1 Introduction

> The principles of physics, as far as I can see, do not speak against the possibility of maneuvering things atom by atom. It is not an attempt to violate any laws; it is something, in principle, that can be done; but in practice, it has not been done because we are too big.
>
> Richard Feynman

1.1 Nanoscience and Nanotechnology

The widely used International System of Units (the SI, short for Système International) was adopted in 1889. It is based on seven base units – second (s), meter (m), kilogram (kg), ampere (A), kelvin (K), mole (mol), and candela (cd) – for measuring time, length, mass, electric current, temperature, amount of substance, and luminous intensity, respectively. Prefixes representing integer powers of 10 are added to these base units to produce multiples and submultiples of the original unit. The SI system also specifies that Latin terms should be used for negative powers of 10 – e.g., milli (m), micro (μ), nano (n) – and Greek terms should be used for positive powers of 10 – e.g., kilo (k), mega (M), giga (G). The prefix nano was adopted in 1958 to precisely mean 10^{-9} SI units. According to the Oxford English Dictionary, the word *nano* originates from the classical Latin *nanus*, or its ancient Greek etymon *nanos* (νάνος), meaning dwarf [1].

In 1974, Norio Taniguchi introduced the term *nanotechnology* to describe his work on ultrafine machining and its potential for engineering devices at a submicrometer scale [2]. The modern usage of this term extends well beyond this simple machine metaphor and corresponds to a transformational technology capable of assembling, manipulating, and controlling individual atoms, molecules, or their interactions on a nanometer scale (1 to 100 nm). Even though this usage captures the essence of present-day nanotechnology, it is based on the size of objects involved and thus has many deficiencies. International Organization for Standardization (ISO), for example, has proposed to broaden the scope by including materials having at least one internal or surface feature, where the onset of size-dependent phenomena differs from the properties of individual atoms and molecules. Such structures enable novel applications and lead to improved materials, devices, and systems by exploiting nanoscale properties.

Nanoscience can be described as the science of nanoscale devices. Essentially, one can consider nanoscience as the bridge between classical and quantum physics – it is a scale

1
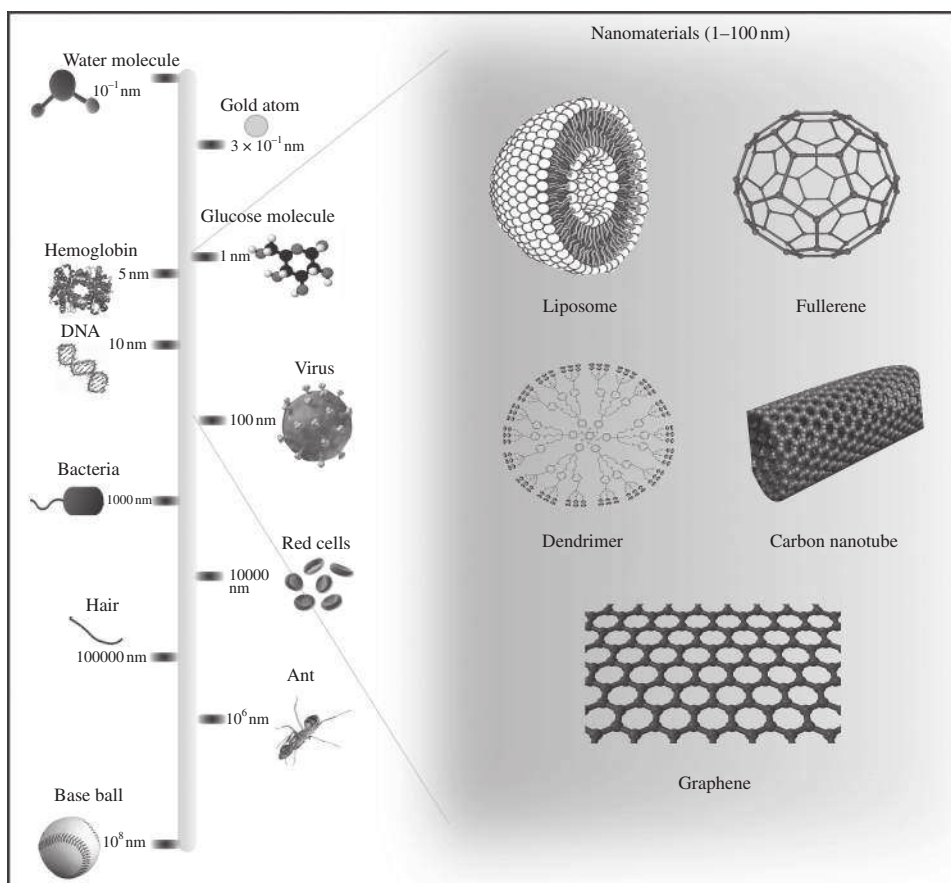
Fig. 1.1 Making sense of length scales involved in nanotechnology. Typical bond lengths are 0.12 to 0.15 nm, and a DNA double-helix has a diameter of 2 nm. The smallest cellular life-form, bacteria of the genus Mycoplasma, is 200 nm in length, and the diameter of a human hair is about 10^5 nm (or 0.1 mm). *Source:* https://commons.wikimedia.org/w/index.php?curid=32395880 ©Sureshbup. (The color version of this figure is available online.)

where we can make use of both aspects to harness collective rather than the individual properties of atoms and molecules. As we shall see later, these collective properties of individual building blocks predominantly define the novel aspects of nanostructures. Figure 1.1 shows a variety of objects covering length scales ranging from 0.1 nm to 1 cm. An expanded view of a few nanoscale (1 to 100 nm) objects involved in the development of nanotechnology is shown on the right side of this figure.

1.1.1 Historical Perspective

Historically, James Clerk Maxwell proposed in 1867 the use of tiny machines to violate the second law of thermodynamics, which states that entropy of a closed system cannot decrease. According to this law, heat must flow from hot to cold, and a perpetual motion

machine cannot be built. The gedanken experiment he proposed is known as Maxwell's demon and involves a machine (or demon) guarding a tiny hole between two gas reservoirs at the same temperature. The demon can measure the speed of individual molecules and let through only the fast ones, which would create a temperature difference between the two reservoirs without doing any work. As the second law of thermodynamics has stood the test of time, it is unlikely that Maxwell's demon would succeed, but it is fascinating to see that molecular-level sensing and manipulation ideas were conceived more than 150 years ago.

More recently, in a 1959 lecture titled "There's plenty of room at the bottom" and given to the American Physical Society, the physicist Richard Feynman alluded to the possibility of having miniaturized devices, made of a small number of atoms and working in compact spaces, for exploiting specific effects unique to their size and shape to control synthetic chemical reactions and to produce useful devices or substances.

Historical evidence exists showing that humans have exploited the interaction of light with nanoparticles, without understanding the physics behind it. An intriguing example is provided by the Lycurgus Cup shown in Figure 1.2. It is thought to have been made by Roman craftsmen during the fourth century. The cup contains gold and silver nanoparticles embedded in the glass and exhibits a color-changing property that makes its glass take on different hues, depending on the light source. It appears jade-green when observed in reflected light. However, when light is shone into the cup, it appears translucent-red from the outside. The second object in Figure 1.2, a stained glass window at Lancaster Cathedral showing Edmund and Thomas of Canterbury, uses trapped gold and silver nanoparticles in the glass to generate the ruby-red and deep-yellow colors, respectively. These visual effects can be explained using modern theories on plasmon generation, but it is still a puzzle how

Fig. 1.2 Historical evidence of the use of plasmonic effects. The Lycurgus Cup (left), thought to have been made during the fourth century, exhibits a color-changing property. It appears jade-green when looked at in reflected light but translucent-red when light is shone into the cup because it is made of glass containing gold and silver nanoparticles. The stained glass window (right) at Lancaster Cathedral in which gold and silver nanoparticles were used to generate the ruby red and deep yellow colors. (The color version of this figure is available online.)

ancient blacksmiths knew the precise material properties and compositions to realize them in practice.

Regardless of the current advances that allow humans to harness the power of nanotechnology, the striking reality is that natural processes have cleverly utilized nanotechnology effects for billions of years. Examples include harvesting of solar energy though photosynthesis, accurate replication of the DNA structure, and repair of any damage to DNA incurred because of endogenous or exogenous factors. Discovery of such effects unique to the nanoscale is the prime task of *nanoscience*. Theoretical know-how and understanding developed through nanoscience is used for *nanotechnology* that benefits the society through its specific applications such as longer lasting tennis balls, more efficient solar cells, and cleaner diesel engines. However, from prehistoric times to now, there are numerous cases where the use of a technology preceded the underlying science; practitioners were unaware of the reasons for peculiar behavior they found in materials and devices much different from familiar individual atoms, molecules, and bulk matter, yet proceeded to use them in applications – a model that modern engineers and scientists appear to emulate even now!

1.1.2 New Features Appearing at the Nanoscale

Materials interacting with electromagnetic and other fields exhibit phenomena on a broad range of spatial and temporal scales. A basic postulate in physics is the independence of any observation with respect to the choice of time, place, and units. It requires that physical quantities rescale by the same amount throughout space-time, yet it does not imply physics to be scale invariant. It is very clear that, at the smallest scale, physics demands a quantized treatment, and Planck's constant h sets the smallest observable limit. The standard model of elementary particles identifies four fundamental forces that govern our universe: gravity, the weak force, the strong force, and the electromagnetic force. Each of these forces has a specific coupling strength and a specific distance dependence. The gravitational and electromagnetic forces scale as $1/r^2$ (called the inverse-square law) and they can act over long distances, but the weak and strong forces act only over short distances. At distances greater than 10^{-14} m, the strong force is practically unobservable, and the weak force has no influence over distances greater than 10^{-18} m. All this suggests that we need to pay attention to the scale and units used for measuring different quantities.

A common feature of all forces is that they fade away as one moves from the source. Quantum field theory explains any force between two objects using exchange particles, which are virtual particles emitted from one object (source) and absorbed at the other (sink). Four types of exchange particles – photons, gluons, weak bosons, and gravitons – give rise to four forces; they all have a spin of 1 in units of $\hbar \equiv h/(2\pi)$ and transfer momentum between the two interacting objects. The rate at which momentum is exchanged is equal to the force created between the two objects. Quantum field theory shows that this force weakens as the distance between objects increases. For example, electromagnetic force between two charge particles decreases as $1/r^2$, but this dependence becomes

$1/r^4$ for dipole–dipole interactions. As the origin of most physical or chemical properties can be traced back to the interactions among the atomic or molecular constituents, all such properties tend to carry remnants of the inverse-distance dependence and manifest as size-dependent features for the nanoscale objects. For example, the following material properties become size dependent (to various degrees) when at least one of the dimensions goes below 100 nm:

- Mechanical properties [3]: elastic moduli, adhesion, friction, capillary forces;
- Thermal properties [4, 5]: melting point, thermal conductivity;
- Chemical properties [6, 7]: reactivity, catalysis;
- Electrical properties [8]: quantized conductance, Coulomb blockade;
- Magnetic properties [9]: spin-dependent transport, giant magnetoresistance;
- Optical properties [10]: band structure, band-gap energy, nonlinear response.

A practical and useful aspect of this size dependence is that it gives engineers the ability to tune one or more properties of bulk materials by resizing them to the nano-regime (1 nm to 100 nm). It is this feature that lies behind the concept of metamaterials – artificially designed materials that allow one to use nanotechnology for practical applications.

1.1.3 Surface-to-Volume Ratio

In nanosize objects, surface atoms behave somewhat differently from their bulk atoms. A simple way to judge whether the surface or bulk effects dominate is to consider the ratio of surface area A and volume V of a nanostructure. Table 1.1 compares the ratio A/V for three solids in the shape of a sphere, a cube, and a right-square pyramid. It shows that this ratio scales as $1/r$, where r is a measure of linear size. This scaling is found to hold for all regular, simple structures. Even for a complicated structure, if we can identify a single size parameter (e.g., by enclosing the structure inside a sphere of radius r), the same scaling holds approximately.

Physically, the $1/r$ scaling implies that, when the size of a three-dimensional structure shrinks, the ratio A/V increases. The drastic effect on surface area can be seen in the example shown in Figure 1.3. The cube A has 1 m sides with a surface area of $6\,\text{m}^2$. If it is divided into smaller 1 cm-size cubes (part B), each cube has an area of $6\,\text{cm}^2$ but there are 10^6 such cubes, resulting in a total surface area of $600\,\text{m}^2$. If 1 nm-size cubes are made of cube A (part C), the total surface area would become $6000\,\text{km}^2$. Even though the total volume remains the same in all three cases, the collective surface area is greatly increased with a reduced size of each cube.

A drastic increase in the area of surfaces (or interfaces) can lead to entirely new electronic and vibrational states associated with each surface. Indeed, surface effects are responsible for melting initiated at a surface (premelting) and a lower melting temperature of a compact object (compared to its bulk counterpart) [11, 12, 13]. Also, considerable variations in the thermal conductivity of nanostructures can be partially attributed to an enhanced surface area [14]. For example, thermal conductivity of a nanowire can be much

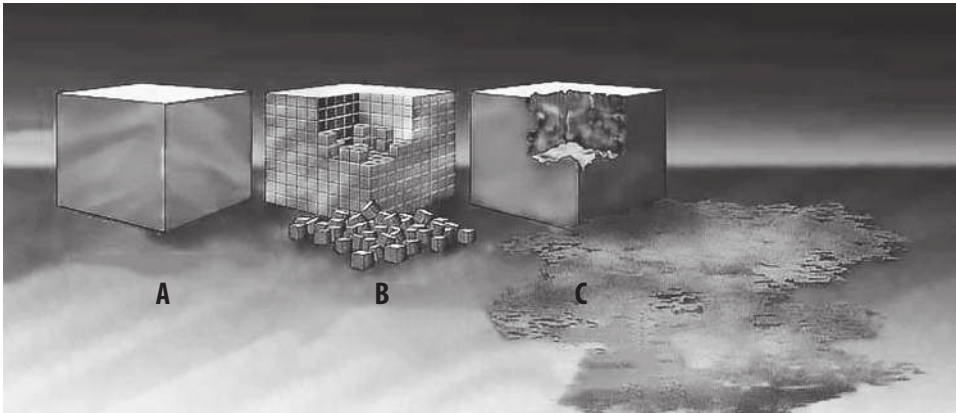| Table 1.1 Surface-to-volume ratio of three regular solids.			
	A	V	A/V
Cube of side r	$6r^2$	r^3	$\frac{6}{r}$
Sphere of radius r	$4\pi r^2$	$\frac{4}{3}\pi r^3$	$\frac{3}{r}$
Pyramid with side r	$3r^2$	$\frac{1}{3\sqrt{2}}r^3$	$\frac{9\sqrt{2}}{r}$

Fig. 1.3 Dramatic increase in total surface area of nanostructures for a given volume. The cube in part A has 1-m sides, smaller cubes in part B have 1-cm sides, and the smallest cubes in part C have 1-nm sides. The total surface area of all cubes first increases from 6 m^2 to 600 m^2 and then to 6000 km^2 for the same volume.

lower than that of the corresponding bulk material [15]; similarly, carbon nanotubes exhibit much higher thermal conductivity compared to diamonds [16].

As an another example of the surface effects, even though bulk gold is relatively inert chemically, it shows high chemical reactivity in the form of a nanosize cluster. This can be partially attributed to the abundance of surface atoms in a gold nanocluster that behave like individual atoms [17]. Similarly, bulk silver tends not to react with hydrochloric acid. However, high reactivity of silver nanoparticles with hydrochloric acid has been observed and is attributed to the electronic structure of the surface states [18].

In addition to the thermal and chemical properties, mechanical and electrical properties are also affected by an increased surface area. For example, indium arsenide (InAs) nanowires exhibit a monotonic decrease in mobility as their radius is reduced to below 10 nm. The low-temperature transport data show clearly that it is surface-roughness scattering that leads to mobility degradation [19]. Moreover, a reduced coordination of surface atoms and the presence of surface charges can exert a relatively high stress that is well beyond the elastic regime [20, 21]. Peculiar enough, charges present on the polar surfaces of thin zinc-oxide (ZnO) nanobelts can cause spontaneous formation of rings and

coils [22]. A large surface area and the resulting surface effects can also explain the occurrence of unexpected phenomena at the nanoscale such as shape-based memory and pseudoelasticity. For example, Young's modulus of films that are thinner than 10 atomic layers is found to be 30% smaller than the bulk value. All these observations indicate that the enhancement of surface-to-volume ratio plays an important role for nanosize objects.

1.2 Characteristic Length Scales

To understand the physics at the nanoscale, it is useful to have a clear understanding of a few length scales that have fundamental meaning associated with them. Whenever a characteristic dimension of a nanosize object becomes comparable to one of these length scales, we expect to see effects peculiar to the nanoscale, normally absent from its bulk counterpart. If we control the movement of elementary particles (electrons, holes, excitons, . . .) by shaping a nanosize object, the quantum effects appear especially in those regions that have dimensions comparable to a specific length scale.

Quantum mechanics shows that moving particles of matter, whether large or small, can be described both as waves and as particles (wave–particle duality). Early work in this area concentrated on demonstrating the wavelike properties of fundamental particles such as electrons, protons, and neutrons. More recently, massive particles such as C_{60} fullerene (a molecule with 60 carbon atoms, size 1.1 nm), tetraphenylporphyrin (a biodye molecule, size 2 nm), and $C_{60}F_{48}$ (a fluorinated buckyball of 108 atoms) have been shown to comply with the wave–particle duality description.

When dealing with nanosize objects, it is often not clear whether a classical or quantum model should be used to describe their behavior, and there are situations when both are needed. One good example is the photoelectric effect of light, for which both wave and particle aspects of light are required for a complete description. Another example is diffraction of electrons from crystals. The length scale that governs the wave nature of particles is known as the de Broglie wavelength, denoted as λ_{DB} and defined in Aside 1.1. This length scale plays an important role in nanoscience. We will see later that when motion of a nanosize object is restricted to dimensions smaller than its de Broglie wavelength, not only quantum-size effects appear but the associated density of states is also modified.

Further insight could be gained by looking at the analogous situation in optics. The geometrical-optics approximation of light does not take into account its wave nature. It provides us with a strong clue when we could discard the wave effects. Geometrical optics works well for analyzing interaction of light with objects whose physical dimensions are larger than the wavelength λ of light. For example, if light passes through a slit whose width is much larger than λ, we can use geometrical optics. However, when the slit width becomes comparable to λ, the wave nature of light cannot be ignored, and the diffraction effects will be pronounced and noticeable. If we replace the wavelength of light by the de Broglie wavelength of a particle, the same reasoning can be used to decide when the wave nature of the particle becomes relevant. We shall occasionally use the term *quantum*

particle to remind the reader that we intend to use both properties judiciously as needed to model certain effects.

Aside 1.1   de Broglie Wavelength

The de Broglie wavelength, associated with any moving particle with momentum p, is defined as

$$\lambda_{\mathrm{DB}} = h/p, \tag{1.1}$$

where the Planck constant has a value $h = 6.626 \ 10^{-34}$ J-s. It is useful to express p in terms of relativistic energy E and mass m of the moving particle:

$$E^2 = p^2 c^2 + m^2 c^4, \tag{1.2}$$

where c is the speed of light in vacuum. This relation implies that λ_{DB} can be written as

$$\lambda_{\mathrm{DB}} = \frac{hc}{\sqrt{E^2 - m^2 c^4}}. \tag{1.3}$$

When a particle is traveling close to the speed of light, its rest-mass energy mc^2 can be discarded compared to E to obtain

$$\lambda_{\mathrm{DB}} \approx hc/E. \tag{1.4}$$

If the particle is travelling much slower than the speed of light ($pc \ll mc^2$), we can use the approximation $E \approx mc^2 + p^2/2m$. Such a particle has a kinetic energy given by $E_k = p^2/2m$, and its de Broglie wavelength is given by

$$\lambda_{\mathrm{DB}} \approx \frac{h}{\sqrt{2mE_k}} = \frac{h}{mv}, \tag{1.5}$$

where v is the particle's speed.

The magnitude of λ_{DB} is immeasurably small for macroscopic objects such as humans, but it is of the same order (0.1 nm) as chemical bonds for subatomic particles. As its value depends on the momentum, λ_{DB} can vary considerably for any particle, depending on its mass and speed (or equivalently its kinetic energy). For example, consider a situation where electrons have a kinetic energy of about 100 eV, resulting in a de Broglie wavelength of about 0.1 nm. This value is comparable to typical spacing between atoms in a crystal. As a result, electrons behave as waves and are diffracted by a crystal. However, if the kinetic energy of electrons is increased to 100 MeV, λ_{DB} becomes so small that electrons behave as particles, showing no diffraction effects in the same crystal.

Depending on the context, de Broglie wavelength has different names. The *thermal length* λ_T, for example, is a length scale related to the de Broglie wavelength of a gas consisting of noninteracting, slowly moving particles in equilibrium at temperature T. The

kinetic energy E_k for such particles is of the order of $\pi k_B T$, where k_B is the Boltzmann constant. Using this value in Eq. (1.5), thermal length of this gas is obtained from

$$\lambda_T = \frac{h}{\sqrt{2\pi m k_B T}}. \tag{1.6}$$

Consider a gas in thermal equilibrium. The characteristic interparticle distance d for such a gas can be estimated from the packing density of cubes of volume d^3 filling a unit volume, giving $d = n^{1/3}$, where n is the number density of particles. As an example, oxygen at room temperature has d around 3 nm, whereas λ_T is estimated to be around 0.02 nm from Eq. (1.6). Since $\lambda_T \ll d$, it follows that this gas does not require a quantum-mechanical treatment (except for collisions that do require a quantum treatment).

As a second example, consider electron–hole pairs inside a semiconductor. Since the effective mass of electrons and holes can vary from 1% to 100% of a free electron's mass, the corresponding de Broglie wavelength at 300 K is in the range of 73 nm to 7.3 nm. As the temperature approaches 3 K, these values increase by a factor of 10, as seen clearly from Eq. (1.6). Apart from temperature, the de Broglie wavelength can also be manipulated in semiconductors through impurity doping. The main point to stress is that λ_{DB} of charged carriers in semiconductors can vary from 7 nm to 1 μm, depending on the experimental conditions, and a quantum treatment may be needed in some cases.

The situation becomes somewhat different for metals, which are characterized by nearly free-moving electrons whose density can be controlled thermally, chemically, or optically. Metals exhibit high electrical conductivity as well as high thermal conductivity compared to other materials. Metals also generally obey the Wiedemann–Franze law at high temperatures (which states that the ratio of thermal to electrical conductivity is proportional to the temperature). Owing to the abundance of electrons with little interaction among them, the simplest model assumes that metals are a collection of positive ions in a sea of non-interacting electrons and treats metals as a free-electron gas. The positive ions are assumed to form a regular lattice and provide a periodic potential in the Schrödinger equation. As a consequence of the Bloch theorem, the electron's wave function is proportional to $\exp(i\mathbf{k} \cdot \mathbf{r})$, where $\hbar\mathbf{k}$ represents the electron's momentum and $\mathbf{r}$ is its position vector. The Fermi energy for a free-electron gas can be written as $E_F = \hbar^2 k_F^2 / 2m$, where m is the effective mass of electrons in the metal and k_F is the wave number for this energy.

The energy distribution for a free-electron gas is governed by the Fermi–Dirac distribution; it is similar in nature to the Maxwell–Boltzmann distribution of an ideal gas (or Planck's blackbody radiation). The Fermi–Dirac distribution depends on the chemical potential μ of the material and its temperature, both of which can be experimentally measured. The significance of chemical potential is most readily seen in the limiting case of $T = 0$ K, for which the ground state is obtained by placing all electrons into the lowest available energy levels up to the energy E_F. As a result of this filling, there is a sharp boundary in the three-dimensional $\mathbf{k}$-space between the filled and empty states. This surface is known as the *Fermi surface* and its shape is spherical for a free-electron gas.

The Fermi wavelength λ_F is a length scale defined as the de Broglie wavelength of electrons at the Fermi energy E_F. If the momentum of electrons at this energy is $p_F \hat{\mathbf{p}}$,

where $\hat{\mathbf{p}}$ is a unit vector, then $E_F = p_F^2/2m$ is the kinetic energy of each electron. Thus, the Fermi wavelength is given by

$$\lambda_F = \frac{h}{p_F} = \frac{h}{\sqrt{2mE_F}}. \tag{1.7}$$

This relation provides good agreement with experimentally observed behavior because only electrons in the vicinity of the Fermi surface participate in most physical processes. For this reason, the Fermi energy and the shape of the Fermi surface are important in practice. Typically, λ_F for metals is around $0.6\,\text{nm}$, and this value can be controlled by changing the density of free electrons. This dependence is widely exploited in semiconductors by changing the type and density of dopants. The Fermi wavelength as a length scale plays an important role in both the design and analysis of metal-based nanoscale devices. For example, as we will see later, λ_F is an important parameter for understanding quantum conductance of nanostructures.

Sometimes a characteristic temperature, called *Fermi temperature* and denoted by T_F, is also defined as $T_F = E_F/k_B$. For most metals, its value lies in the range 10^4 K to 10^5 K and is much higher than room temperature or typical temperatures at which metals are used. Thus, under typical operation, temperature changes do not affect the Fermi momentum p_F much. Suppose the perturbation to p_F is given by $\delta\mathbf{p}$. Then, the dispersion relation providing the electron's energy E_e as a function of its momentum can be written as

$$E_e = \frac{1}{2m}(p_F\hat{\mathbf{p}} + \delta\mathbf{p}) \cdot (p_F\hat{\mathbf{p}} + \delta\mathbf{p}) = E_F + \mathbf{v}_F \cdot \delta\mathbf{p} + o(\delta p^2), \tag{1.8}$$

where $\mathbf{v}_F$ is the *Fermi velocity* with the magnitude p_F/m. Since $\delta\mathbf{p}$ is relatively small, we retain only the first-order term. This approximation is appropriately called semiclassical approximation and it enables us to interpret the electron's motion analogous to a classical particle moving with the Fermi velocity. Because of a relatively large value of the Fermi temperature for most metals, they can be described with sufficient accuracy assuming zero-temperature conditions.

Another widely used length scale is the electron's *mean free path* l_e. Its value is calculated by averaging the distance different electrons travel between two scattering events. Clearly, this value is not precise unless the type of scattering (elastic, inelastic, . . .) is specified. For example, one can calculate this quantity for elastic scattering, occurring when electrons come near impurities or collide with solid boundaries. Noting that only electrons close to the Fermi surface participate in such scattering, the average time τ_e between two elastic scattering events, known as the *mean collision time*, can be used to find the electron mean free path as $l_e = \tau_e v_F$. The collision time τ_e is also called the relaxation time for elastic collisions of electrons.

However, inelastic collisions of electrons are more prevalent in solids. These can occur when electrons interact with other electrons, phonons, plasmons, or change their energy through interband transitions. The fraction of electrons subjected to such inelastic processes depends on the energy of electrons. Inelastic scattering affects phase coherence of electrons and, if it occurs multiple times, all coherence is lost. In this situation, electrons obey a diffusion law with a diffusion coefficient D_{ie}, and it becomes possible to define another length scale, the *inelastic scattering length*, as