = \lambda f / \pi w_L$, where λ wavelength of laser radiation. If D is the lens diameter and the whole aperture is illuminated by laser radiation (i.e., $w_L = D/2$) then $r_s = 2\lambda f / \pi D$ or $r_s = 2\lambda F / \pi$, where $F = f/D$ is f number of the lens.

In the case of a particular Nd:YAG laser operating at 1064 nm wavelength and 35 mJ/pulse energy, and 10 ns pulse width. It is focused by a convex lens of $F = 5$, and whole of the lens area is illuminated by laser. The spot diameter at the focal point is given by $r_s = 2 \times 1.064 \times 10^{-6} \times 5/\pi = 3.4 \mu\text{m}$. The irradiance of the laser beam is given by $I = E(J)/\text{pulse width (s)} \times (\pi r_s^2) = 35 \times 10^{-3} / (10 \times 10^{-9} \text{ s} \times \pi (3.4 \times 10^{-6})^2) \approx 10^{16} \text{ W m}^{-2}$.

1.5

Modification in Basic Laser Structure

Addition of some electronic, optical, or electro-optical systems between the active media and mirror to modify the pulse width, pulse shape, and energy/pulse and generation of integral multiple of laser frequency is important for advanced technological applications. Mode locking or phase locking, Q-switching, pulse shaping, pulse compression and expansion, frequency multiplication, and so on, are some commonly used methods in advanced laser technology.

1.5.1

Mode Locking

1.5.1.1 Basic Principle of Mode Locking

Mode locking is a technique in optics by which a laser can be made to produce light pulses of extremely short duration of the order of picoseconds (10^{-12} s) or femtoseconds (10^{-15} s). The basis of this technique is to induce constant phase relationship between the modes of laser cavity. Simply, same phase of δ can be chosen for all laser modes. Such a laser is called *mode-locked* or *phase-locked laser*, which produces a train of extremely narrow laser pulses separated by equal time intervals. Let N modes are oscillating simultaneously in the laser cavity with $(A_0)_n$, ω_n , and δ_n being the amplitude, angular frequency, and phase of the n th mode. All these parameters vary with time, therefore modes are incoherent. The output of such laser is a linear combination of n different modes and is given by $A(t) = \sum_{n=0}^N (A_0)_n e^{i(\omega_n t + \delta_n)}$ expression. For simplicity, frequency of the n th mode can be written as $\omega_n = \omega - n\Delta\omega$, where ω_n is the mode with highest frequency and $\Delta\omega = c\pi/L$ is the angular frequency separation between two modes. If all the modes have same amplitude and we force the various modes to maintain same relative phase δ to one another, that is, we mode lock the laser such that $\delta_n = \delta$, then the expression for resultant amplitude will be $A(t) = A_0 e^{i(\omega t + \delta)} \sum_{n=0}^N e^{-i\pi nct/L} = A_0 e^{i(\omega t + \delta)} \sin(N\phi/2) / \sin(\phi/2)$, where $\phi = \pi ct/L$. The irradiance of the laser output is given by $I(t) = A(t)A(t)^* = A_0^2 \sin^2(N\phi/2) / \sin^2(\phi/2)$, which is the periodic function of the period $\Delta\phi = 2\pi$ in

the time interval $t = 2L/c$ (time of round-trip inside the cavity). The maximum value of irradiance is $N^2 A_0^2$ at $\phi = 0$ or $2p\pi$ (p is integer). Irradiance has zero value for $N\phi/2 = p\pi$, where p is an integer with values neither zero nor a multiple of N . This makes $\phi = 2p\pi/N = \pi ct/L$ or $t = (1/N)(2L/c)p$. Therefore, separation between two consecutive minima, that is, pulse width of a single laser pulse is $\Delta t = (1/N)(2L/c)$. Hence, the output of a mode-locked laser has sequence of short pulses of pulse duration $(1/N)(2L/c)$ separated in time by $2L/c$. The ratio of pulse separation to the pulse width is equal to the number of modes N , which shows that there should be a large number of modes in the cavity in order to get high-power short duration (picoseconds and femtoseconds) laser pulses.

1.5.1.2 Mode Locking Techniques

1.5.1.2.1 Active Mode Locking

We have discussed that mode locking is achieved by inducing the longitudinal mode to attain the fixed phase relationship, which may be exploited by varying the loss of the laser cavity at a frequency equal to the intermode separation $c/2L$. Let us consider a shutter between the active medium and output mirror, which is closed for most of the time and is opened after every $2L/c$ seconds (corresponding to the time of round-trip) and remains open for short duration of $(1/N)(2L/c)$ seconds. If the laser pulse train is as long as the shutter remains opened and arrives at the shutter exactly at the time of its opening, the pulse train is unaffected by the presence of shutter. The segment of the pulse, which arrives before opening and after closing of the shutter, will be clipped. Thus phase relationship of the modes is maintained by periodic oscillation of the shutter. An electro-optical or acousto-optical crystal operating on the principle of Pockels or Kerr effect respectively, may be used as a periodic shutter.

In the former case of electro-optical switching, a polarizer and an electro-optical crystal are arranged in between the active medium and laser exit mirror so that laser beam passes through the polarizer before entering into the crystal. Laser light from the active medium passes through the polarizer and gets plane polarized. When this polarized beam passes twice through the crystal (with appropriate electric field along the direction of light propagation) before returning to the polarizer, the plane of polarization is rotated by an angle of 90° , which does not allow the light beam to enter into the active media through the polarizer. In other words, the shutter is effectively closed. If there is no field along the crystal in the direction of propagation of light, there will be no rotation of the plane of polarization of light and it can pass though the polarizer and enter into the active medium (shutter is open). Similarly in acousto-optical switching, a 3D grating pattern is established in a medium (water or glass, not active medium of lasing) by the incident and reflected sound waves created by piezoelectric transducer attached at one end of the medium. This grating diffracts a part of the laser beam and creates a high loss.

1.5.1.2.2 Passive Mode Locking

Passive mode locking method consists of placing a saturable absorber inside the cavity. Saturable absorbers are molecules having a nonlinear decrease in absorption coefficients with the increase in the irradiance of light. The saturable absorber is placed between active laser medium and mirror. If an intense pulse of light passes through the saturable absorber placed inside the laser cavity, the low-power tails (weaker modes) of the pulse are absorbed because of the absorption of dye molecules. The high-power peak of the pulse is, however, transmitted because the dye is bleached. Owing to this nonlinear absorption, the shortest and most intense fluctuation grows, while the weaker dies out.

1.5.2

Q-Switching

Q-switching, sometimes known as *giant pulse formation*, is a technique by which a laser can be made to produce a pulsed output beam. The technique allows the production of light pulses with extremely high (of the order of approximately gigawatts) peak power, much higher than would be produced by the same laser if it were operating in a CW mode. Compared to mode locking, another technique for pulse generation with lasers, Q-switching leads to much lower pulse repetition rates, much higher pulse energies, and much longer pulse durations. Both techniques are sometimes applied at once. In contrast to mode locking, where we achieve a train of pulses, Q-switching provides a single strong and short pulse of laser radiation.

Q-switching is achieved by putting a variable attenuator inside the laser's optical resonator. When the attenuator is functioning, light that leaves the gain medium does not return and lasing cannot begin. This attenuation inside the cavity corresponds to a decrease in the *Q factor* or *quality factor* of the optical resonator. A high Q factor corresponds to low resonator losses per round-trip, and vice versa. The variable attenuator is commonly called a "Q-switch," when used for this purpose. Initially the laser medium is pumped while the Q-switch is set to prevent feedback of light into the gain medium (producing an optical resonator with low Q). This produces a population inversion, but laser operation cannot yet occur since there is no feedback from the resonator. Since the rate of stimulated emission is dependent on the amount of light entering the medium, the amount of energy stored in the gain medium increases as the medium is pumped. Owing to losses from spontaneous emission and other processes, after a certain time, the stored energy will reach some maximum level; the medium is said to be *gain saturated*. At this point, the Q-switch device is quickly changed from low to high Q, allowing feedback and the process of optical amplification by stimulated emission to begin. Because of the large amount of energy already stored in the gain medium, the intensity of light in the laser resonator builds up very quickly; this also causes the energy stored in the medium to be depleted almost as quickly. The net result is a short pulse of light output from

the laser, known as a *giant pulse*, which may have very high peak intensity. Similar to mode locking, active and passive techniques are used for Q-switching. Electro-optical and opto-acoustic (active Q-switching) switches, and saturable absorbers (passive Q-switching) are used to prevent feedback signal into the active media.

1.5.3

Pulse Shaping

Pulse shaping is a technique of optics, which modifies temporal profile of a pulse from laser. It may lengthen or shorten pulse duration, can generate complex pulses, or generate multiple pulses with femtosecond/picosecond separation from a single laser pulse. A pulse shaper may act as a modulator. Modulating function is applied on the input pulse to get the desired pulse. The modulating function in pulse shapers may be in time or a frequency domain (obtained by Fourier transform of time profile of pulse). In the pulse shaping methods optical signal is converted into electronic signal, where the presence and absence of pulses are designated by 1 and 0, respectively. There are two well-known pulse shaping techniques (i) direct space to time pulse shaper (DST-PS) and (ii) Fourier transform pulse shaper (FT-PS). In DST-PS, electrical analog of the desired output signal is used as a modulator with the electrical analog of input pulses. In contrast to DST-PS, the FT-PS uses a modulating function, which is a Fourier transform of the required sequence. In other words, a specific temporal sequence of pulses, the modulating function, is its Fourier transform and acts in frequency domain.

Optical design is a common experimental arrangement consisting of a combination of two sets of grating, a pair of mirrors, a pair of convex lenses, and a pulse shaper (see Figure 1.21a). By placing different types of optical components we can design various types of laser pulses. For example, employing suitable filter at focal plane can remove optical radiation of undesired frequency. A slab of transparent material with varying thickness can offer different path lengths for different frequency components and can decide which component of light will come out first. A concave lens at the focal plane provides longer path length for higher frequency component, and vice versa (Figure 1.21b). This causes the lower frequency component to come first and the higher frequency one comparatively later, resulting in positively chirped¹⁾ output pulse (frequency increases with time). In contrast to this, a planoconvex or biconvex lens makes longer path length for lower frequency, and vice versa, which makes higher frequency to come out first and shorter frequency after that, that is, negative chirped output. Both positive and negative chirping lengthens the pulse duration, that is, stretching of the laser pulse. Removing or reducing the chirp causes compression (reduction of pulse duration) of the pulse.

1) Chirping is a mechanism by which different components of light frequency of a pulse from laser comes out at different time. When frequency of light components of pulse increases with time it is called *positive chirping*, decreases with time negative chirping and if constant with time no chirping (un-chirped).

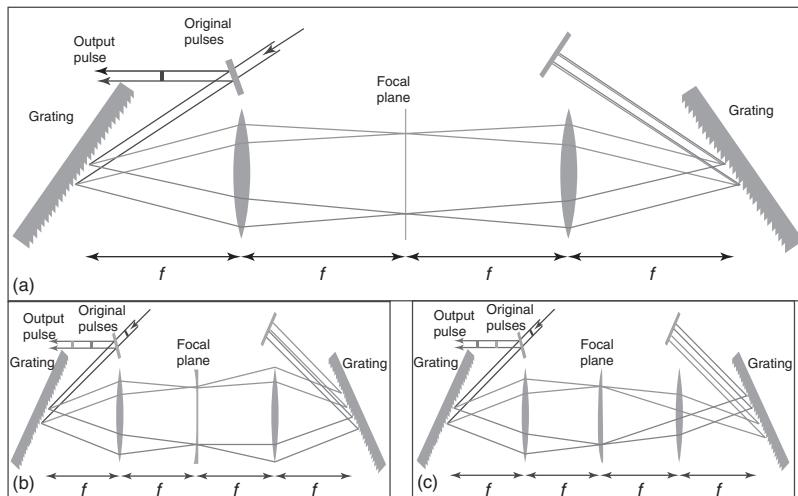


Figure 1.21 (a) Basic optical geometries for laser pulse shaping, (b) positive chirped, and (c) negative chirped pulse shaping.

References

1. Wilson, J. and Hawkes, J.F.B. (eds) (1987) *Lasers: Principles and Applications*, Prentice Hall Publications. ISBN: 013523697-5.
2. O'Shea, D.C., Callen, W.R., and Rhodes, W.T. (eds) (1977) *Introduction to Lasers and their Applications*, Addison-Wesley Publishing Company, Inc., Philippines. ISBN: 020105509-0.
3. Ifflander, R. (ed.) (2001) *Solid State Laser Material Processing: Fundamental Relations & Technical Realizations*, Springer-Verlag, Berlin, Heidelberg. ISBN: 354066980-9.
4. Weber, M.J. (ed.) (1991) *Handbook of Laser Science and Technology*, CRC Press, Boca Raton, FL, Ann Arbor, MI, Boston, MA. ISBN: 084933506-X.
5. Letokhov, V.S., Shank, C.V., Shen, Y.R., and Walther, H. (eds) (1991) *Interaction of Intense Laser Light with Free Electrons*, M.V. Fedorov; *Laser Science and Technology and International Handbook*, Harwood Academic Publishers GmbH, Switzerland. ISBN: 3718651262.
6. Dattoli, G. and Renieri, A. (1984) Experimental and theoretical aspects of Free-Electron Lasers, in *Laser Handbook*, vol. 4 (eds M.L. Stich and M.S. Bass), North Holland, Amsterdam.
7. Marshall, J.C. (1985) *Free Electron Lasers*, Mac Millan Publishing Company, New York.
8. Luchini, P. and Motz, H. (1990) *Undulators and Free-Electron Lasers*, Clarendon Press, Oxford.
9. Colson, W.B., Pellegrini, C., and Renieri, A. (eds) (1990) *Laser Handbook*, Free-Electron Lasers, vol. 6, North Holland, Amsterdam.
10. Brau, C.A. (1990) *Free-Electron Lasers*, Academic Press, Oxford.
11. Dattoli, G., Renieri, A., and Torre, A. (1995) *Lectures in Free-Electron Laser Theory and Related Topics*, World Scientific, Singapore.
12. Dunn, J., Osterheld, A.L., Shepherd, R., White, W.E., Shlyaptev, V.N., and Stewart, R.E. (1998) Demonstration of x-ray amplification in transient gain nickel-like palladium scheme. *Phys. Rev. Lett.*, **80** (13), 2825–2828.
13. Franken, P.A., Hill, A.E., Peters, C.W., and Weinreich, G. (1961) Generation of optical harmonics. *Phys. Rev. Lett.*, **7**, 118.

Introduction to Laser Theory

*John Suárez
Princeton REU Program
Summer 2003*



What we'll talk about . . .

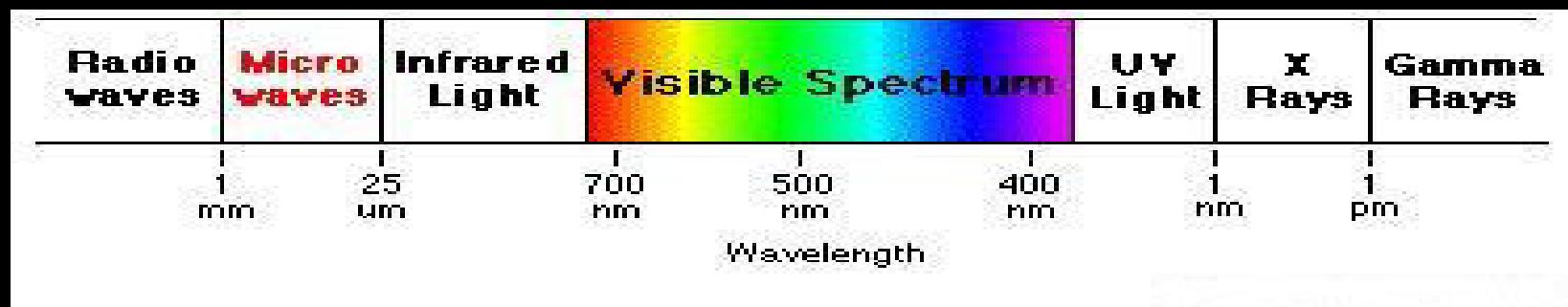
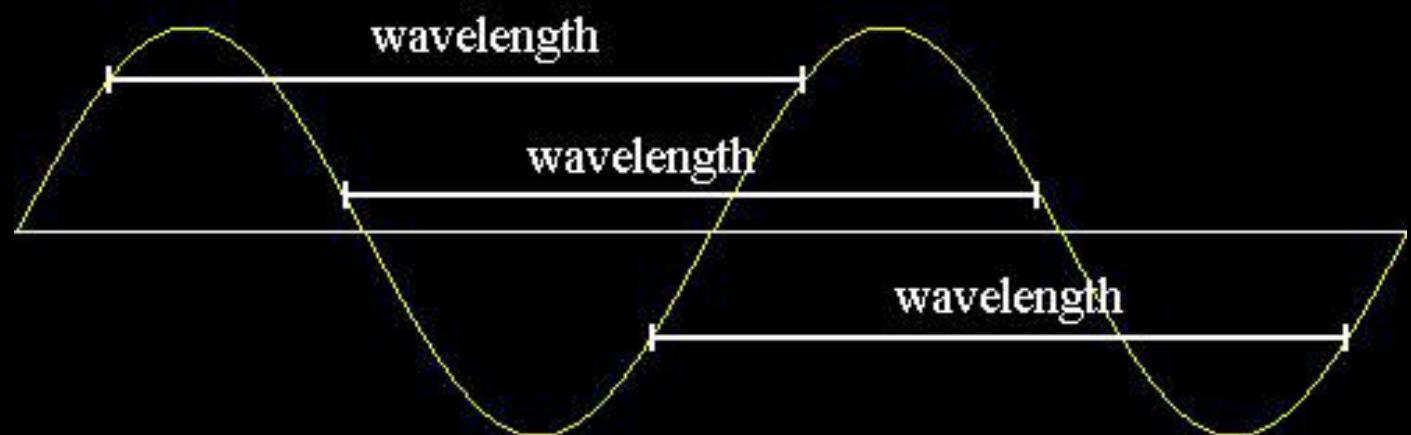
- 
- Light amplification by stimulated emission of radiation
 - Electronic energy levels
 - Energy transfer processes
 - Classifications of lasers and properties of laser beams
 - Detecting and characterizing very fast laser pulses

The Invention of the Laser

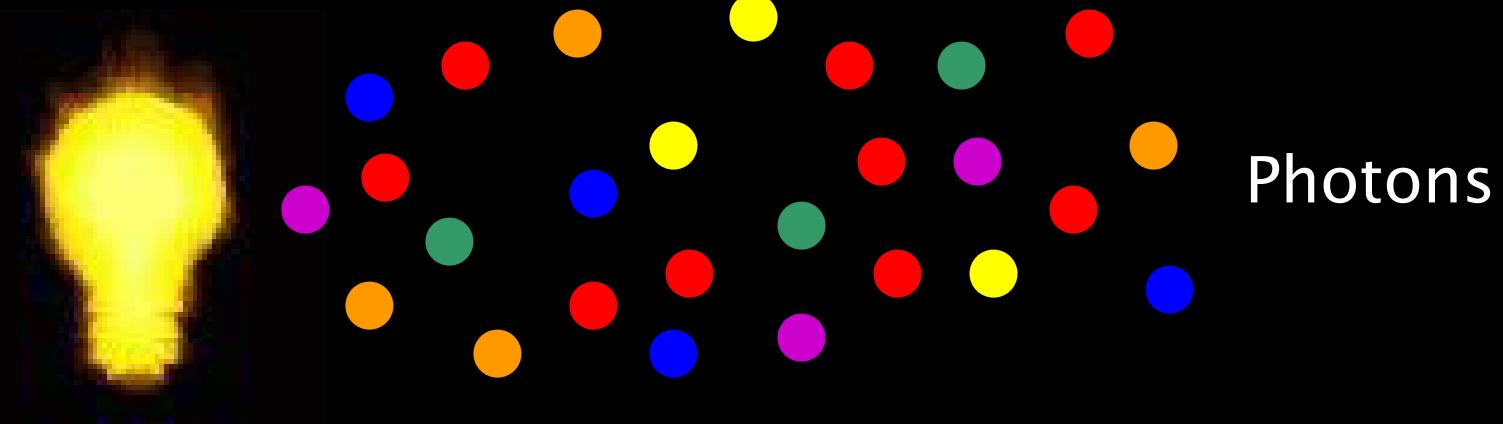
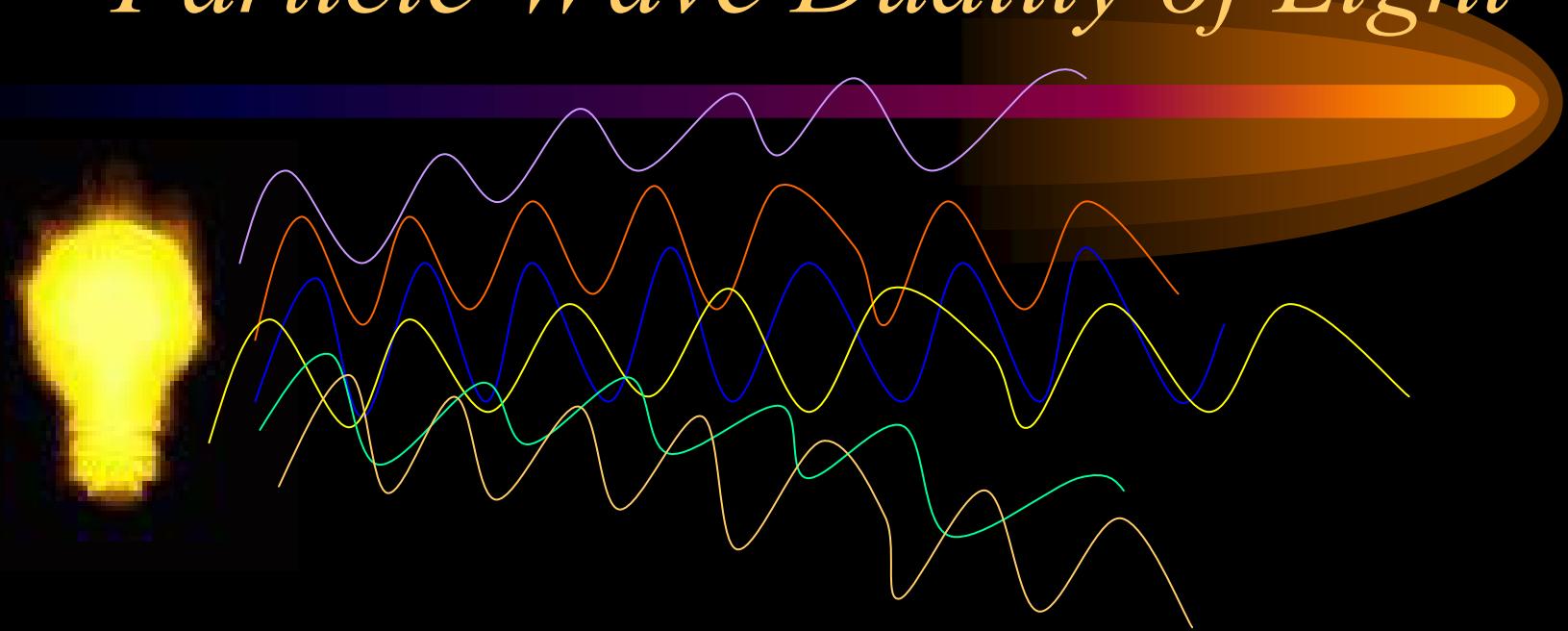


- Invented in 1958 by Charles Townes and Arthur Schawlow of Bell Laboratories
- Was based on Einstein's idea of the "particle-wave duality" of light, more than 30 years earlier
- Originally called MASER (m =“microwave”)

The Electromagnetic Spectrum

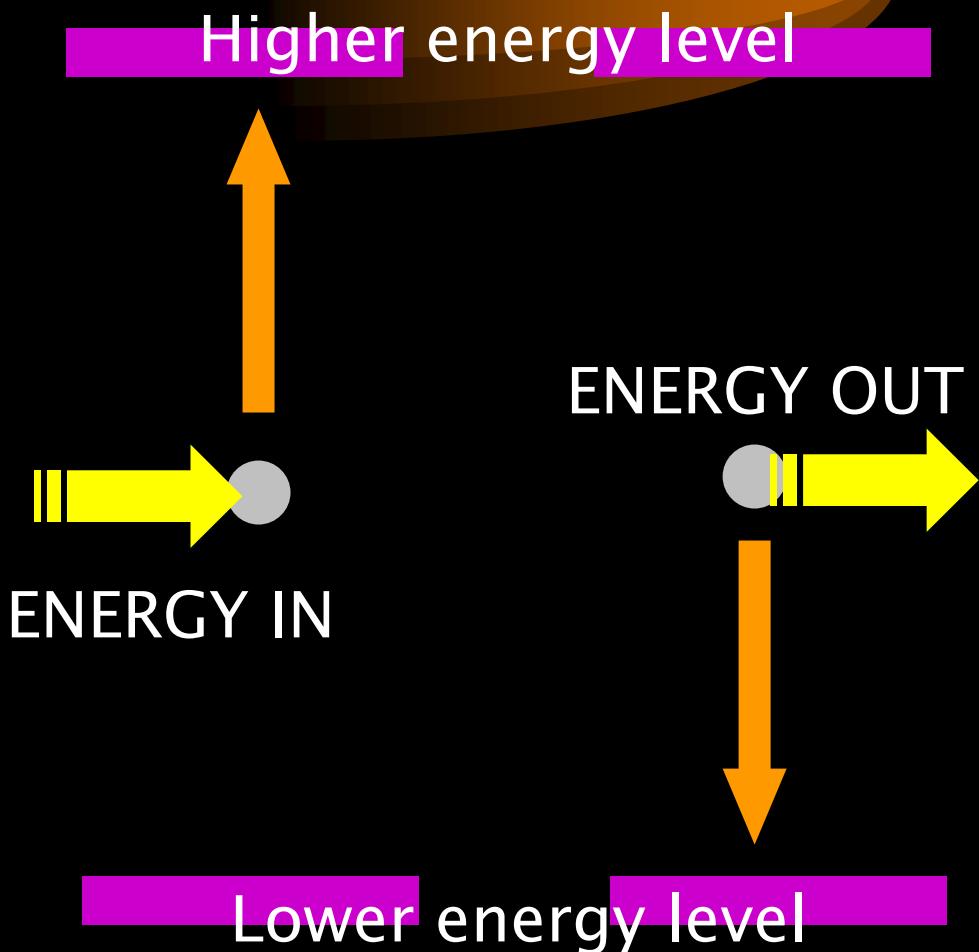


Particle-Wave Duality of Light



Energy is “quantized”

- To raise an electron from one energy level to another, “input” energy is required
- When falling from one energy level to another, there will be an energy “output”
- Theoretically: infinite number of energy levels.



Spontaneous Emission

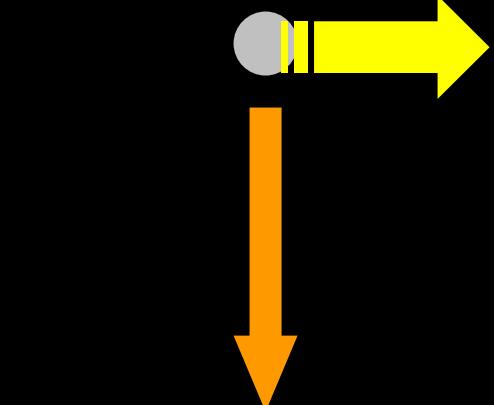
Electron initially in level 2 “falls” to level 1 and gives off energy (just happens spontaneously)

Energy is emitted in the form of a photon:

$$E = hf$$

Higher energy level

ENERGY OUT

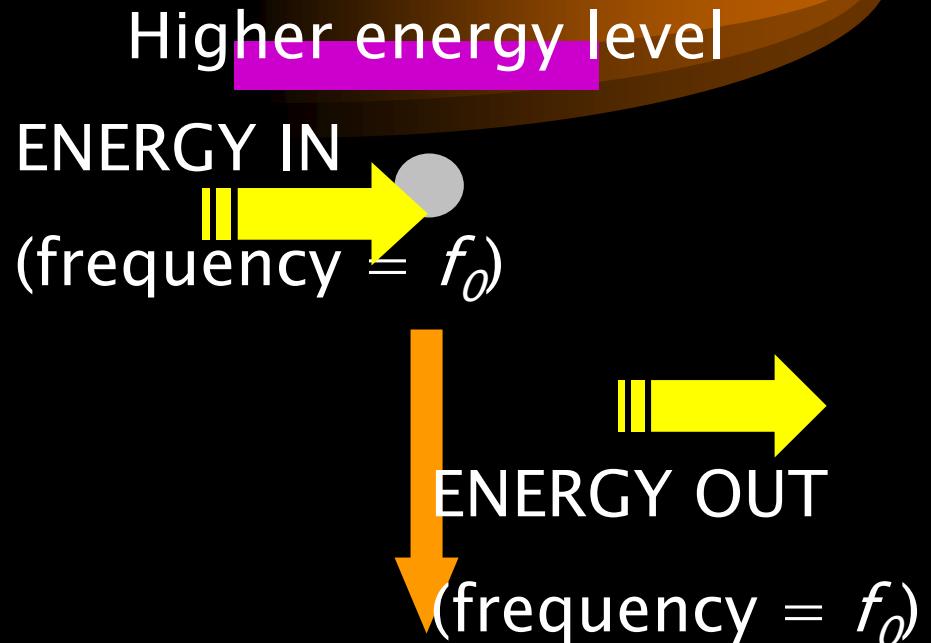


Stimulated Emission

Same idea as spontaneous emission except we MAKE it happen by sending in an EM wave of frequency f_0

A photon is given off with the energy

$$E = hf_0$$

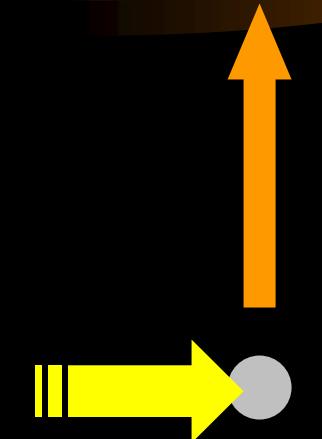


Lower energy level

Stimulated Absorption

Electron is initially in level 1. We send in an EM wave and the electron goes UP from level 1 to level 2.

Higher energy level



Lower energy level

Energy Transfer Processes



- Spontaneous emission
 - electron “naturally” falls down from level 2 to level 1

Send in an electromagnetic (EM) wave:

- Stimulated emission
 - electron can be knocked down from level 2 to level 1
- Stimulated absorption
 - electron can be raised from level 1 to level 2

Energy Transfer Processes: Quantitatively

- Define N_i , the population of level i
- $N_i(t)$ = the # of electrons, per unit of volume, occupying energy level i at time t

Energy Transfer Processes: Quantitatively

For spontaneous emission:

$$\frac{dN_2}{dt} = -AN_2$$

For stimulated emission:

$$\frac{dN_2}{dt} = -W_{2 \text{ to } 1} N_2$$

For stimulated absorption:

$$\frac{dN_2}{dt} = +W_{1 \text{ to } 2} N_1$$

Level 2

$$W_{2 \text{ to } 1} = \sigma_{2 \text{ to } 1} \Phi_{\text{photon}}$$

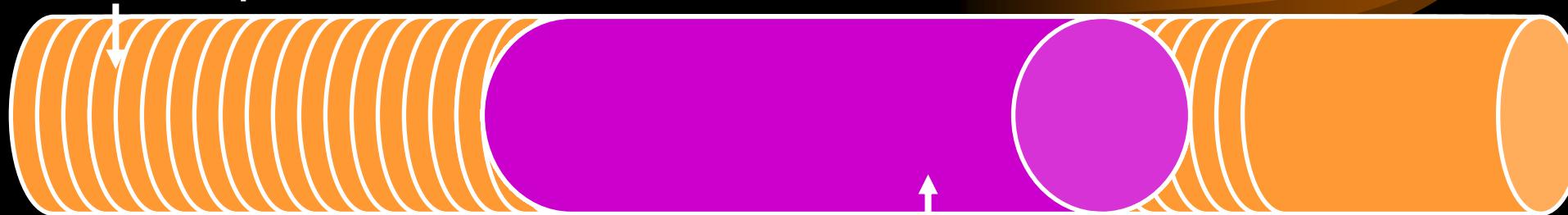
$$W_{1 \text{ to } 2} = \sigma_{1 \text{ to } 2} \Phi_{\text{photon}}$$

$$g_2 W_{2 \text{ to } 1} = g_1 W_{1 \text{ to } 2}$$

Level 1

The Laser Concept

Beam of photons



Beam has a cross-sectional
(surface) area of S

Material

As the beam travels, it will cause stimulated emission and absorption within the material.

The Laser Concept

Beam of photons

Cross-sectional
(surface) area = S

dz

Volume of this orange
region is Sdz

The Laser Concept

How would we express the incremental change in photon flux ($d\Phi$) in the material due to the photon beam?

$$Sd\Phi = \left(\begin{vmatrix} \text{change in # of photons} \\ \text{due to stim. emission} \\ (\text{per unit of time}) \end{vmatrix} - \begin{vmatrix} \text{change in # of photons} \\ \text{due to stim. absorption} \\ (\text{per unit of time}) \end{vmatrix} \right) Sdz$$

$$Sd\Phi = \left(\left| \frac{dN_2}{dt} \right|_{\text{stim. emiss.}} - \left| \frac{dN_2}{dt} \right|_{\text{stim. absorp.}} \right) Sdz$$

$$W_{2 \text{ to } 1} = \sigma_{2 \text{ to } 1} \Phi_{\text{photon}}$$

$$W_{1 \text{ to } 2} = \sigma_{1 \text{ to } 2} \Phi_{\text{photon}}$$

$$g_2 W_{2 \text{ to } 1} = g_1 W_{1 \text{ to } 2}$$

The Laser Concept

$$Sd\Phi = \left(\left| \frac{dN_2}{dt} \right|_{\text{stim. emiss.}} - \left| \frac{dN_2}{dt} \right|_{\text{stim. absorp.}} \right) Sdz$$

$$Sd\Phi = (W_{2 \text{ to } 1} N_2 - W_{1 \text{ to } 2} N_1) Sdz$$

$$\frac{d\Phi}{dz} = (W_{2 \text{ to } 1} N_2 - W_{1 \text{ to } 2} N_1)$$

$$\frac{d\Phi}{dz} = \left[\sigma_{2 \text{ to } 1} \left(N_2 - \frac{g_2}{g_1} N_1 \right) \right] \Phi$$

The Laser Concept

$$\frac{d\Phi}{dz} = \left[\sigma_{2 \rightarrow 1} \left(N_2 - \frac{g_2}{g_1} N_1 \right) \right] \Phi \quad \begin{array}{l} \frac{d\Phi}{dz} > 0 \Rightarrow \text{amplifier} \\ \text{“Population inversion”} \end{array}$$

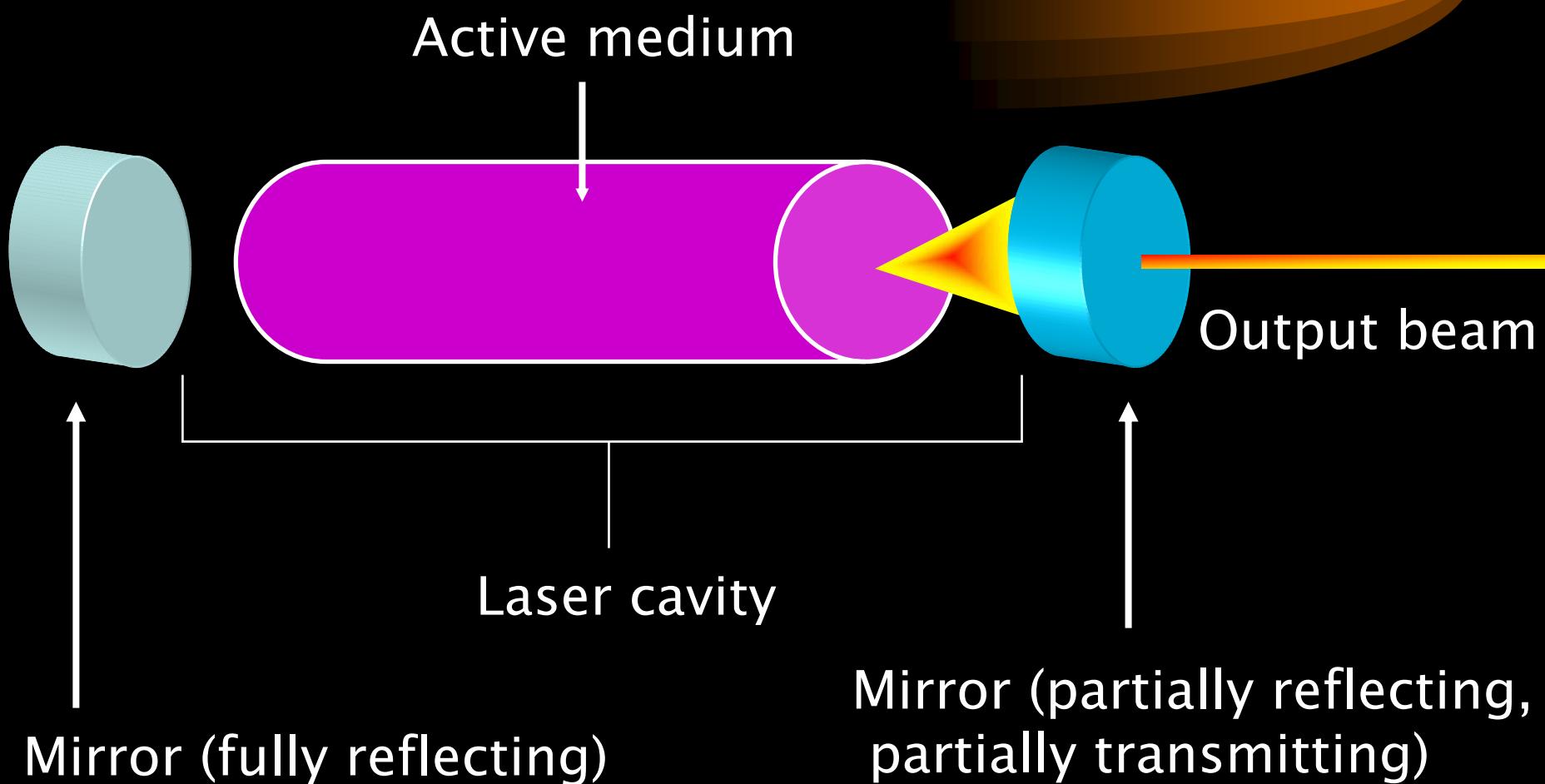
In thermal equilibrium: material acts as an absorber

Some materials: amplifiers when not in thermal equilibrium

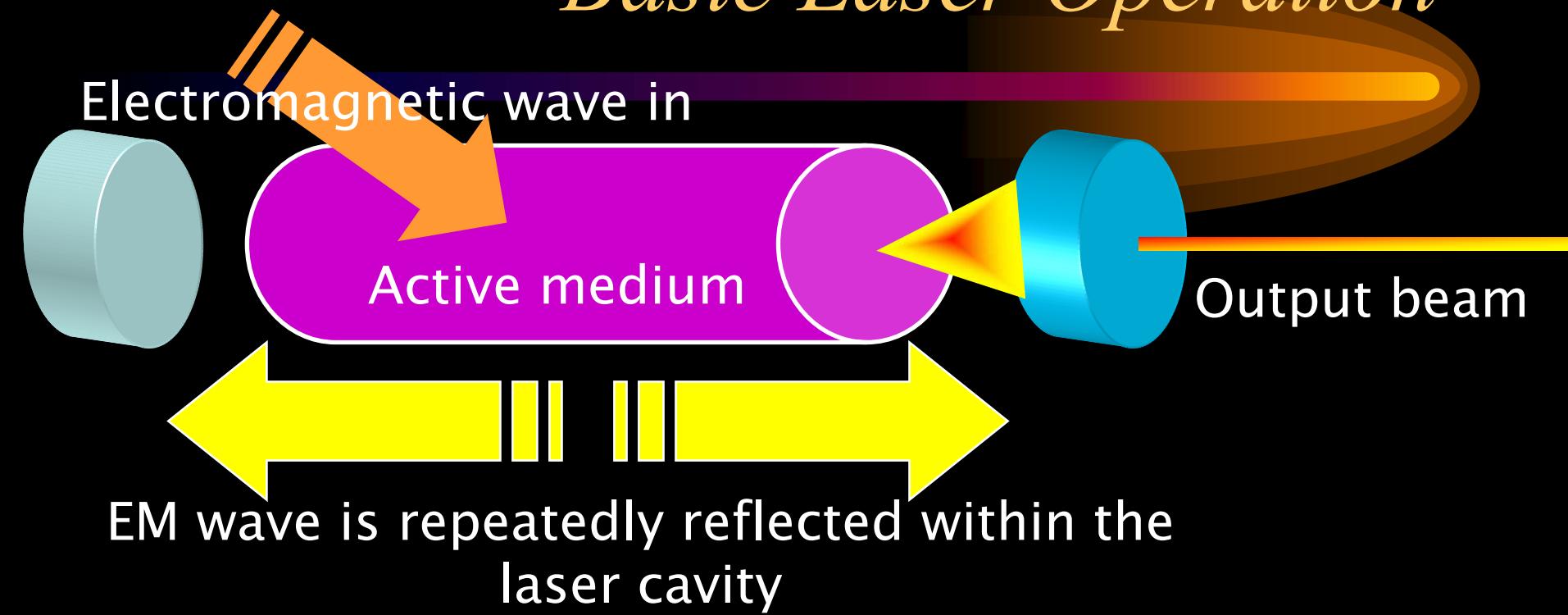


ACTIVE MEDIUM

Components of a Laser

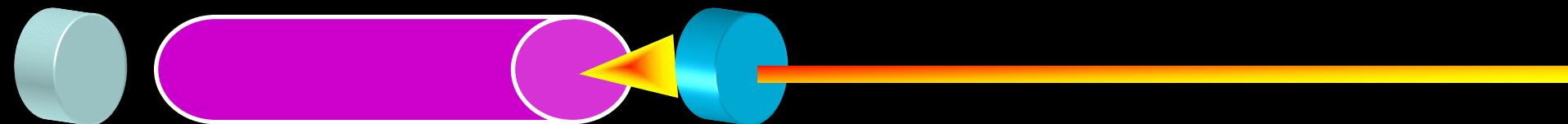
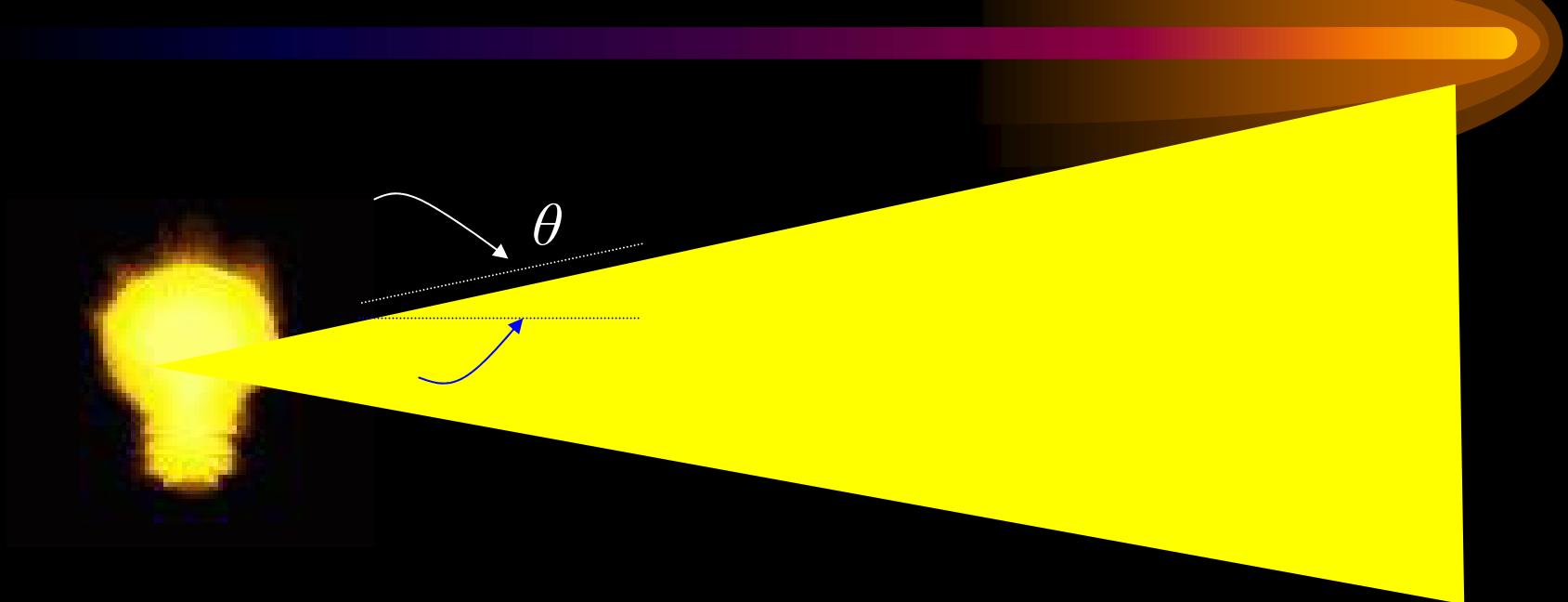


Basic Laser Operation

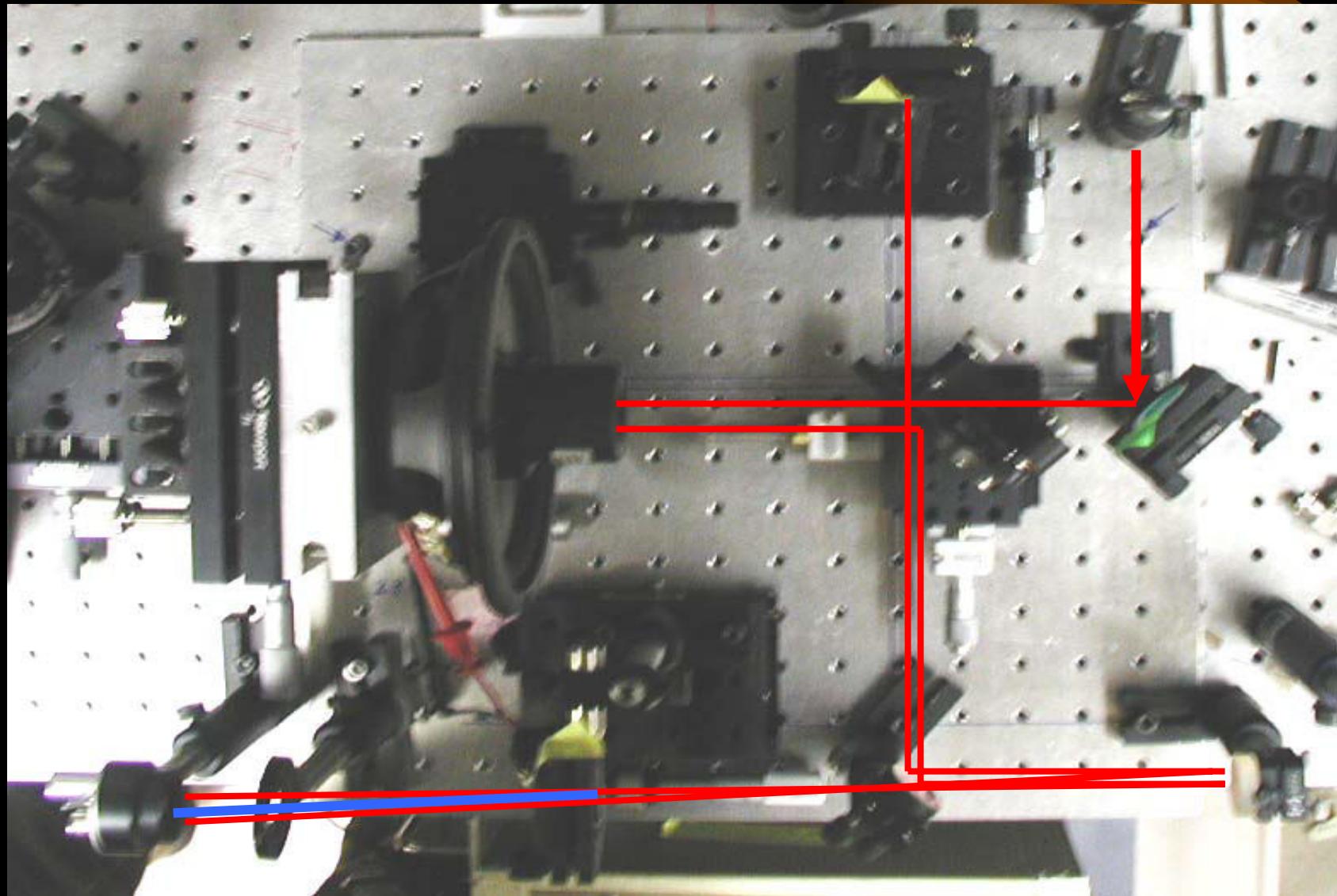


If the frequency (f) of the output beam is . . .
microwave region (1 GHz – 30 THz) . . . MASER
optical region (430 THz – 750 THz) . . . LASER

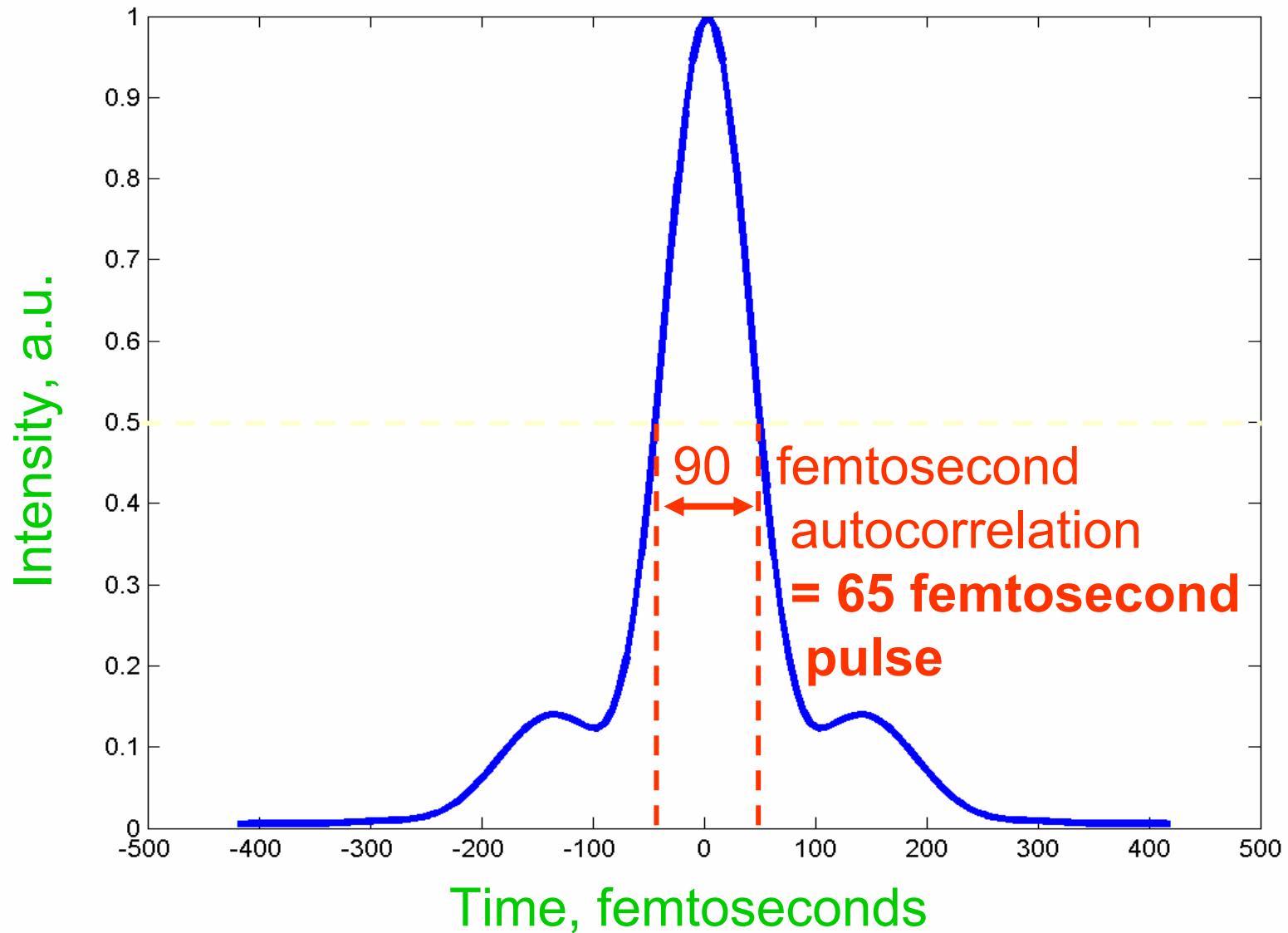
Properties of Laser Beams



Characterizing Fast Laser Pulses: Autocorrelation



Characterizing Fast Pulses: Autocorrelation



Summary

- The electromagnetic spectrum
- Energy level diagram representation
- Spontaneous and stimulated emission, absorption
- The laser concept: rate equations, population inversion, amplification
- Properties of laser beams
- Autocorrelation: characterizing femtosecond laser pulses



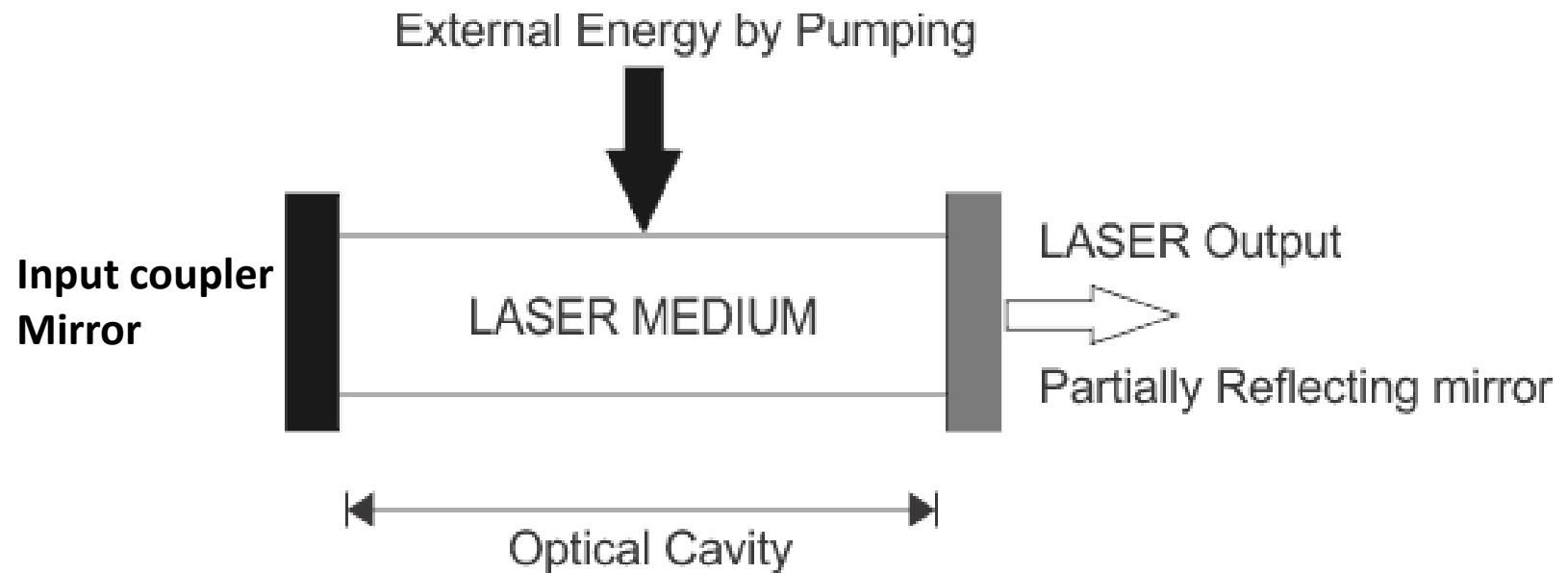
Thank You

- Dr. Warren, Wolfgang, and the entire Warren group
- Dr. Jay Benziger, Ms. Soonoo Aria, Mr. Don Schoorman
- National Science Foundation

Introduction to Laser Theory

*John Suárez
Princeton REU Program
Summer 2003*





Components of LASER

Essential components of a laser system :

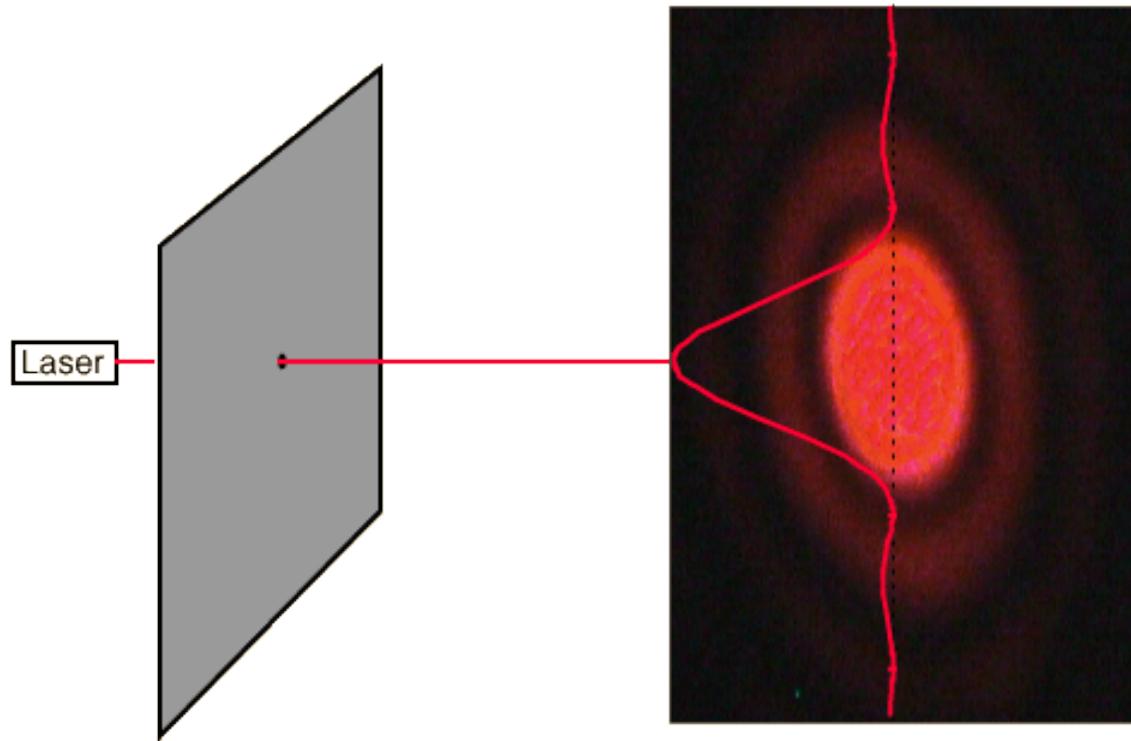
Active medium or Gain medium : It is the system in which population inversion and hence stimulated emission (laser action) is established.



Pumping mechanism : It is the mechanism by which population inversion is achieved.
i.e., it is the method for raising the atoms from lower energy state to higher energy state to achieve laser transition.

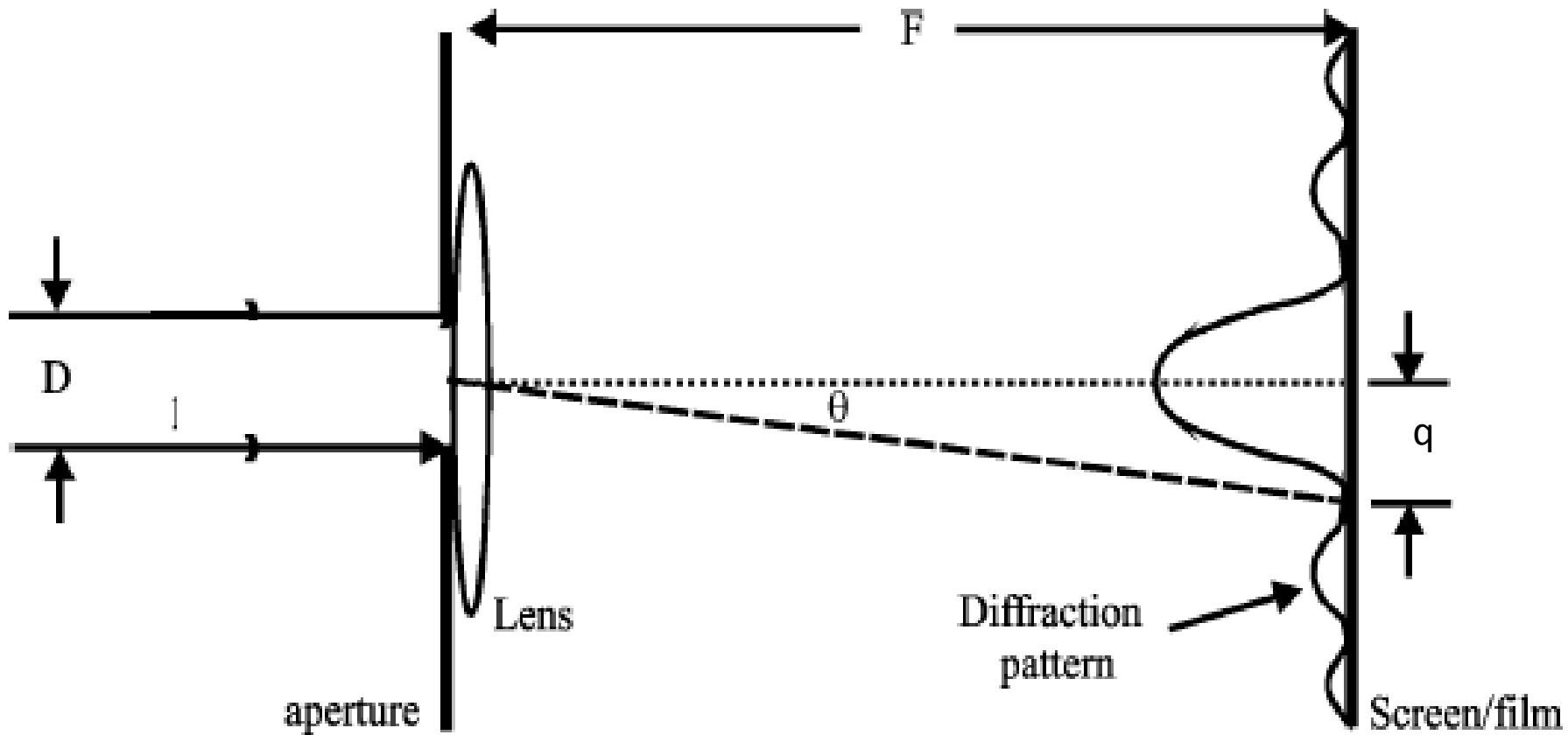
DIFFERENT PUMPING MECHANISMS :

- i. *Optical pumping* : Exposure to electromagnetic radiation of frequency $\nu = (E_2 - E_1)/h$ obtained from discharge flash tube results in pumping Suitable for solid state lasers.
- ii. *Electrical discharge* : By inelastic atom-atom collisions, population inversion is established.
Suitable for Gas lasers
- iii. *Chemical pumping* : By suitable chemical reaction in the active medium, population of excited state is made higher compared to that of ground state Suitable for liquid lasers.
- iv. *Optical resonator* : A pair of mirrors placed on either side of the active medium is known as optical resonator. One mirror is completely silvered and the other is partially silvered. The laser beam comes out through the partially silvered mirror.



Airy disk : high irradiance circular spot

Diffraction limit of Camera



Two point sources cannot be resolved if their separation is less than the radius of the Airy disk.

According to Rayleigh's criterion

Angular limit of resolution :

$$\Delta\theta = \frac{1.22\lambda}{D}$$

Limit of resolution :

$$\Delta l = \frac{1.22F\lambda}{D}$$

Δl =centre to centre separation between images

Round trip gain with losses

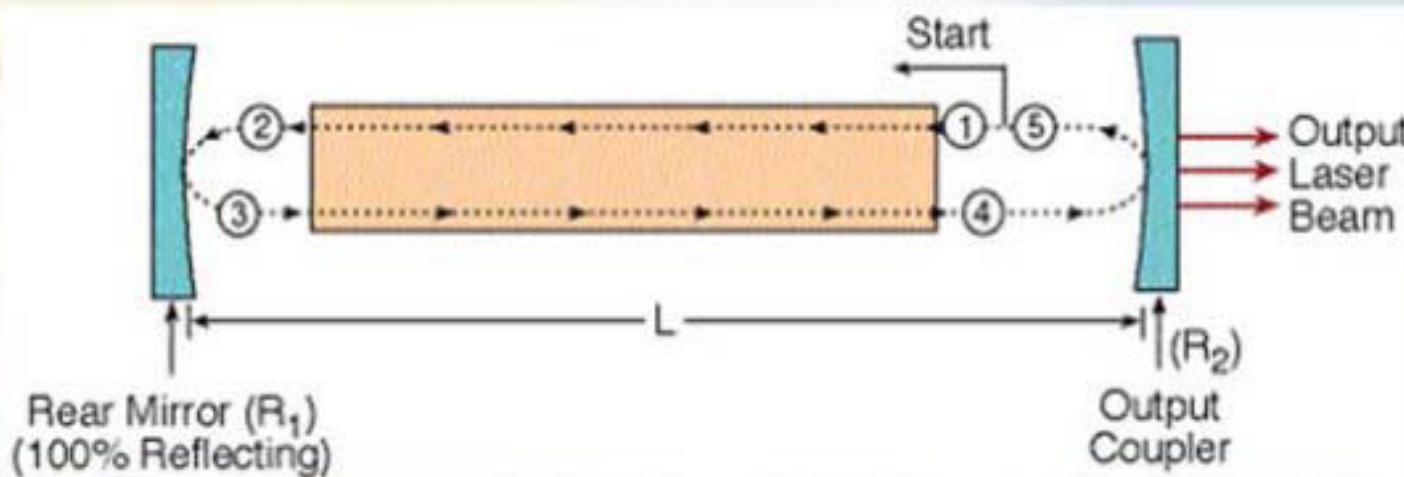
The total losses of the laser system is due to a number of different processes these are:

1. Transmission at the mirrors
2. Absorption and scattering by the mirrors
3. Absorption in the laser medium
4. Diffraction losses at the mirrors

All these losses will contribute to reduce the effective gain coefficient to ($\gamma_0 - k$)

Round trip Gain (G)

Figure below show the round trip path of the radiation through the laser cavity. The path is divided to sections numbered by 1-5, while point "5" is the same point as "1".



Round trip path of the radiation through the laser cavity.

By definition, Round trip Gain is given by:

$$G = I_5 / I_1$$

G = Round trip Gain.

I₁ = Intensity of radiation at the beginning of the loop.

I₅ = Intensity of radiation at the end of the loop.

Gain (G) Without Losses

17.Threshold_gain_22-Feb-2021_Reference_Material_I

$$I_5 = R_1 * R_2 * G^2 * I_1$$

Gain (G) With Losses

We assume that the losses occur uniformly along the length of the cavity (L). In analogy to the Lambert formula for losses, we define loss coefficient (α), and using it we can define absorption factor k:

$$k = \exp(-2\alpha L)$$

k = Loss factor, describe the relative part of the radiation that remain in the cavity after all the losses in a round trip loop inside the cavity.

All the losses in a round trip loop inside the cavity are $1-k$ (always less than 1).

α = Loss coefficient (in units of 1 over length).

2L = Path Length, which is twice the length of the cavity.

Adding the loss factor (k) to the equation of I_5 :

$$I_5 = R_1 * R_2 * G_A^2 * I_1 * k$$

From this we can calculate the **round trip gain**:

$$G = I_5 / I_1 = R_1 * R_2 * G_A^2 * k$$

As we assumed uniform distribution of the loss coefficient (α), we now define **gain coefficient (γ)**, and assume **active medium gain (G_A)** as distributed uniformly along the length of the cavity.

$$G_A = \exp(+\gamma L)$$

$$G(v) = e^{\gamma(v)L}$$

Substituting the last equation in the Loop Gain:

$$k = \exp(-2\alpha L)$$

$$G = R_1 * R_2 * \exp(2(\gamma - \alpha)L)$$

$$G = R_1 * R_2 * \exp(2(\gamma - \alpha)L)$$

When the loop gain (G) is greater than 1 ($G > 1$), the beam intensity will increase after one return pass through the laser.

When the loop gain (G) is less than 1 ($G < 1$), the beam intensity will decrease after one return pass through the laser. laser oscillation decay, and no beam will be emitted.

Conclusion:

There is a threshold condition for amplification, in order to create oscillation inside the laser.

$$G_{th} = 1$$

This Threshold Gain is marked with index "th".

For continuous laser , the threshold condition is:

$$G_{th} = 1 = R_1 R_2 G_A^2 k = R_1 * R_2 * \exp(2(\gamma - \alpha)L)$$

Example

Active medium gain in a laser is 1.05. Reflection coefficients of the mirrors are: 0.999, and 0.95. Length of the laser is 30cm. Loss coefficient is: $\alpha = 1.34 \times 10^{-4} \text{ cm}^{-1}$.

Calculate:

1. The loss factor k .
2. The round trip gain G .
3. The gain coefficient (γ).

Solution

1. The loss factor k:

$$k = \exp(-2\alpha L) = \exp[-2(1.34 \times 10^{-4}) \times 30] = 0.992$$

2. The Loop gain G:

$$G = R_1 R_2 G_A^2 k = 0.999 \times 0.95 \times 1.052 \times 0.992 = 1.038$$

Since $G_L > 1$, this laser operates above threshold.

3. The gain coefficient (γ):

$$G = \exp(\gamma L)$$

$$\ln G = \gamma L$$

$$\gamma = \ln G / L = \ln(1.05) / 30 = 1.63 \times 10^{-3} [\text{cm}^{-1}]$$

The gain coefficient (γ) is greater than the loss coefficient (α), as expected.

Example

Helium Neon laser operates in threshold condition. Reflection coefficients of the mirrors are: 0.999, and 0.97. Length of the laser is 50 cm. Active medium gain is 1.02.

Calculate:

- 1. The loss factor k .**
- 2. The loss coefficient α .**

Solution

17.Threshold gain_22-Feb-2021_Reference_Material_I_-

Since the laser operates in threshold condition, $G = 1$.
Using this value in the round trip gain:

$$G = 1 = R_1 R_2 G_A^2 k$$

1. The loss factor k :

$$k = 1 / (R_1 R_2 G_A^2) = 1 / (0.999 * 0.97 * 1.02^2) = 0.9919$$

As expected, $k < 1$.

Since $G > 1$, this laser operates above threshold.

2. The loss coefficient (α) is calculated from the loss factor:

$$k = \exp(-2\alpha L)$$

$$\ln k = -2\alpha L$$

$$\alpha = \ln k / (-2L) = \ln(0.9919) / (-100) = 8.13 * 10^{-5} [\text{cm}^{-1}]$$

Attention:

If the loss factor was less than 0.9919, then $G < 1$, and the oscillation condition was not fulfilled.

Example

**Reflection coefficients of the mirrors are: 0.999,
and 0.95. All the losses in round trip are 0.6%.**

Calculate the active medium gain.

Solution

17.Threshold gain_22-Feb-2021_Reference_Material_I_-

For finding the active medium gain G_A , the loss factor (k) must be found.

All the losses are $1-k$.

$$1-k = 0.006$$

$$k = 0.994$$

Using this value in the threshold loop gain:

$$G_{th} = 1 = R_1 R_2 G_A^2 k$$

$$(G_A)_{th} = 1/\sqrt{R_1 R_2 k} = 1/\sqrt{0.999 \cdot 0.95 \cdot 0.994} = 1.03$$

The active medium gain must be at least 1.03 for creating continuous output from this laser.

Summary

G = round trip Gain, determines if the output power of the laser will increase, decrease, or remain constant. It include all the losses and amplifications that the beam have in a complete round trip through the laser.

$$G_L = R_1 R_2 G_A^2 k$$

R₁, R₂ = Reflection coefficients of the laser mirrors.

G_A = Active medium gain as a result stimulated emission.

$$G_A = \exp(+\gamma L)$$

γ = Gain coefficient.

L = Active Medium length.

k = Optical Loss Factor in a round trip path in the laser cavity.

$$k = \exp(-2\alpha L)$$

α = Loss coefficient.

Summary

When $G = 1$, The laser operate in a steady state mode, meaning the output is at a constant power. This is the threshold condition for lasing, and the active medium gain is:

$$(G_A)_{th} = 1/\sqrt{R_1 R_2 k}$$

The round trip Gain is:

$$G_L = R_1 * R_2 * \exp(2(\gamma - \alpha)L)$$

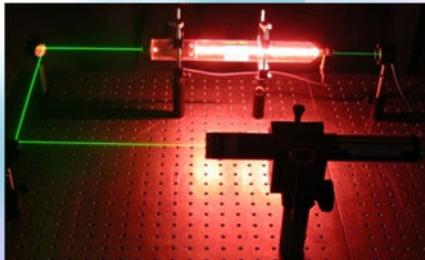
Physics Academy

Al-Azhar University - Gaza

Laser Physics

Round trip gain with losses

Lecture 13



Dr. Hazem Falah Sakeek
www.hazemsakeek.com
www.physicsacademy.org

Physics Academy

Round trip gain with losses

درسنا في محاضرات سابقة الحصيلة الناتجة عن دورة كاملة لليزر خلال المادة، وعلمنا أن انقلاب التعداد شرط أساسى لكي يستمر الحصول على ليزر ولكن فقد الناتج عن عدة عوامل يسبب في نقصان الحصيلة. ولكن نحصل على ليزر فإن الحصيلة لكل دورة يجب أن تكون على الأقل أكبر من الخسارة في كل دورة. في هذه المحاضرة سنذكر على العلاقة بين الحصيلة والخسارة.

The total losses of the laser system is due to a number of different processes these are:

1. Transmission at the mirrors
2. Absorption and scattering by the mirrors
3. Absorption in the laser medium
4. Diffraction losses at the mirrors

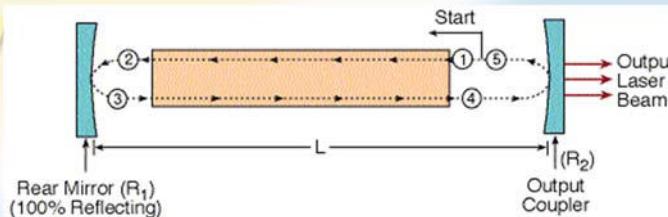
All these losses will contribute to reduce the effective gain coefficient to ($\gamma_0 - k$)

Dr. Hazem Falah Sakeek

2

Round trip Gain (G)

Figure below show the round trip path of the radiation through the laser cavity. The path is divided to sections numbered by 1-5, while point "5" is the same point as "1".



Round trip path of the radiation through the laser cavity.

By definition, **Round trip Gain** is given by:

$$G = I_5 / I_1$$

G = Round trip Gain.

I₁ = Intensity of radiation at the beginning of the loop.

I₅ = Intensity of radiation at the end of the loop.

Gain (G) Without Losses

From [lecture \(9\)](#) we found that the intensity after one round trip is given by the equation

$$I_5 = R_1 * R_2 * G^2 * I_1$$

Gain (G) With Losses

Physics Academy

We assume that the losses occur uniformly along the length of the cavity (L). In analogy to the Lambert formula for losses, we define loss coefficient (α), and using it we can define absorption factor k:

$$k = \exp(-2\alpha L)$$

k = Loss factor, describe the relative part of the radiation that remain in the cavity after all the losses in a round trip loop inside the cavity.

All the losses in a round trip loop inside the cavity are $1-k$ (always less than 1).

α = Loss coefficient (in units of 1 over length).

$2L$ = Path Length, which is twice the length of the cavity.

Dr. Hazem Falah Sakeek

5

Adding the loss factor (k) to the equation of I_5 :

$$I_5 = R_1 * R_2 * G_A^2 * I_1 * k$$

From this we can calculate the round trip gain:

$$G = I_5 / I_1 = R_1 * R_2 * G_A^2 * k$$

As we assumed uniform distribution of the loss coefficient (α), we now define gain coefficient (γ), and assume active medium gain (G_A) as distributed uniformly along the length of the cavity.

$$G_A = \exp(+\gamma L)$$

$$G(\nu) = e^{\gamma_o(\nu)L}$$

Substituting the last equation in the Loop Gain:

$$k = \exp(-2\alpha L)$$

$$G = R_1 * R_2 * \exp(2(\gamma - \alpha)L)$$

Dr. Hazem Falah Sakeek

6

$$G = R_1 * R_2 * \exp(2(\gamma-\alpha)L)$$

When the **loop gain (G)** is greater than 1 ($G > 1$), the beam intensity will **increase** after one return pass through the laser.

When the **loop gain (G)** is less than 1 ($G < 1$), the beam intensity will **decrease** after one return pass through the laser. laser oscillation decay, and no beam will be emitted.

Conclusion:

There is a threshold condition for amplification, in order to create oscillation inside the laser.

$$G_{th} = 1$$

This Threshold Gain is marked with index "th".

For continuous laser , the threshold condition is:

$$G_{th} = 1 = R_1 R_2 G_A^2 k = R_1 * R_2 * \exp(2(\gamma-\alpha)L)$$

Example

Active medium gain in a laser is 1.05. Reflection coefficients of the mirrors are: 0.999, and 0.95. Length of the laser is 30cm. Loss coefficient is: $\alpha = 1.34*10^{-4} \text{ cm}^{-1}$.

Calculate:

1. The loss factor k.
2. The round trip gain G.
3. The gain coefficient (γ).

Solution

1. The loss factor k:

$$k = \exp(-2\alpha L) = \exp[-2(1.34 \times 10^{-4}) \times 30] = 0.992$$

2. The Loop gain G:

$$G = R_1 R_2 G_A^2 k = 0.999 \times 0.95 \times 1.052 \times 0.992 = 1.038$$

Since $G_L > 1$, this laser operates above threshold.

3. The gain coefficient (γ):

$$G = \exp(\gamma L)$$

$$\ln G = \gamma L$$

$$\gamma = \ln G / L = \ln(1.05) / 30 = 1.63 \times 10^{-3} [\text{cm}^{-1}]$$

The gain coefficient (γ) is greater than the loss coefficient (α), as expected.

Example

Helium Neon laser operates in threshold condition. Reflection coefficients of the mirrors are: 0.999, and 0.97. Length of the laser is 50 cm. Active medium gain is 1.02.

Calculate:

1. The loss factor k.
2. The loss coefficient α .

Solution

Since the laser operates in threshold condition, $G = 1$.
Using this value in the round trip gain:

$$G = 1 = R_1 R_2 G_A^2 k$$

1. The loss factor k :

$$k = 1 / (R_1 R_2 G_A^2) = 1 / (0.999 * 0.97 * 1.02^2) = 0.9919$$

As expected, $k < 1$.

Since $G > 1$, this laser operates above threshold.

2. The loss coefficient (α) is calculated from the loss factor:

$$k = \exp(-2\alpha L)$$

$$\ln k = -2\alpha L$$

$$\alpha = \ln k / (-2L) = \ln(0.9919) / (-100) = 8.13 * 10^{-5} [\text{cm}^{-1}]$$

Attention:

If the loss factor was less than 0.9919, then $G < 1$, and the oscillation condition was not fulfilled.

Example

Reflection coefficients of the mirrors are: 0.999, and 0.95. All the losses in round trip are 0.6%.

Calculate the active medium gain.

Solution

For finding the active medium gain G_A , the loss factor (k) must be found.

All the losses are 1-k.

$$\begin{aligned}1-k &= 0.006 \\k &= 0.994\end{aligned}$$

Using this value in the threshold loop gain:

$$G_{th} = 1 = R_1 R_2 G_A^2 k$$

$$(G_A)_{th} = 1/\sqrt(R_1 R_2 k) = 1/\sqrt(0.999 * 0.95 * 0.994) = 1.03$$

The active medium gain must be at least 1.03 for creating continuous output from this laser.

Summary

G = round trip Gain, determines if the output power of the laser will increase, decrease, or remain constant. It include all the losses and amplifications that the beam have in a complete round trip through the laser.

$$G_L = R_1 R_2 G_A^2 k$$

R_1, R_2 = Reflection coefficients of the laser mirrors.

G_A = Active medium gain as a result stimulated emission.

$$G_A = \exp(+\gamma L)$$

γ = Gain coefficient.

L = Active Medium length.

k = Optical Loss Factor in a round trip path in the laser cavity.

$$k = \exp(-2\alpha L)$$

α = Loss coefficient.

Summary

When $G = 1$, The laser operate in a **steady state mode**, meaning the output is at a constant power. This is the **threshold condition for lasing**, and the **active medium gain** is:

$$(G_A)_{th} = 1/\sqrt{R_1 R_2 k}$$

The **round trip Gain** is:

$$G_L = R_1 * R_2 * \exp(2(\gamma - \alpha)L)$$

Dr. Hazem Falah Sakeek

Physics Academy

15

Nd-YAG laser

The lasers having the solid material as an active medium is called the solid state lasers.

Generally there are two classes of solid state lasers.

- Continuous Wave Types
- Pulsed Solid State Lasers

The output characteristics of the two are different while the construction and function of all solid state lasers is same.

Nd:YAG (neodymium-doped yttrium aluminium garnet; Nd:Y₃Al₅O₁₂) is a crystal that is used as a lasing medium for solid-state lasers.

The dopant, triply ionized neodymium, Nd³⁺, typically replaces a small fraction (1%) of the yttrium ions in the host crystal structure of the yttrium aluminium garnet (YAG), since the two ions are of similar size.

It is the neodymium ion which provides the lasing activity in the crystal.

Characteristics of Solid State Lasers

Continuous Wave Type

The term Continuous wave or CW normally indicates that a laser has continuous output. Usually in CW case of laser is continuously pumped.

There are very few solid state crystals that can produce laser light and resist the extreme heat generated by a CW pumping source.

Only common laser of this type is the ND: YAG. Output wave of the Nd: YAG laser is 1.064 microns which is in the near infrared spectrum.

The beam diameter varies from 0.75mm to 6mm while beam profile can be either $\text{TEM}_{0,0}$ or multimode.

The $\text{TEM}_{0,0}$ beams tend to have narrower beam diameters. Importantly note that the beam divergence from a solid state laser is not constant.

As the laser rod gets heated by the light source during pumping, thermal expansion occurs.

This causes the rod to act as a lens and expand beam. When the rod gets hotter then the rod expands even more and so causing greater divergence.

The beam divergence for the CW solid state laser can be as low as 1 milli radian for TEM_{0,0} mode lasers are as high as 20 milli radians fro multimode lasers.

The output power for the CW solid state laser varies from a low of 0.4 watt to a high of 600watts.

Since a Q switch can be inserted between the laser mirrors, therefore a very rapidly pulsed output can be obtained.

In addition the CW pumped solid state lasers can be mode locked for the production of ultra short pulses. In fact these lasers can produce a CW, modulated CW, Q switched or mode locked output.

Pulsed Type

The pulsed solid state laser produces a pulsed output due to the pulsing of the input energy.

In pulsed lasers the material used can be cooled between pulses therefore the active medium does not exposed to the extreme temperature rise, experienced by a CW laser.

That is why there are more choices of active media for pulsed laser rather than for CW solid state lasers.

When Nd:YAG is used as a active medium then it can be pulsed very rapidly because it can even withstand the extreme thermal loads of CW operation.

Nd:YAG lasers typically emit light with a wavelength of 1064 nm, in the infrared. However, there are also transitions near 940, 1120, 1320, and 1440 nm.

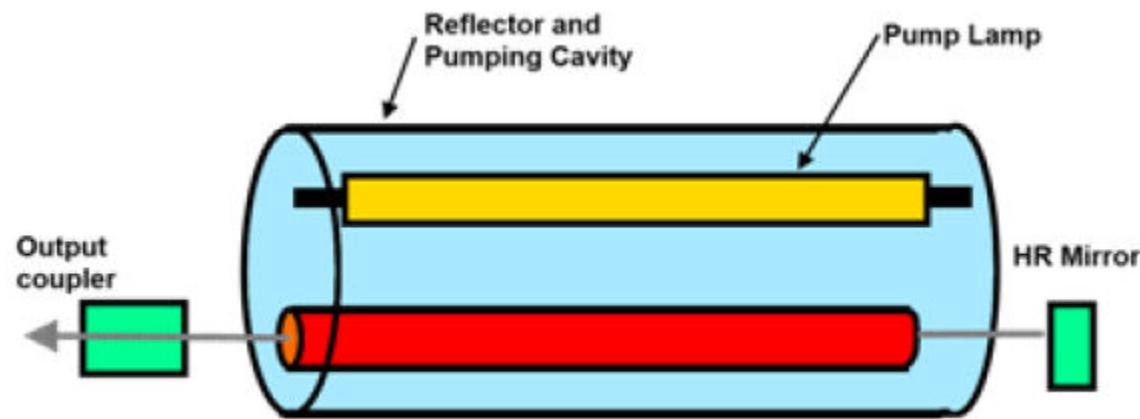
It is very common to pulse YAG lasers a rate of up to 200 pulses per second. However if ruby or Nd: glass is used as the active medium then pulse rates are limited to 2 or 3 pulses per second.

The additional time between pulses is needed to cool the crystal to prevent fracturing.

Beam diameter of pulsed laser is in 5 to 10mm range.
Beam divergence ranges from 1 milli radian to 10 milli radian.

Power output from pulsed laser averages about 400 watts although the peak power of individual pulse is much higher.

Construction of Solid State Lasers



Usually all solid state lasers have similar design.

A laser rod is mounted near an arc or flash lamp.

The lamp is connected to DC power supply that maintains a controlled current through lamp.

The laser rod and lamp are placed parallel to each other and are surrounded by a reflector.

The high reflective (HR) mirror and the output coupler are placed at either ends of the laser cavity.

For simplicity we do not show the DC power supply and cooling system, however they exists in solid state lasers.

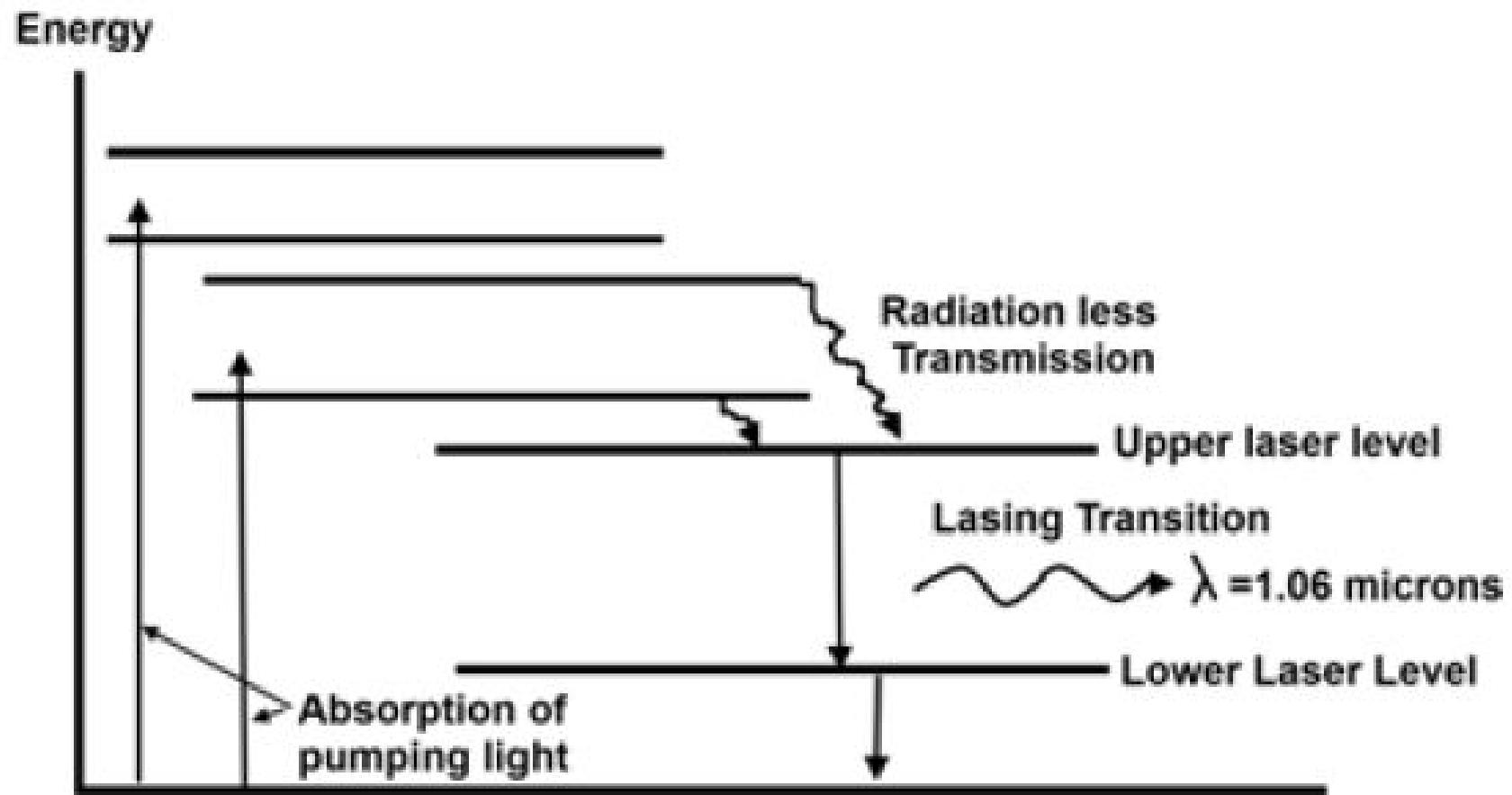
Function of Solid State Lasers

We know that active medium used for solid state lasers is a solid material.

Usually all solid state lasers are pumped optically means light source is used as energy source of solid state lasers.

When the solid material rod absorbs the light energy from light source then becomes excited.

The upper energy level of the ions are radiation less , however when the energy transition takes place then the meta stable upper laser level is soon reached. At this point the emissions occurs which obviously results in lasing.



Applications of Solid state Lasers

Nd: YAG solid state lasers usually used when drilling holes in metals.

Nd: YAG pulsed type solid state lasers can be used in medical applications such as in endoscopy etc.

As military application, Nd:YAG is used by target destination system.

Advantages of Solid State lasers

The solid state lasers have the following advantages over other types of lasers.

No chance of wasting material in the active medium because here material used is in solid form not in gas form, where this occurs.

Both continuous and pulsed output is possible from solid state lasers.

Solid state lasers have high efficiency compared to some of gas lasers such as HeNe lasers and Argon Lasers.

Efficiency of solid state Nd: YAG laser is 2% to 3%.

Construction of solid state laser is comparatively simple.

Beam diameter of solid state laser is very less than CO₂ lasers.

Output power ranging from very low value of about 0.04 watts to high value of about 600 watts.

Cost of solid state lasers is economical.

Disadvantages of Solid State Lasers

Efficiency of solid state laser is very low as compared to CO₂ lasers.

Great disadvantage of solid state lasers is the divergence, which is not constant and ranges 1 milli radian to 20 milli radian.

Output power is also not very high as in CO₂ lasers.

Due to thermal lasing in solid state lasers, the power loss occurs when the rod gets too hot.

Producing solid state Rods for Lasers

The active medium in solid state lasers can be one of different crystals.

These crystals are not found in nature but rather they are produced commercially such as Ruby, Nd: YAG (Neodymium: Yttrium, Aluminum garnet), Nd: glass (Neodymium: glass), erbium etc.

The crystals used in the lasers are made by doping a highly transparent host material with a metal that will lase.

For example YAG host material that has been heated and is in melton form can be doped with Nd. When the mixture cools, a crystal begins to form.

A cylindrical crystal is then carefully drawn from the melton material when it continues to cool. This process is called growing a crystal.

Once the crystal is completely grown then ends of the cylinder are polished to perfection.

The final result is a Nd: YAG rod, that is used as an active medium in a solid state laser.

Other solid state lasers rods are produced in similar manner. During the above process that any variation in temperature of the melton material can cause distortion in the crystal.

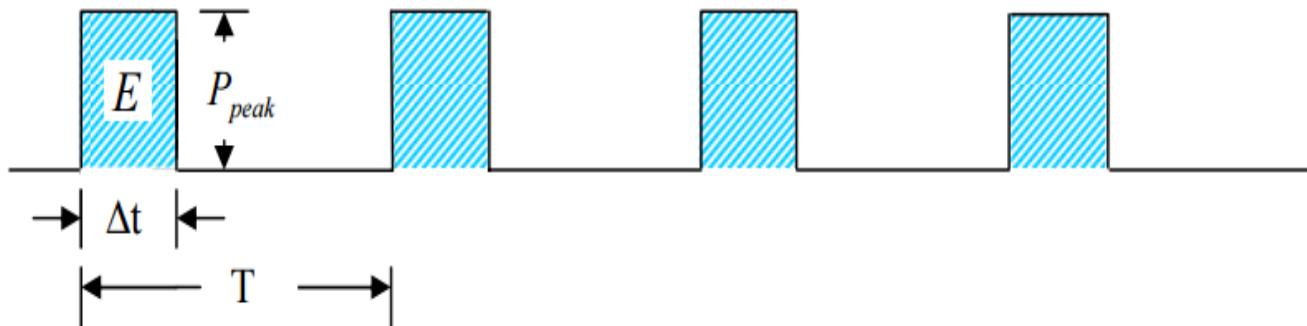
The Nd comprises between 0.7 & 1.25 of the total weight of the Nd: YAG active medium.

It must be evenly distributed throughout the crystal or "hot spots" will appear in crystal. These two can cause distortion.

Average and Peak Power – A Tutorial

It is easy to calculate the power or energy of optical pulses if the right parameters are known. Presented here are the relationships among some basic quantities often needed when working with laser pulses and power or energy meters.

Consider a regularly repeating train of optical pulses with repetition rate $f = 1/T$ as shown below.



Assume the energy, E , contained in every pulse is constant. Power is just the time rate of change of the energy flow (energy per unit time). So this leads us to define two different types of power.

1. Definition of **peak power**:

Rate of energy flow in every pulse.

2. Definition of **average power**:

Rate of energy flow averaged over one full period (recall that $f = 1/T$).

2. Definition of average power:

Rate of energy flow in every pulse.

Rate of energy flow averaged over one full period (recall that $f = 1/T$).

$$P_{peak} = \frac{E}{\Delta t}$$

$$P_{avg} = \frac{E}{T} = Ef$$

Solve both for E and equate:

$$P_{peak} \Delta t = P_{avg} T$$

Rearranging variables allows us to define a new quantity called Duty Cycle, the fractional amount of time the laser is “on” during any given period.

$$\text{Duty Cycle} \equiv \frac{\Delta t}{T} = \frac{P_{avg}}{P_{peak}}$$

Therefore the peak power of a pulse can easily be calculated if the average power and the Duty Cycle are known:

$$P_{peak} = \frac{P_{avg}}{\text{Duty Cycle}}$$

For example, ND-YAG laser might have a 10 ns pulse width, energy of 10 mJ per pulse, and operates at a repetition rate of 10 pulses per second. This laser has a peak power of

$$P_{\text{PEAK}} = 10 \text{ mJ} / 10 \text{ ns} = 1 \text{ MW},$$

and average power of:

$$P_{\text{AVG}} = 10 \text{ mJ} \times 10 \text{ (1/s)} = 100 \text{ mW}.$$

The pulse length can be very short (i.e. picoseconds or femtoseconds) resulting in very high peak powers with relatively low pulse energy, or can be very long (i.e. milliseconds) resulting in low peak power and high pulse energy.