

Chapter 1

The hubris of classical physics

1.1 Explaining nature: Apples and the moon

This book is about quantum physics which deals with, by and large, phenomena at very small scales. The way microscopic bodies like atoms, electrons etc. behave is quite different from what we are used to in our everyday life, based on the behaviour of tennis balls or motor cars. It came as a great surprise to physicists in the early part of the twentieth century that the microscopic world behaves in such a spooky way, so to speak, and it took some time for them to digest and accept it. Given this complexity, it is important to first understand how physical laws operate in the more familiar setting before plunging ourselves into the strange world of the quantum. This will be the aim of the present chapter.

It is probably best to approach this task through a series of simple examples. Let us begin by throwing a ball vertically up and catching it on its return. The ball starts moving upwards with some initial speed (which we will denote by v) and, as it goes up, its speed keeps decreasing. After some time, it reaches a maximum height (when its speed is instantaneously zero) and then it starts its downward journey. Its speed now increases and eventually it comes back to you with (roughly) the same speed as you threw it upwards. We are all familiar with this phenomenon. *The role of the physicist is to explain it.*

What does one mean by “explain” ? Well, to begin with you should be able to *predict* beforehand how high the ball will rise, how long it will take for the ball to come back etc.; more generally, we want predict its height x at any given time t ; that is, you need to predict the function $x(t)$. You can’t do any of these unless you have a ‘rule’. (One usually says ‘laws of physics’ but ultimately they are just rules for calculating numbers which

can be verified against observations). In this particular case, this is done using the rules of mechanics, which we will now describe.

The first question is what makes a body — which, in this case, is a ball — move. It turns out that you need a force to act on the body if its speed or direction of motion has to change. To simplify the description it is convenient to introduce at this stage the notion of *velocity*, which is essentially speed in a particular direction. The velocity can change when the speed of the body changes; but it can also change when the speed remains the same but its direction of motion changes. If you bounce a good rubber ball on the ground, its speed just before and just after hitting the ground will be the same but the direction of motion will change, thereby changing the velocity. Our rule says that you need to apply a force to change the velocity of a body; more importantly, it says you do *not* need any force to keep a body moving with a uniform velocity. This result is actually quite a deep one and has the names of Galileo and Newton associated with it. In contrast, Aristotle thought (wrongly) that you need a force to act on a body even to keep it moving with uniform velocity.

The rate at which the velocity changes is called *acceleration* and this is what you provide when you step on the gas in your car. The so called Newton's second law, which is the cornerstone of mechanics, says that the acceleration induced on a body by a force is proportional to the force. To state this more precisely, we introduce a quantity called the *momentum* for a moving body which — according to Newton — is given by the product of the mass of the body and its velocity; $p = mv$ where m is the mass of the body and v is its velocity. The rate at which momentum changes will be equal to the force that is applied. Taking the mass of the body to be a constant, this tells you that the force changes the velocity of the body; that is, force causes acceleration. So if you want to change the velocity quickly, you need to apply more force. Note that it is the rate at which the velocity changes which is important, not the velocity itself. You need to apply the same force to change the velocity of a body from 100 m s^{-1} to 101 m s^{-1} in 0.1 sec as to change the velocity from 1000 m s^{-1} to 1001 m s^{-1} in 0.1 sec.

Coming back to the example of the ball which you have thrown upwards, it is clear that you gave it an initial velocity. If there was no force acting on it, this velocity cannot change and the ball will continue to move upwards forever with the same velocity. However, we know that Earth is exerting a gravitational force on this ball, which changes its velocity. In this case, the force is acting downwards — in the direction opposite to the direction

of initial velocity — and hence the force will decrease the velocity. (This is called *deceleration* as opposed to acceleration). Eventually it will reduce the velocity to zero which is the instant at which the ball has reached the maximum height. From the next instant onwards, the downward pull on the ball gives it a velocity in the downward direction. Now the ball will start accelerating towards Earth and eventually you will catch it.

Given the acceleration induced on the ball by Earth's gravitational field — which happens to be about $g = 9.8 \text{ m s}^{-2}$ — you can compute how long it will take for the body to reach its maximum height. It is given by $t = v/g$; if you throw the ball with 9.8 m s^{-1} speed, it will take about 1 sec to reach its peak. We have done our first computation using a law of physics and predicted a result; someone can now check it with a ball and stop watch and either verify or disprove our law.

This example tells you the role played by physics in “explaining” nature. Here we start with the simple observation of how a ball behaves when it is thrown up. We then introduce certain laws: one to connect its motion with the force that is acting on it; and another to describe the force itself which — in this case — is the gravity of Earth. Given these two, we can explain the way any ball thrown upwards will behave! For our purpose, we only needed to say gravity produces an acceleration of 9.8 m s^{-2} on the ball but this is, of course, a very approximate statement. In different contexts, we need to use different laws.

The same analysis, of course, can be used to determine the behaviour of *falling* bodies. Suppose you drop a ball towards the ground from a tall tower of height h (Or suppose an apple falls from a tree of height h). How long will it take for the apple or ball to reach the ground and how fast will it be moving when it hits the ground? Since the downward acceleration produced on the apple by Earth's gravity is g , it would have acquired a speed $v = gt$ in a time interval t . (Remember that acceleration is just the rate at which speed changes so that $g = v/t$). If the apple took an amount of time T to hit the ground, its final speed will be $V = gT$ while initially it had zero speed. So the average speed with which it moved is $(1/2)(gT + 0) = (1/2)gT$. With this average speed, it would have covered a total distance $(1/2)gT \times T = (1/2)gT^2$ which we know is the height h . So we get the relation $h = (1/2)gT^2$. If we know h and g , this relation tells you how to determine the time it takes for an apple to fall to the ground.

A more nontrivial example is given by the motion of Moon around Earth. The Earth exerts a gravitational force on the Moon making it fall towards Earth, just as it does to an apple. The notion of Moon ‘falling’

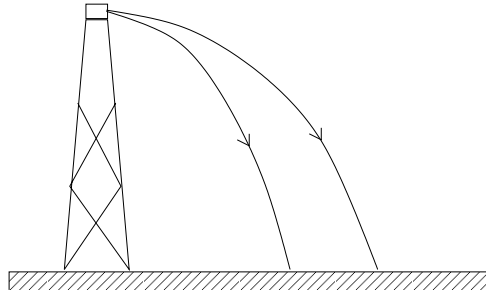


Fig. 1.1 The trajectory of a ball thrown horizontally from the top of a tower. It will fall at some distance away from the foot of the tower. If the initial speed is increased it will fall farther away. In this figure we have assumed that Earth is flat.

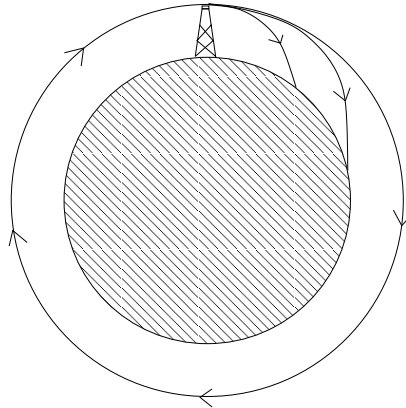


Fig. 1.2 The trajectory of a ball thrown horizontally from the top of a tower taking into account the curvature of Earth. (The dimensions are quite exaggerated.) As the initial horizontal speed of the ball is increased, the ball travels a larger distance before hitting the ground. At a critical value of the speed, the ball misses the ground completely because the Earth ‘curves away’. The ball now will be in a circular orbit around the Earth. This shows that an artificial satellite (or moon) orbiting the Earth is actually falling towards the Earth all the time in its orbit.

towards Earth may be a bit strange; so we will explain that idea first. To see what is involved, consider a stone thrown *horizontally* from the top of a tower (see Fig. 1.1). We know from experience that it will fall at some distance away from the foot of the tower. If you throw it with larger speed, it will fall farther away from the foot of the tower. Figure 1.2 shows the same situation but now taking into account the fact that Earth is not flat.

It is now clear that, as we increase the speed of the stone, a limit will be reached when the stone simply goes round and round the Earth! Of course, the stone is all the time trying to fall to the ground but the Earth is ‘curving away’ preventing this from happening. This is precisely how you place a satellite in an orbit. Instead of throwing it from a tower, we take the satellite up in a rocket and then give it the appropriate horizontal speed so that it will keep falling to Earth without ever hitting it. The Moon, of course, is just a natural satellite orbiting the Earth in the same manner. It is constantly falling to the Earth, just like an apple from the tree.

Earth induces an acceleration of 9.8 m s^{-2} on the apple but a much smaller acceleration on the Moon. It turns out that, according to Newton’s law of gravity, the acceleration induced by a body of mass M on another body at a distance r is $g = GM/r^2$ where G is a constant. So the acceleration increases in proportion to the mass but decreases as a square of the distance. That is, when the distance is doubled, the force falls by a factor $2 \times 2 = 4$; if the distance is tripled, it falls by a factor $3 \times 3 = 9$ etc. Such a force law is said to satisfy the “inverse square” relation. Since the Moon is nearly 60 times farther away from the centre of the Earth than the falling apple, the acceleration is only about 0.0027 m s^{-2} (see Fig. 1.3). You can now use this rule to calculate how exactly the Moon should move around Earth and — of course — the result matches with the observations.

You should pause for a moment to think through the miracle we have achieved. Given the two laws (one relating motion to force and another giving how the acceleration due to gravity varies with distance) we have unified a purely terrestrial phenomenon (falling of apple or a ball) to an awesome celestial phenomenon (motion of Moon around the Earth). *This is the true spirit of physics*: It provides verifiable links between apparently unrelated phenomena and allows one to describe a large number of facts using a handful of ‘rules’ and, of course, a lot of maths which we have skipped! We no longer need two explanations for the apple falling towards Earth and Moon going around the Earth — they are fundamentally the same.

While the real drama of nature in the above examples happens through changes (changing position of bodies, changing velocities) one of the fundamental considerations in physics is to sift through this and determine physical quantities that remain constant. In the cases we considered above it turns out that there is one physically important quantity, viz. the total energy, which is conserved — that is, it does not change during the motion. Take the case of the ball that is thrown up which has two different kinds

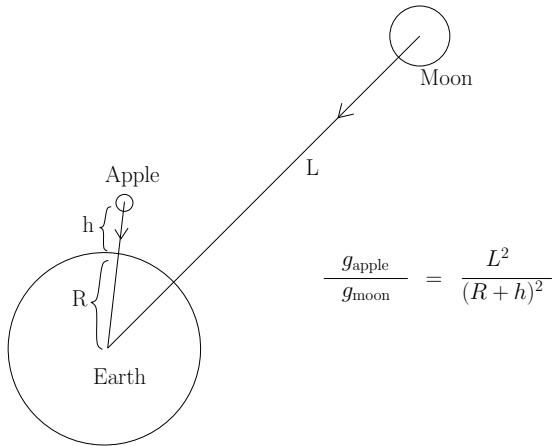


Fig. 1.3 Earth's gravitational force affects both the apple falling toward the Earth and the Moon which is falling toward the Earth at each instant in its orbit (see the previous figures). Given the distance to the Moon, one can compute the gravitational forces and relate the motion of the Moon to the motion of the apple.

of energies. The first is the kinetic energy that arises due to the motion and is given by the formula $K = (1/2)mv^2$ where m is its mass and v is its speed. Since the speed of the ball keeps changing, this kinetic energy is not a constant. The second one is called the potential energy which is due to the location of the ball with respect to the ground and is given by $V = mgx$ where x is the height above the ground (which is taken to be positive) and g is the acceleration due to gravity (see Fig. 1.4). Initially, $x = 0$ making the potential energy zero while the ball has some kinetic energy due to its initial speed. As the ball rises to its maximum height, its speed and kinetic energy decrease; but its potential energy increases because x increases. The miracle is that the sum of kinetic and potential energies, which is the total energy, remains constant during the motion. When the ball is in its downward trip, its kinetic energy keeps increasing while its potential energy keeps decreasing keeping the total energy again constant. We will see in later chapters that a good part of physics is a quest for such conserved quantities. (Incidentally, the expressions given above for kinetic and potential energies are approximate — as we will see in the next chapter — but the fact that the total energy is conserved is an exact statement.)

Energy is just one of the many quantities which are conserved during

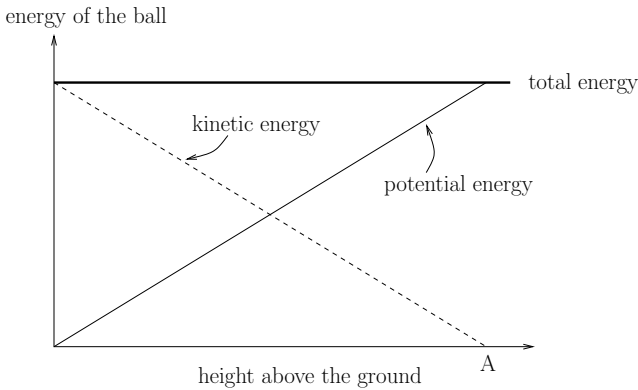


Fig. 1.4 Conservation of energy in the motion of a ball thrown upwards from the ground. As it rises up its kinetic energy decreases and potential energy increases keeping the sum constant. On reaching the maximum height at A , the kinetic energy is zero and the total energy is equal to the potential energy. On its way down, the potential energy decreases with the kinetic energy increasing.

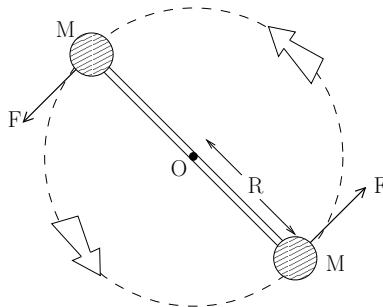


Fig. 1.5 The concept of angular momentum for a simple rotating system. When a force F is applied on each of the two masses connected by a rod as shown in the figure, the system will rotate about the centre O . If each of the masses move with a speed v , the angular momentum of either mass is defined to be MvR .

the time evolution of a system. Another quantity which is often conserved — and will play a crucial role in our description of quantum mechanics — is called *angular momentum*. Angular momentum of a rotating body, in Newtonian mechanics, is essentially the sum of the momenta of the particles in the body but weighted by the distance of the particles from the axis of rotation. Consider, for example, the dumbbell shaped body in Fig. 1.5 made of two masses M connected by a rigid rod of length $2R$. The system is free

to rotate about the centre O . Suppose we apply a force F to the two masses in the direction shown in the figure. Since the forces are equal and opposite, the net force acting on the body is zero. Nevertheless, from our experience, we know that the body will rotate with the masses moving in a circle of radius R . This is described mathematically by noting that the forces exert a torque on the system about the centre O . This torque produces a quantity called angular momentum of the system (just as unbalanced forces give rise to the momentum of the system). The angular momentum of either mass is given by the product of the momentum of the mass and the distance of the mass from the centre. In this case we get the angular momentum of either of the masses to be MvR where v is the velocity with which the mass will move. To be precise, we need to mention that the angular momentum is also a vector and — in this case — it is directed along the direction of the axis of rotation.

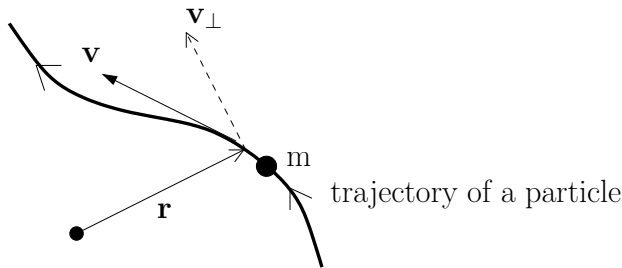


Fig. 1.6 For a particle moving along some arbitrary trajectory, the angular momentum at any given instant is defined as the product of the magnitude of the position vector \mathbf{r} of the particle and the momentum of the particle in the direction perpendicular to the position vector. In the figure v_{\perp} denotes the magnitude of the velocity in the direction perpendicular to the position vector and hence the angular momentum is $mv_{\perp}r$.

As a second example consider a ring of mass M and radius R , rotating uniformly about an axis passing through the centre and normal to the plane of the ring. If each particle in the ring is moving with speed v we attribute to the rotating ring the angular momentum MvR . You see that one can increase the angular momentum by increasing any of the three quantities M , v or R . In the case of a more complicated motion like the one shown in Fig. 1.6, the angular momentum of the particle at any given instant is calculated by the product of the mass, the distance from some chosen origin and the magnitude of the velocity in the direction perpendicular to its position vector. In Fig. 1.6 the quantity \mathbf{r} denotes the position vector

of the particle, \mathbf{v} denotes the velocity and the dashed arrow indicates the magnitude v_{\perp} of the velocity in the direction perpendicular to \mathbf{r} . For an isolated system, the angular momentum is a conserved quantity, just like energy or momentum.

The energy and angular momentum are just two examples of quantities which are conserved in physical systems. The term “conservation” in physics means that a physical quantity retains its numerical value when the system evolves in time. In the example we saw above, a stone moves upwards from Earth, reaches a maximum height and falls back to Earth causing its position, speed, kinetic energy etc. to change from instant to instant. But the total energy of the stone does not change during its motion. In other areas of physics, one comes across much more complicated conservation laws leading to other physical quantities (like energy and angular momentum in mechanics) which remain conserved during physical processes. These conservation laws and their discovery are very important in the development of physical theories. Quite often one can characterize the properties of a theory by the quantities which remain conserved in the theory. We shall see several examples of more complicated conservation laws in the later chapters.

1.2 More laws of classical physics

Mechanics, which essentially deals with the motion of terrestrial and celestial bodies, was one of the earliest areas of natural phenomena in which physics made a mark. However, it was known right from antiquity that there are other phenomena in Nature which are not purely mechanical. These observations have also played a crucial role in the development of physics and we will now take a quick look at some of them. Electric and magnetic phenomena probably deserves the centrestage in such a discussion.

Electrical forces, which act between particles carrying electric charge, were known to mankind for a very long time. You probably will recall from your high school days that a glass rod, when rubbed with a cloth will attract small pieces of paper — which is an example of the electric force in action. A body can have positive, negative or zero electric charge. (The last category is called “neutral”.) A positively charged particle will repel another positively charged particle while attracting a negatively charged particle; similarly a negatively charged particle will repel a negatively charged par-

ticle but will attract a positively charged particle. Neutral particles do not feel any electrical force. The force of attraction or repulsion increases in direct proportion with the electric charge of the particle. That is, a particle carrying twice the amount of electric charge will exert a force which is twice as large. The force decreases with increasing distance between the charged particles as the square of the distance between them, just as in the case of gravity.

Like the electric forces, magnetic forces were also known to humanity for centuries. For example, many ancient civilizations knew that a compass needle will stay north-south because of its magnetic properties. For several centuries these electric and magnetic phenomena were treated as separate entities: Certain phenomena required electrical forces for their explanation and others needed magnetic forces. What was not known in the early days was that electric and magnetic forces are essentially the same. It turns out that magnetic force is produced by moving electric charges and is not a separate kind of force caused by some new kind of “magnetic charge”. The moving charges constitute what is usually called an electric current and it is this electric current which produces the magnetic field. Thus one should view the electromagnetic force as a single entity.

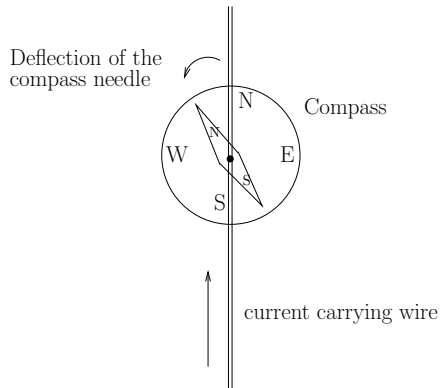


Fig. 1.7 Electric current can produce a magnetic field, as discovered by Ampere. The current which flows through the wire produces a magnetic field that acts on the compass needle and deflects it.

The first two breakthroughs indicating this inter-relationship between electric and magnetic forces came when it was realized (by the scientists Ampere and Faraday, among many others) that electricity can produce

magnetism and vice versa. For example, if you hold a straight piece of wire above a compass needle (which responds to magnetic forces) and send a current through the wire, the compass needle will be deflected. This is a clear proof that an electric current creates a magnetic field around it (see Fig. 1.7). On the other hand, if a magnet is moved in and out of a coil of wire, it generates an electric current which will flow through the coil — which lies behind most procedures to generate commercial electricity (see Fig. 1.8).

These two features, called Ampere's law and Faraday's law, certainly show that electric and magnetic phenomena are closely related and are — in a precise, specific sense — interconvertible. There is, however, one crucial difference. Faraday's law says that a *changing* magnetic field produces an electric field. On the other hand, Ampere's law connects a *steady* current to a magnetic field without the existence of any changing electric field. To see this difference more clearly, note that there are two sources for the electric field: one is just the charged particles (as in the case of a glass rod rubbed by a cloth) and the second is the changing magnetic field, discovered by Faraday. If there has to be some symmetry between electric and magnetic phenomena, we should expect two possible sources for the magnetic field as well. The first is a steady current, which is what Ampere found. The second, which we are now contemplating based purely on symmetry considerations, should be a changing electric field. That is, we would expect a *changing* electric field to produce a magnetic field just as a changing magnetic field leads to an electric field. The first part of the story — electric current producing the magnetic field — was already in place but what was not clear was whether a changing electric field can also produce a magnetic field.

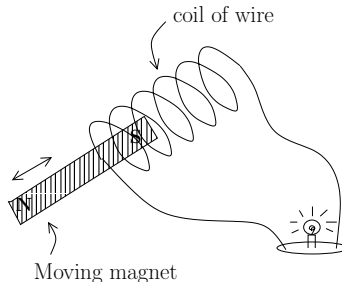


Fig. 1.8 Faraday's law: A magnet moving through a coil can produce an electric current flowing through the coil.

It took the genius of J.C. Maxwell, the British physicist, to assert that it does. Somewhat indirectly, Maxwell was looking for symmetry in the mathematical equations describing the electric and magnetic phenomenon. The set of laws, as available at that time, had one major problem: They were not consistent with the law for conservation of charge. Roughly speaking the conservation of charge means the following: Suppose you find that there is a certain amount of charge in a particular region of space at a given instant of time. If the charges move in or out of this region, you keep tabs on the amount of charge transfer. Since electric current is nothing but moving charge, this is easily done by calculating the electric current flowing through the surface enclosing this region of space. When you count the total amount of charge in the region at a later time, it should be precisely accounted for by the initial amount of charge plus the amount which flowed into the region minus the amount which flowed out of the region. In other words, you cannot create or destroy charge but can only make it move from place to place in the form of an electric current. Since the currents and charges are intimately linked to electric and magnetic fields, you can imagine that any law obeyed by the latter will have something to say about charge conservation. It turns out that to maintain the conservation of charge (which is a fact established by observations) we need to change Ampere's law and introduce the new feature: viz., that a changing electric field can lead to a magnetic field. This is precisely what Maxwell did. The symmetry between electric and magnetic fields was now complete.

Figure 1.9 gives the laws governing the electric and magnetic fields in empty space in modern notation. The symbol \mathbf{E} stands for the electric field and the symbol \mathbf{B} stands for the magnetic field. The other funny symbols in the figure refer to mathematical operations which you are free to ignore! Even without knowing what they are, it is obvious that these equations have a sense of symmetry about the interchange of \mathbf{E} and \mathbf{B} . Except for a minus sign, they remain the same if you replace \mathbf{E} by \mathbf{B} and vice versa. A physicist will — quite correctly — make a song and dance about the beauty in these equations. And in the physicist's eye, beauty is invariably associated with some form of symmetry.

You might think this is all a bit preposterous and wonder why what Maxwell did was such a big deal. "Fine, he made the equations which have weird symbols more symmetric in electric and magnetic fields, so what?" you might ask. It *is* a very big deal; if Maxwell's new term was not there much of the technology we take for granted in the modern world will disappear! The new term led to the existence of electromagnetic waves —

$\nabla \cdot \mathbf{B} = 0$ $\nabla \times \mathbf{B} = \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t}$	$\nabla \cdot \mathbf{E} = 0$ $\nabla \times \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t}$
---	--

Fig. 1.9 The Maxwell's equations written in modern notation. These equations govern the behaviour of source-free electromagnetic field and show the intimate relationship between electric and magnetic phenomena. Even without knowing what the symbols mean, you can see that these equations remain the same, except for a minus sign, if you replace \mathbf{E} by \mathbf{B} and vice versa.

something we use in every form of communication today, to say the least.

Let us try to see how this come about. Suppose you set up an electric field in some region of space, possibly using currents and charges. Now suppose the electric field tends to die down when you switch off the sources. Maxwell's equation — because of the crucial new term he added — will now predict the generation of a magnetic field because of the *changing* electric field. So when the electric field dies down it would have generated a magnetic field; now if the magnetic field dies down, it will generate an electric field due to Faraday's law! So we now open up the possibility of the existence of oscillating electric and magnetic fields, each generating the other and surviving to all eternity! What you now have is an *electromagnetic wave* (see Fig. 1.10) with a life of its own. R.P. Feynman describes this phenomenon as the caterpillar becoming a butterfly!

There is more to the genius of Maxwell. When you work out the maths, it turns out that such an electromagnetic wave will move with a speed that can be calculated from Maxwell's equations and known physical constants. When this is done, one finds that the speed of the electromagnetic wave is precisely equal to the speed with which ordinary light was (known to be) travelling! At that time no one had a clue what light was — and, least of all, no one suspected that light has something to do with electricity or magnetism. So this was a great moment for physics. Not only did Maxwell unify electricity and magnetism into a single structure but — in the process — he explained light as an electromagnetic wave.

Since we will be dealing extensively with the wave phenomena later on,

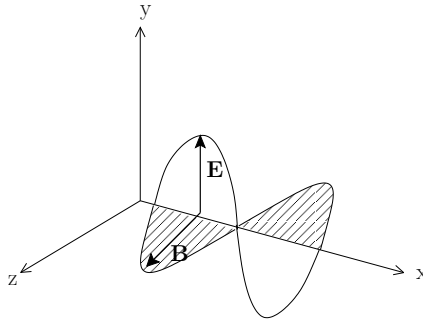


Fig. 1.10 The electromagnetic wave is made of oscillating electric and magnetic fields orthogonal to each other. The wave propagates in a direction perpendicular to both the fields.

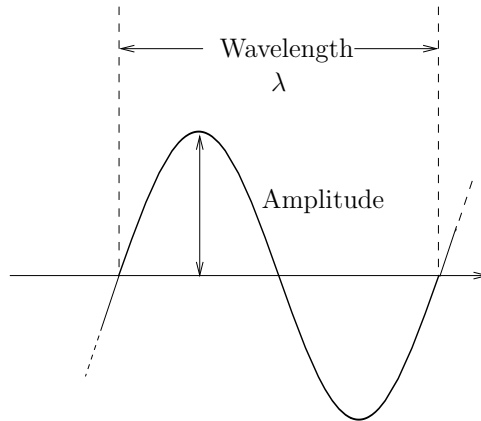


Fig. 1.11 Basic parameters which describe a wave.

this is a good place to get acquainted with some elementary properties of waves. One can usually associate with any wave four different physical quantities which are illustrated in Fig. 1.11. (For the sake of definiteness, let us consider an electromagnetic wave which is essentially an oscillating electric or magnetic field in space; but most of our discussion is applicable to any wave.) The first quantity is the wavelength which is the distance over which the wave property repeats itself; it can be measured as the distance between two consecutive crests (or two consecutive troughs) and is usually denoted by the Greek letter λ ('lambda'). The next is the frequency of the

wave which measures the number of oscillations per second taking place at any given location and we will denote it by the symbol ν ('nu'). The third quantity is the speed with which the wave is propagating forward; in case of electromagnetic radiation, this speed is the same as speed of light, c , — remember that light is a form of electromagnetic radiation. The last physical attribute for a wave is called amplitude which describes how high the crests of the wave rise or the troughs deepen (see Fig. 1.11).

From the definitions, it is easy to see that the first three quantities are related by $c = \nu\lambda$. Since c is a constant, the frequency of electromagnetic radiation will increase when its wavelength decreases and vice versa. The nature of the electromagnetic wave and its physical attributes depend sensitively on its wavelength so that it is often difficult to imagine that they are all the manifestation of the same underlying phenomenon. Among the well known types of electromagnetic radiation the ones with largest wavelength are radio waves without which modern communication will come to a standstill; next comes the microwaves which are produced, for example, in your kitchen oven; then we have infrared radiation by which heat is transmitted to you from your fireplace; next is the normal visible light which our eyes are sensitive to with the wavelength decreasing from red light to blue light; the ultra violet radiation is that which tans your skin on a summer holiday; at still higher frequencies we have the x-rays which your dentist uses and finally, the radiation with the tiniest of the wavelengths called gamma rays which are fairly harmful to our existence. This short summary should caution you that the same physical entity — an electromagnetic wave — can have completely different properties depending on its wavelength or, equivalently, its frequency.

We described above how an electromagnetic wave — once created — has a life of its own and is self-sustaining. The decay of either the electric or magnetic field leads to the other. But that did not answer the question of how we generate an electromagnetic wave in the first place. It turns out that, at the basic level, any charged particle that is accelerated will emit electromagnetic radiation. So if you push a charged particle, making it move with some acceleration, you would have generated an electromagnetic wave. If you now stop the charge from accelerating, it will stop generating electromagnetic radiation but what has already been generated will keep propagating outwards at the speed of light. In more practical contexts (like in a radio transmitter which produces electromagnetic radiation), the acceleration of charges will be more complicated and precisely coordinated but the basic principle is the same.

This is also probably a good place to mention the serious trouble this result causes in describing microscopic, atomic systems. Early experiments suggested that the best description of an atom is as a system with a positively charged nucleus at the centre and electrons (which carry negative charge) orbiting around it. An electron orbiting in a circular path, say, is of course accelerating. (Recall that, in a circular motion the *direction* of the velocity is continuously changing due to the acceleration produced by the force of attraction between the electron and the positively charged nucleus at the centre.) Our theory says that such an electron should be radiating electromagnetic waves and thus will be losing energy. As this happens, the radius of the electron orbit will shrink and the electron will spiral down to the nucleus. Calculations show that this will happen in an extraordinarily small time interval. So, according to classical electromagnetic theory the atoms will collapse in no time and no structure can exist! We will see later in Chapter 3 that this is one of the reasons we need the classical theory to be modified by quantum mechanical considerations.

1.3 Laws of physics in moving cars

Now that we have introduced the laws of mechanics and laws of electrodynamics, it is time to take a look at some of the broader issues related to the nature of physical laws. Let us begin with mechanics. The key feature in Newton's law of motion is that motion with a uniform velocity does not require any force. This innocuous looking statement has some deep implications which we will now highlight.

Consider a car moving on a straight road with constant speed. Suppose you are inside the car and drop a coin vertically down. If the car was not moving, the coin will, of course, fall vertically down on a spot A directly below the point at which it is released (see Fig. 1.12). The question is where it will fall in the moving car: a little behind A, little to the front of A or still on the spot A? At first you might think as follows: "The coin takes some amount of time to hit the floor of the car. During this time, the floor of the car would have moved forward a little bit; so the coin has to fall at a point which is little behind A." This, however, is incorrect and the coin will fall exactly on A, as though the car is not moving.

The reason is as follows: When you release the coin from your fingers, it has a velocity in the forward direction which is same as that of the car. After the release, Earth's gravity acts downwards and gives the coin

a downward acceleration g . This will lead to a component of velocity in the downward direction which will slowly increase with time and the coin will eventually hit the ground. This vertical motion of the coin is the same in both a stationary car and in a moving car. But note that there is no force acting on the coin in the horizontal direction. This means that the horizontal speed of the coin cannot change; in particular it can't be reduced to zero from the initial value it had before you released it from your fingers. So the coin continues to move forward exactly at the same speed as the car, even after it is released! Obviously, the spot A, which is also moving forward because of the car's motion, will always be directly under the coin as the coin falls. Eventually the coin will land at A!

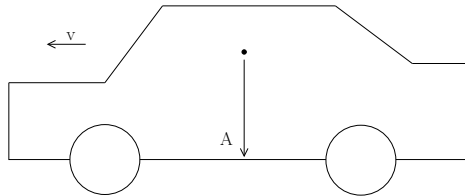


Fig. 1.12 A coin is dropped from inside a car such that it falls on the point A at the floor of the car when the car is not moving. If the experiment is repeated with the car moving with uniform velocity, you will find that the coin again falls at A. If the car was moving forward with an *acceleration*, the situation will change and the coin will fall behind A — that is a little to the right of A in the figure.

If this were not true, you could have determined whether the car is moving with uniform speed or not without looking out of the windows. You first mark the point A when the car is stationary. Then whenever you want to know whether the car is moving or not, you just drop a coin. If it falls behind, you conclude that the car is moving. Our analysis above shows that this procedure won't work. The coin will fall on A whether the car is moving with uniform speed or not. You cannot distinguish between a car at rest and a car moving with uniform velocity by dropping a coin. You may ask: "Fine, dropping coins won't be enough; but is there some other experiment I can perform inside the car to determine whether the car is at rest or moving with uniform velocity?" It turns out that, because forces care only about accelerations and not about velocities, you cannot devise any mechanical experiment inside a car to determine whether the car is moving with uniform velocity or not.

Note that all these comments are only applicable to motion with *uniform*

velocity. If the car is *accelerating* forward, the coin will fall slightly behind A. Now the coin will still move forward with the speed it had when it was released (which is the same as the speed of the car at the instant when you released your fingers) but the car is accelerating forward. So the spot A will move forward with an acceleration while the coin will move forward with only the initial speed it had and no acceleration. Obviously, the spot A will be ahead of the coin all the time; in other words, the coin will now fall behind the spot A.

All these show that motion with uniform velocity is rather special. Physicists like to talk in terms of *frames of reference*, which is just an abstract way of thinking about moving cars. Suppose there is some region of space in which Newton's laws of motion hold and we introduce three coordinate axes (x, y, z) to measure positions of particles. This will constitute what a physicist will call an inertial frame of reference. Consider now a coordinate grid rigidly attached to the car moving with uniform velocity with respect to the original inertial frame. We just saw that such a frame is indistinguishable from the original one. This is usually stated as follows: *Laws of mechanics are the same in all inertial frames*. (This is called the Galilean principle of relativity.) Such a result constitutes what physicists call a symmetry principle or an invariance principle. Broadly speaking if your results do not change when you change some external circumstance, you have a symmetry principle in operation.

Laws of mechanics remain the same in all inertial frame but what about other laws? For example, does the laws of electrodynamics — encoded in Maxwell's equations — remain the same in all inertial frames? Well, the answer turns out to be complicated and — in fact — will form the cornerstone of relativity we will discuss in the next chapter. The laws of electrodynamics are indeed invariant when you go from one inertial frame to another; but it turns out that the *relation* between the two inertial frames needs to be defined in a new manner for this to work. If you use this new definition with laws of mechanics, then it turns out that laws of Newtonian mechanics are *not* the same in all inertial frames! In other words, you cannot define the idea of inertial frames mathematically such that *both* the laws of Newtonian mechanics and electrodynamics remain the same in all inertial frames.

Another way of stating this result is as follows: Let us assume that Newtonian mechanics is correct and that laws of mechanics are the same in all inertial frames. We will then find that laws of electrodynamics are not the same in the inertial frames, defined using the results for mechanics. In

practical terms, this means that you can use electromagnetic phenomenon — like the propagation of light — to determine whether a car is at rest or moving uniformly. People tried to do this using the Earth itself treated as a car moving through space. All these experiments led to a null result! That is, one simply could not determine the motion of a frame using laws of electromagnetism even though theory said this should be feasible. It took the genius of Einstein to change the theory and with it the mathematical relation between the inertial frames. The result was the special theory of relativity which we will discuss in the next chapter.

1.4 Heat as a form of motion

There is another domain of great practical importance in which physics could make significant progress before the twentieth century. This has to do with the the nature of heat and thermal phenomena which, when we look closely, turn out to be quite nontrivial.

The simplest of these examples deals with the properties of bulk matter. If you take a solid body and heat it, its temperature will increase — something even a caveman knew after the invention of fire. The concept of a body being hot or cold and even the idea of putting heat to use (like in, say, cooking) is as old as fire and was known from prehistoric days. What is more, the body can eventually melt and become a liquid if you heat it enough. We are all familiar with a chunk of ice turning into water when heated. If you heat it further, the liquid can vaporise and become a gas. (In the case of water it becomes steam.) Though ice, water and steam have completely different physical properties, it is clear that — in a fundamental sense — they are all the same. For one thing, if you freeze the water obtained by melting ice, you get back ice; but not, say, silver. So there is some kind of a ‘memory’ running through these three states.

The real question is: what makes a hot body hot? To be concrete let us consider two identical blocks of, say, copper, one kept at 30°C and another at 100°C . Anyone can see that the physical effects of these two on the surrounding will be different (just touch them!) but what exactly is different in these two blocks? For a long time people thought that heat is some kind of ‘fluid’ and heating a metal increases the content of this fluid in the metal. If you dip a hot copper rod into a pan of water at room temperature, say, the metal will cool and the water will get heated; eventually both will reach the same temperature. This was thought to be due to the transfer of the ‘heat fluid’ from copper to water.

Very soon experiments showed that such a view leads to several difficulties. Most importantly, this ‘heat fluid’ needed to have all kinds of peculiar properties if it should be able to move invisibly from one body to another and remain in it without altering any of its properties. Further, in this approach, it is not very easy to explain several simple relations that were known about heated gases. For example, consider a gas kept in a container of volume V at temperature T . The pressure P exerted by this gas will vary inversely as the volume of the gas if the temperature is constant. That is if you halve the volume, the pressure will double, when temperature is kept constant. (This is known as Boyle’s law). Similarly, if you double the temperature the pressure will double, if the volume is kept constant (which is known as Charles’ law). One can combine these results to say that $PV \propto T$ for a gas. The real question is how do we understand these laws? Are these fundamental laws we need to add to our collection or can one understand these from something which we already know? Remember that this is the real spirit behind physics. Once we understood the falling apple using gravity we could also explain the motion of Moon around Earth — we did not need one law for the apple and one for the Moon. In the same way, can we understand the thermal behaviour of a gas from something which we already know?

Indeed we can and this correct interpretation of the nature of heat was arrived at by Boltzmann. (Unfortunately his views came under a lot of flak, eventually driving him to suicide; physics is a dangerous vocation at times.) Boltzmann suggested that heat is just random motion of microscopic constituents of a body. We know *today* that all matter is made of atoms or molecules but, of course, this was far too radical during Boltzmann’s days. In a solid, one can imagine these atoms to be arranged in some kind of a lattice structure but these atoms are not stationary. They vibrate about their equilibrium position with a small amplitude. One can therefore associate some amount of random energy with these atoms in the solid lattice. The temperature of the solid is essentially a measure of the average random kinetic energy of the atoms; the larger their random kinetic energy is, the larger is the temperature. If you raise the temperature too much the random vibrations will increase to such an extent that the lattice structure will be destroyed and the solid will melt becoming a liquid. Of course, the atoms still remain in the liquid, with random motions and if you further heat the liquid, their random motion will increase. Eventually, the liquid will vaporise into a gaseous state (like water becoming steam). In

the gaseous state, each atom (or molecule) will again have random motion which will continue to be a measure of its temperature.

As an aside, let us mention that the lowest energy state for any matter occurs at around -273 degree centigrade and it is impossible to cool any body to temperatures lower than this value. (To cool a body, we have to extract the heat energy from the body; which is impossible if it is already in the lowest energy state). In physics, it is often convenient to use a different unit so that the lowest temperature corresponds to zero degree. This unit is called “Kelvin” and is defined so that zero degree Kelvin (0 K) corresponds to -273 degree centigrade. To obtain the temperature in Kelvin, we merely have to add the number 273 to the temperature expressed in centigrade. Thus the temperature -183 degree centigrade (at which oxygen becomes a liquid) corresponds to $-183 + 273 = 90$ Kelvin.

Boltzmann’s insight was to reduce the — until then mysterious — thermal phenomenon into one of mechanics. If a gas is made of zillions of randomly moving molecules and the temperature of the gas is linked to the random kinetic energy of motion, then the study of thermal properties of the gas reduces to the study of the random motion of the molecules. That is, thermodynamics now becomes part of Newtonian mechanics. Since we will be dealing with a huge number of molecules at any given time, the emphasis shifts from describing the trajectories of each molecule or atom to describing the statistical properties of the system. Hence this branch of physics, which originated from Boltzmann, is often called *statistical mechanics*. It tells you how to link up the bulk properties of matter — like, for example, the pressure of the gas — to the microscopic motion of the molecules.

Just to see how this works, let us consider pressure exerted by a gas of N molecules kept in a cubical box of size L , using nothing more than high school algebra. (If it scares you, skip ahead to next paragraph!). We will concentrate on the side of the box perpendicular to the x-axis. The gas exerts a pressure on the wall of the box because the collision of the molecules with the wall exerts a force on the latter. Let us assume that each molecule is moving randomly with an average speed v . In a time interval t , the molecules can travel a distance vt on the average; so all the molecules which are at a distance less than vt from a wall will collide with a wall in the time interval t (see Fig. 1.13). This number is given by the molecules which are inside a volume $(1/6) \times vt \times L^2$ where the factor $1/6$ takes into account the fact that only a third of the molecules will be moving along the x-axis and only half of these will be moving *towards* the

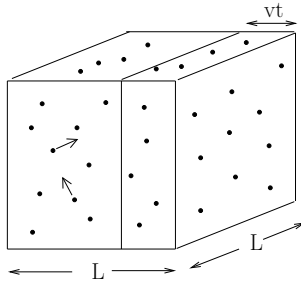


Fig. 1.13 One can compute the pressure exerted by a gas from Newton's laws of mechanics, if we assume that the gas is made of a large number of molecules in random motion. This hypothesis, made by Boltzmann, interpreted heat as a form of motion and essentially unified the laws of thermodynamics with laws of mechanics.

wall. Since there are (N/L^3) molecules per unit volume, this number is $(N/L^3) \times (1/6)vtL^2 = Nvt/6L$. On colliding with the wall the molecules will bounce back so that the momentum of each one changes from mv to $-mv$; that is by an amount $2mv$. The total change of momentum for all the molecules is $(Nvt/6L) \times 2mv = (N/3L)mv^2t$. Since this occurs in the time interval t , the rate of change of momentum, which is the force exerted on the wall by the molecules, is obtained by dividing this expression by t giving $F = (N/3L)mv^2 = (2N/3L)K$ where $K = (1/2)mv^2$ is the kinetic energy of each molecule. The pressure of the gas is defined as the force it exerts per unit area of the surface; here the area that is involved is L^2 and hence the pressure is $P = F/L^2 = (2/3)K(N/V)$ where $V = L^3$ is the volume of the box. This 'law' can be written as $PV = N(2/3)K$ which relates the pressure of the gas, volume of the gas and the mean kinetic energy of the molecules. It was known, on the other hand, that for a gas with fixed number of particles, the product PV is actually proportional to its temperature. Boltzmann's insight was in identifying the notion of temperature with the mean kinetic energy of the particles and suggest $K = (3/2)k_B T$ where T is the temperature and k_B is called — quite appropriately — Boltzmann's constant. This direct relation between kinetic energy of molecules and the temperature allows us to write our earlier relation as $PV = Nk_B T$ which, some of you might recognize as the equation describing an ideal gas.

This is again something truly remarkable and one needs to savor the moment. In thermodynamics, one would have described a gas in a container without ever worrying about whether it is made of molecules or not. The gas can be a continuous substance with no substructure and we can still de-

fine its pressure, volume etc. and study its properties. In this case, however, we will have no fundamental understanding of why the product PV for a gas is proportional to its temperature; we need to postulate it separately as a new law. Boltzmann's idea was that if we assume the gas is made of large number of particles and we use for these particles the usual laws of mechanics (recall that we just computed the rate of change of momentum due to collisions with the walls) then we can describe how such systems behave. In the process, we realize that temperature has a simple interpretation as the average kinetic energy of the molecules and we no longer need to introduce any mysterious 'heat fluid'. Just as Maxwell's unification allowed one to understand light as an electromagnetic phenomenon, Boltzmann's work allows us to understand thermal phenomena as a part of Newtonian mechanics.

Given all these, the current picture of bulk matter can be understood as follows. Consider, for the sake of definiteness, a solid piece of ice with which you are quite familiar in everyday life. Ice, like most other solids, has a certain rigidity of shape because, in a solid, atoms are arranged in a regular manner. Such a regular arrangement of atoms is called a "crystal lattice" and one may say that most solids have "crystalline" structure (see Fig. 1.14). Atoms, of course, are extremely tiny and they are packed fairly closely in a crystal lattice. Along one centimeter of a solid, there will be about one hundred million atoms in a row. The standard notation for this is to say that there are 10^8 atoms along one centimeter of ice. (The notation 10^8 stands for ten multiplied eight times, which is a number with 1 followed by 8 zeros.) This means that the typical spacing between atoms in a crystal lattice will be about one part in hundred millionth of a centimeter, i.e., about $1/100,000,000$ centimeter. This number is usually written as 10^{-8} cm. The symbol 10^{-8} , with a minus sign before 8, stands for one part in 10^8 ; i.e., one part in 100,000,000. (In this notation — which we shall use frequently — one tenth will be 10^{-1} , one hundredth will be 10^{-2} etc. To denote two parts in a thousand, we will write 2×10^{-3} . Note that 3×10^3 stands for 3000 while 3×10^{-3} stands for 3/1000, viz. three parts in a thousand.) To describe the phenomenon at atomic scales, we have to use such small lengths and it is useful to introduce new units to describe very small lengths. The length, 10^{-8} cm, is a convenient unit and is called one Angstrom (1\AA , for short). So the separation between the atoms in a crystal is about one Angstrom. A small cube of a solid, 1 cm in the side, will contain about 10^{24} atoms. It is because of this tiny separation between the atoms that solids appear continuous and people were skeptical when

Boltzmann introduced the substructure to matter.

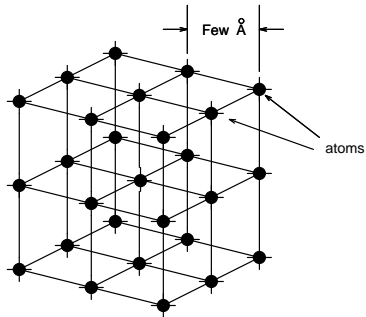


Fig. 1.14 Many solids are made of a regular array of atoms in the form of a lattice. The figure shows the ‘crystalline structure’ of such a simple solid in which one atom is located at each vertex of a cube. Much more complicated, but periodic, arrangements of atoms are possible in a crystal. The atoms are held in place by electromagnetic forces and the typical separation between the atoms is a few Angstroms.

What keeps these atoms in such a regular array in a solid? Incidentally, it is the electromagnetic forces which determine the overall structure of matter and the force between atoms in a crystal lattice is of electrical nature. Even though individual atoms are electrically neutral, the currents produced by the orbiting electrons inside the atom exert a force on the nearby lattice atom. This force is the cause of the binding energy of a crystal lattice. You know from experience that it takes fair amount of effort with an ice pick to break an ice cube into smaller pieces. In this process, one has overcome the forces that hold the crystal together by supplying energy to the block of ice.

What will happen if we keep supplying more and more energy to a solid body? At first the atoms in the crystal lattice will vibrate with larger and larger amplitude and soon — if we supply sufficient amount of heat energy to a block of ice — it will melt and become water. Heating provides enough energy to break the binding between the atoms in a crystal lattice, thereby overcoming the rigidity of the ice. The same phenomenon takes place in any other solid. As we heat a solid — and supply more and more energy — there occurs a moment when the solid melts and forms a liquid. Different solids, of course, will require different amount of heat energy to be supplied for their melting; it is comparatively easier to melt an ice cube than a piece of iron. It takes about 80 calories of energy to melt one gram of ice.

This quantity, “calorie”, is a unit for measuring the energy. (It is related to the unit you keep hearing in connection with exercise and diets, all of which have to do with supplying or burning of energy.) In physics, it is also usual to use several other units for energy, of which the most frequently encountered one is “erg”. One erg is about 2.38×10^{-8} calories; so we may say that it takes 3.4×10^9 ergs to melt one gram of ice. The rigidity of ice is then lost and water “flows” easily.

We can, of course, continue the above experiment by supplying more and more heat energy to the liquid. At first, the water will just get hotter and hotter. Soon, another change will occur and the liquid will become a gas. That is, water will change to steam. You must have noticed that steam is nearly invisible, compared to water. This is because the atoms forming the steam are lot more sparsely distributed compared to the atoms in water. [One cubic centimeter of steam will have nearly a thousand times less atoms than one cubic centimeter of water]. The gaseous phase is also common to any other solid; by supplying sufficient energy to molten iron we can produce a gaseous vapour of iron atoms.

This description covers three of the most familiar states of matter: solid, liquid and gas. In nature, there is also one more phase which occurs at sufficiently high temperature, called the plasma state. When you supply sufficient amount of energy, one can rip apart the electrons from the atoms leading to positively charged ions and negatively charged electrons both of which exists in a state of random motion. This, of course, involves some amount of atomic physics and hence will become clearer when we describe the structure of the atom in Chapter 3.

1.5 Entropy: A new beast

In the last section we described how the behaviour of bulk matter, like a gas of hydrogen atoms or a crystal lattice, can ultimately be understood in terms of the laws of microscopic physics. We illustrated this by showing that the ideal gas law which governs the behaviour of gaseous systems can be obtained by applying the simple laws of Newtonian mechanics to the molecules which are randomly moving in a volume of space. Similar considerations apply to more complex systems and one can always obtain the behaviour of macroscopic systems from the laws of microscopic systems.

This fact suggests that macroscopic bodies do not bring in any new phenomena which are not inherently present at the microscopic level. While

this is true at some fundamental level, it is not of much practical value in studying the physics of macroscopic systems. Very often, it is convenient to introduce concepts in macroscopic systems which have no interpretation at the microscopic level. Some of these concepts, as it will turn out, play a crucial role in our discussions in later chapters. In view of this, we shall take a closer look at such concepts before proceeding further.

To illustrate the ideas in a simple context, let us again consider a gas made of, say, hydrogen molecules confined in a large volume V . For any macroscopic volume, the number of molecules contained in it will be enormous; for example, in one litre of hydrogen gas under normal room temperature and pressure, there will be about 10^{22} molecules. Each of these molecules will be moving randomly with some average speed. The temperature of the gas — as we described in the previous section — is essentially a measure of the average kinetic energy associated with the random motion of these molecules. Obviously, one needs a large number of molecules in order to define a useful notion of an average. This is, of course, true in any situation in which one wants to introduce the concept of an average. For example, to compute the average monthly income of the people in a city we add up the total income of the people in the city and divide it by the number of people living there. This makes sense only when the number of people in the city is sufficiently large. If there are only 2 or 3 persons living in a hypothetical city, you can still compute the average but it may not mean much if the income of the 3 people differs widely. In a similar manner, the concept of temperature in a gas makes sense only when large number of molecules are involved and there is no such thing as a temperature of an individual molecule. Therefore, temperature is the first example of a physical concept which makes practical sense when large number of particles are involved but cannot be defined usefully at the microscopic, molecular level.

Similar comments apply to even the idea of density of a gas — or even that of a solid. When we say that the density of iron is 7.87 gm/cm^3 we refer to a piece of iron containing a large number of iron atoms. Most of the region inside the solid however is empty and — in fact — even most of the region inside the atom is empty. The mass of an atom is entirely concentrated on the nucleus which has an extraordinarily high value of density. So we again see that the notion of 7.87 gm/cm^3 as the density of iron is a completely macroscopic concept and cannot be defined usefully at the microscopic level.

These two examples illustrate the point but they are, in some sense,

easy to understand and not surprising. We shall next introduce a concept called *entropy* which plays an important role in virtually all areas of physics and has some counterintuitive aspects associated with it. Instead of giving a mathematical description, we shall try to introduce it in terms of simple examples.

To do this we need to bring in the concepts of microscopic state (or *microstates*, for short) and the macroscopic state (or *macrostate*, for short). Consider an array of N light bulbs located in a line each of which can be switched on or off. Suppose you switch on a set of them, say, the 1st, 3rd, 17th, 123rd and tell me which of them are turned on and which of them are off. This completely specifies the microstate of the system in this case. You now know precisely which bulb is on and which bulb is off. Each of the bulbs can exist in two different states (on or off) and hence the total number of possible states of the system is $2 \times 2 \times \dots N$ times $= 2^N$. This is a very large number even for reasonable N . For example, for $N = 270$, the total number of possible states for these bulbs is larger than the number of atoms in the observed universe.

Let us now consider a slightly different situation. You again switch on certain number of bulbs (say, k bulbs) but do not tell me which bulbs are switched on. (Alternatively, you can imagine that I measure the total amount of illumination produced by the array of bulbs from some distance and figure out that k bulbs must be switched on; but I have no way of accurately determining *which* k bulbs are switched on.). I now have less information compared to the previous situation in which you told me the exact state of each bulb. The information I now have — viz. that some k bulbs are switched on — specifies the macrostate of the system. The key point is that, there are large number of microstates which are consistent with the information I have about the macrostate. Suppose $k = 2$; that is I only know that 2 bulbs out of N are switched on. The first of these could be any one of the N bulbs. Having fixed that, for each of the choices, the second bulb could be any of the $(N - 1)$. So there are $N(N - 1)$ possible ways of doing this. But the final state of the system in which I first switch on 3rd bulb and then switch on 17th bulb is the same as the one obtained by first switching on the 17th bulb and then switching on the 3rd. So only half of the $N(N - 1)$ possibilities are really different. We thus conclude that there are $(1/2)N(N - 1)$ distinct microstates of the system which lead to the same macrostate in which two bulbs are switched on. The macrostate contains considerably less information than the microstate.

This lack of information is described by a quantity called entropy. If

a given macrostate can arise due to 2^S (that is 2 multiplied S times) microstates, then S is called the entropy of that macrostate. These concepts can be applied to any system with large number of subcomponents (like the bulbs in the above example). For example, in a gas made of N particles, the microstate is specified by the positions and velocities of each of the molecules. But usually we specify the state of such a gas by giving its total energy E and the volume V it occupies. There are ways of computing the total number of possible microstates (positions and velocities of the molecules) which can lead to the same E, V and from this one can determine the entropy of a gas S as a function of its energy and volume. It turns out that once we know this, one can determine all other physical properties of the gas.

There is something very curious about the behaviour of entropy in any physical system which makes it such a useful and important concept. It turns out that when macroscopic systems interact with each other, the entropy always increases or, at best, stays constant. You have to arrange things in a special manner and spend energy in order to decrease the entropy of any given system. Even when you do that, it will turn out that the entropy of some other part of the external environment, which is coupled to the system we are working with, goes up. When you calculate the net change in the entropy of the external environment as well as the system we are interested in, it will turn out that the total entropy has increased. This result — that the entropy keeps increasing during physical interaction — is called the second law of thermodynamics and is one of the key principles of macroscopic physics.

To see how this comes about, let us consider a simple example shown in Fig. 1.15. The part (a) of the figure shows a box of volume V which is separated into two halves by a partition in the middle. On the left half, we have a large number of molecules of a gas at some temperature and pressure in random motion while the right half is initially empty. This system has some amount of entropy which, of course, will depend on the volume of space which is available to each molecule. In any given microstate of the system, each of the molecules could be found anywhere in the volume $V/2$ and just specifying the macrostate by giving the pressure, temperature and the volume of the region occupied by the gas does not give any further information about the microstate. Let us suppose we now remove the partition at a given instant of time. At the next moment, we will have a situation as described in Fig. 1.15(b). All the molecules are still in the left half of the

volume but, of course, they are now free to move throughout the volume. In fact, this is precisely what will happen if we wait for a little while. The gas will expand and very soon the molecules can be found anywhere in the volume V . Each of the molecules can now be found within a larger volume compared to the situation shown in Fig. 1.15(b). Clearly, the number of possible microstates available to the system has gone up in the process and hence the entropy of the gas has gone up.

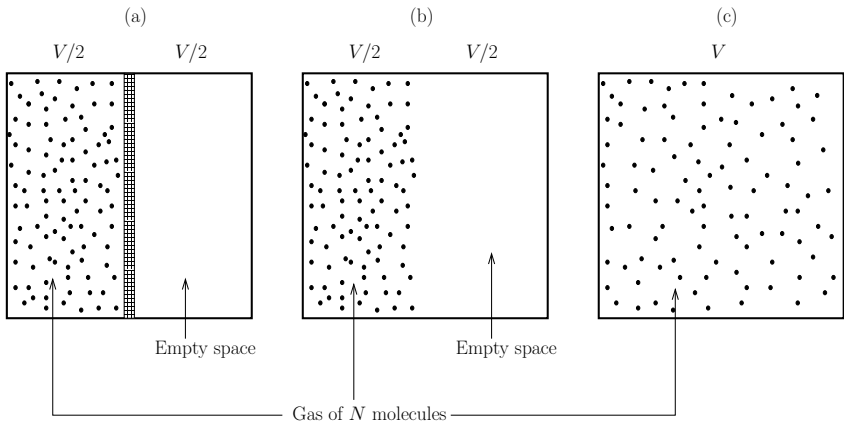


Fig. 1.15 Gas molecules, which were initially confined to the left half of a volume of a box (in Part (a)), are allowed to spread to the full volume by removing the partition. The situation just after the partition is removed is shown in Part (b) and the situation at a much later time is shown in Part (c).

There are two points that need to be stressed regarding the above — relatively simple but instructive — example. First, there is a connection between the entropy and our lack of information about the systems. Let us say that there were k possible microstates available to the system when the gas is confined to the left half of the volume. Our ignorance about which microstate the gas is actually in will be mathematically described by the entropy associated with this state. When the partition is removed and the gas fills the full volume of the box, the total number of available microstates has gone up to, say, k' with $k' > k$. This is what has led to the increase in entropy. But we also have less information about the system now. Originally we knew that the gas is in one of the k possible states, but now the gas is in one of the k' possible states. Since $k' > k$, our information about the system has actually come down.

If you have difficulty in grasping this concept, just compare it with a common situation. Suppose you know that your friend has an office in the 14th floor of a building which has 20 floors with each floor having 10 offices. Then you know that she is in one of the 10 offices in the 14th floor. But suppose you only know that she works in that building, but do not know which floor her office is. Now she could be in any of the $20 \times 10 = 200$ offices. You will certainly agree that when the number of microscopic possibilities goes up from $k = 10$ to $k' = 200$, the amount of information you have about the friend's whereabouts has come down. This shows that the entropy of a system is closely linked to our lack of information about the different possibilities which exists in a given context. This idea will play an important role in a Chapter 6 when we study black holes.

The second feature which is immediately evident from Fig. 1.15(b) and Fig. 1.15(c) is the following. Suppose we were shown these two pictures as the snapshots of the molecules of the gas at two different instants of time and were asked to decide which snapshot was taken earlier and which was taken later. We know that if a large number of gas molecules were confined to the left half of the volume of a box at a given instant, then over a period of time the gas will expand and fill the entire volume. So, clearly, the snapshot in Fig. 1.15(c) was taken at a later time compared to the one in Fig. 1.15(b). In other words, the direction of increase of entropy in some sense allows us to determine the direction of flow of time. One usually says that the above example provides a thermodynamic arrow of time. So, entropy is connected in a mysterious way with the deep question of what determines the irreversible flow of time which we all experience.

The existence of such an arrow of time raises another interesting question. All the fundamental laws of physics — like, for example, the laws of mechanics — remain the same if we change the sign of time; that is, if we interchange past and the future. Suppose we throw a ball up in the air and take snapshots of the trajectory of the ball during a given time interval. If you are shown the sequence of snapshots, there is no way you could have determined in which sequence the snaps were taken. This is because, for any given trajectory of the ball, there exists another equally valid trajectory in which the ball passes through the same spatial location in the reverse order. What is true for a ball is also true for any single molecule which is located on the left half in Fig. 1.15(b). If you were only given the trajectory of any given molecule at different instants in time, you cannot arrive at an arrow of time from it. For any given trajectory there

is another possible trajectory which is obtained by reversing the direction of the time. In spite of this, if you are given the behaviour of very large number of molecules — as in the case of Fig. 1.15(b) and Fig. 1.15(c), one can immediately determine the direction of time.

That is definitely peculiar and people have investigated this thoroughly. The crucial question one needs to answer is the following: Suppose we start with the situation shown in Fig. 1.15(c) with the molecules moving randomly in all directions. Since the motion is random, isn't it possible that after sometime all the molecules end up simultaneously in the left half of the volume thereby reaching the configuration shown in Fig. 1.15(b). Since the laws of microscopic physics are invariant under time reversal, this is certainly possible, in principle. All we need to do is to start with a situation in which the positions of the molecules are as in Fig. 1.15(c) with the direction of motion reversed; then the molecules will end up at the configuration Fig. 1.15(b) after some time. The right question to ask, however, is how probable is it? From everyday experience, we know that gas molecules do not behave like that. (Imagine what would happen if all the oxygen molecules in your room decide to move away from you and stay in an opposite corner for a few minutes!) If there are N molecules in the gas and the motion is truly random, any given molecule has an equal chance to be found either in the left half or in the right half of the box. So the probability that any given molecule is in the left half is $(1/2)$. The chance that all the N molecules end up in the left half simultaneously at any given instant is $(1/2) \times (1/2) \times \dots N \text{ times} = 1/2^N$. For any reasonable N , this is an extremely tiny number and hence the probability of it occurring is utterly negligible. If you take one litre of gas, one can be quite confident that such an event is not going to take place within the entire lifetime of the universe and in fact our intuition — for once, correctly — has grasped that, for all practical purposes, we can say that a gaseous system will not proceed from the state described in Fig. 1.15(c) to the state described in the Fig. 1.15(b). In short, even though such a transition is allowed by the laws of physics, it is an event of such a low probability that we can assume that it does not occur.

If the molecules in the gas proceed spontaneously from the state in Fig. 1.15(c) to the one in Fig. 1.15(b), then the entropy of the gas will decrease during this process. Our conclusion that such an event does not take place *for all practical purposes* shows that the entropy of the gas does not decrease spontaneously for all practical purpose. You see that the laws of thermodynamics have brought in certain probabilistic consideration into

physics unlike the fundamental laws of microscopic physics. The increase of entropy during physical processes occurs with such an extremely high probability that one can take it to be a certainty.

1.6 A peep ahead: Where did it go wrong?

It turns out that the ideas of statistical mechanics has far reaching impact and — with future use in mind — we shall describe some of these. To begin with, let us take a closer look at the relation between the kinetic energy K and the temperature T given by $K = (1/2)mv^2 = (3/2)k_B T$. A gas molecule can move in three different directions in space and one says that the molecule has three degrees of freedom. The above relation shows that one can associate a thermal energy $(1/2)k_B T$ with each degree of freedom. This, of course, only takes care of the translational kinetic energy arising due to the entire molecule moving as a unit. If the gas is made of diatomic molecules, say, then there can be other sources of energy. For example, the two atoms of a diatomic molecule can vibrate along the axis joining them and the vibrational mode stores some amount of energy. This adds another degree of freedom in which energy can be stored. A diatomic molecule can also rotate in the plane of the two atoms which allows for yet another way of storing energy. A more complex molecule can vibrate and rotate in more complicated ways so that it will have more degrees of freedom. A key result in classical statistical mechanics, as formulated by Boltzmann, is that when a system is in thermal equilibrium, each degree of freedom can store $(1/2)k_B T$ of energy. So if we have a gas with N molecules with f degrees of freedom per molecule, then it will have an energy $(f/2)Nk_B T$. In other words, to increase its temperature by one degree, we need to supply an additional energy of $(f/2)Nk_B$. This quantity — the energy that needs to be supplied to increase the temperature by unity — is called the *specific heat* of the system and can be directly measured. We now have a clear prediction about the specific heat of different kinds of gases (monoatomic, diatomic etc.) and one can compare the predictions with observations.

When this was done, the results were surprising. One found that, at sufficiently high temperatures, the results agreed with observations while at lower temperature there were serious discrepancies between the observations and theory. It was as though some of the degrees of freedom were not participating in the energy sharing process when the temperature was low. This was one of the first indications that the energy of microscopic

systems, like molecules, do not behave in the expected manner. We will see in a Chapter 3 that this discrepancy was actually one of the factors that led to the quantum revolution.

There was another — and more striking — discrepancy which statistical mechanics brought into classical physics. To understand this, we need to first introduce the concept of thermal radiation. This is the electromagnetic radiation emitted by any hot body due to its temperature. You know that when an iron rod is heated to a high temperature it emits light (which is a form of the electromagnetic radiation) of predominantly red colour; we say that the iron is “red hot”. This is merely a particular case of a very general phenomenon. When any material object is kept at some nonzero temperature, it emits radiation of a particular characteristic which depends only on the temperature.

A simpler way of thinking about this radiation is as follows: Consider a closed metallic box, the surface of which is kept at some fixed temperature. Then the inside of the box will be filled with electromagnetic radiation with a characteristic pattern in energy density which depends only on the temperature of the box. More precisely, the amount of radiation energy which is in the box at a given frequency can be expressed entirely in terms of the temperature of the box. If you put a small hole on the box and let a little bit of radiation to leak out, we can figure out how much of radiation energy is there at any given frequency. If you make two boxes of different shapes, sizes and made of different metals and keep them at the same temperature, then the radiant energy inside the boxes will have identical form. The question arises as to whether we can understand this phenomenon from theoretical considerations.

We saw earlier that electromagnetic radiation is essentially due to oscillations of electric and magnetic fields and these oscillations can be thought of as different degrees of freedom associated with the oscillating electromagnetic field. Each mode of vibration is capable of storing energy. Using our rule that each degree of freedom should store an energy $(1/2)k_B T$ and counting the number of degrees of freedom associated with a vibration at frequency ν , one can compute the amount of radiation at a given frequency. Such a calculation, originally done by Lord Rayleigh led to the result that, the amount of energy U at a frequency ν should vary as $U \propto \nu^2 T$ at temperature T . This was a rigorously derived result from theory; the trouble is, it is an absurd result!

To see its absurdity, you need to note that the result claims that the energy increases as square of the frequency at any given temperature. So

if you keep a metallic box at room temperature and peep into it, most of the radiation that hits you will have the highest frequencies; lots of X-rays, gamma rays etc. This is manifestly wrong. Further, the results show that any box at any nonzero temperature has an infinite amount of total energy! If you add energies at all the different frequencies you get an arbitrarily large value — mathematically speaking, infinite value — because there is more and more energy at higher and higher frequencies.

While the theoreticians were grappling with these absurdities, the experimentalists were also at work. They found out how exactly a box at some temperature behaves. Of course, nothing strange was happening in real life in which the radiation inside the box had the form in Fig. 1.16. The radiation does contain a wide band of frequencies but observations show that it will be peaked at some frequency with very little radiation at much higher frequencies. The frequency at which one has maximum radiation, increases with temperature of the box. Correspondingly, the wavelength of the peak decreases with the temperature. Bodies at room temperature (which corresponds to 300 K) will emit most of the radiation in the infrared band. If the temperature is increased to about 10^4 K, the radiation will be mostly in the visible range and if we heat the system to about 10^8 K there will be copious production of X-rays. The amount of radiation emitted at different frequencies by a body at some temperature, is called “thermal spectrum” or “Planck spectrum”.

At low frequencies the law $U \propto \nu^2 T$ discovered by Lord Rayleigh was indeed true; the energy keeps increasing with frequency up to a point; but then it starts to decrease rapidly at higher frequencies. At any given temperature the maximum amount of energy was at some frequency which rises in proportion with the temperature. The total amount of energy in the box was, again, finite and increased as T^4 . If you doubled the temperature, the total radiant energy went up by a factor 16 etc. Clearly, the theory falls in its face trying to get the results right!

So what is going wrong? Again the result has to do with quantum theory which we will take up in Chapter 3 but the key point is the following. The results described above assume that the radiation is in equilibrium with matter which makes up the metallic box — that is, the amount of energy the box was absorbing from the radiation should be equal to the amount of energy the box was emitting into the box. In doing the calculation it was assumed that the matter which makes up the box can exchange energy with the radiation by arbitrary amounts. Remember that if the box is kept at some temperature, it will make the atoms of which the box is

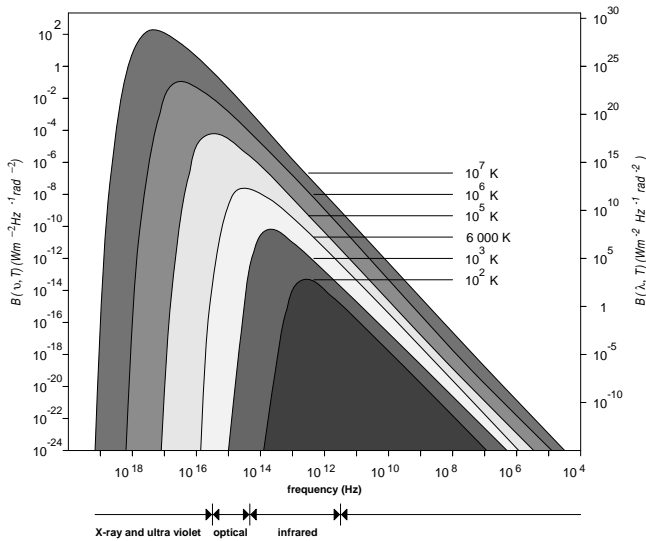


Fig. 1.16 Any object kept at some temperature emits a characteristic spectrum of radiation called the Planck spectrum. As the temperature is increased, most of the radiation will be emitted at shorter and shorter wavelength. Note that, in the figure, the frequency *decreases* to the right. The figure shows the nature of emission for bodies kept at temperatures ranging from 100 K to 10^7 K. Bodies at room temperature (300 K) emit most of the radiation in the infrared. Bodies at higher temperatures, say, between 5000 K to 7000 K (as in stars) emit large amount of visible radiation.

made of, vibrate in a characteristic manner. Assuming that the energy of these vibrations can take all possible values is equivalent to saying that the box can exchange energy with radiation by arbitrary amounts. Quantum theory changes this result and leads to the conclusion that the energy of vibration of the atoms in the box is quantized. This means the radiation can be exchanged only in quantized units. When you redo the calculation with this assumption, you get the correct result; the theory reproduces the experimental result.

It turns out that the same assumption — that the energy is quantized — allows one to understand the strange behaviour of the specific heat as well. At low temperatures, there is not sufficient energy to excite the modes of vibration and hence lots of degrees of freedom do not participate in sharing of energy or contributing to the specific heat. This cannot arise if the

vibrational energy varies continuously (when one can induce vibration with arbitrarily small amount of energy) but can occur if it is discrete.

In summary, statistical mechanics forms the link between microscopic and macroscopic physics by linking the properties of bulk matter to its constituents, like atoms or molecules. In doing so, it attempts to use the classical laws of physics (like Newton's law) to describe the behaviour of atoms and molecules. It works in some cases (like in describing the pressure of a gas) but fails miserably in some other cases. The failure requires us to go beyond the ideas of classical physics — and as we have said before — this leads to the quantum revolution.