CHAPTER 6

NEW FRONTIERS IN PHYSICS

The search for truth: "At early times the magicians used to control people through their knowledge of some scientific facts like Sun and moon eclipses. At present, scientists are no magicians .. They analyze facts for the use and benifit of human kind"

A. SUPERCONDUCTIVITY

Ever since the discovery of **superconductivity** by Onnes in 1911, its unusual scientific challenge and great technological potential have been recognized. The phenomenon of superconductivity is very exciting because of its many technical applications. The discovery of high temperature superconductivity in complex oxides of rare-earth metals with layered perovskite structure has areas. This discovery is thought to be greater than the invention of the transistor, since it might over-throw all the already well-known electrical technology and replace it with another kind of technology that would not need a continuously operating power supply for our electrical machines. For this reason it is very important that all students of science and engineering and those who are not majoring in physics, should know something about the basic electromagnetic properties of superconductors, and become aware of the scope of their current applications which they will surely use after their graduation. In the following section we give some of the well established applications of superconductivity.

1. Power energetics (superpowerful generators and transformers, transmission lines, and electric energy storage).

Presently, generators, transformers and electric transmission lines waste 10-20% of the electric energy in the form of heat. The application of superconductors with zero resistance will make it possible to lower these losses by hundreds of times, and to remarkably miniaturize the generators and transformers. Solenoidal energy storages of vast capacity is created. These miniaturized storage systems will be able to store up to 10^5 kW.h of energy in a volume of 1 cubic meter.

2. In mechanical engineering, levitation could be exploited in the field of transportation. Magnetically levitated trains are now a reality. It is also possible that cars and ships can use the same phenomenon. The frictionless train, sometimes called the bullet train, already operated in Japan and other countries, applies superconductor frictionless bearing shown in Fig. 6.1.

The train has superconducting magnets built into its base. A powerful magnetic field both levitates the train a few inches above the track, and propels it smoothly at speeds of more than 300 km per hour. One can envisage the future society to have all sorts of vehicles gliding above a free way making use of superconducting magnets. It is expected that the first major market for levitation effects will be in toy industry.

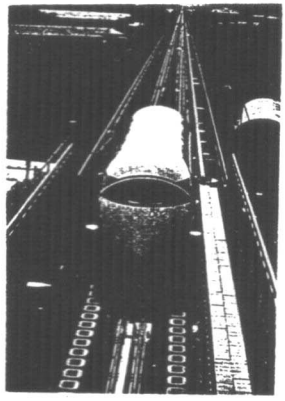


Fig. 6.1. This prototype train, constructed in Japan, has superconducting magnets built into its base. A powerful magnetic field both levitates the train a few inches above the track, and propels it smoothly at speeds of 300 miles per hour or more.

3. One important application that is presently used in the field of diagnostic medicine is the recording of "brain waves". The neurons carrying currents in the brain are associated with very weak magnetic fields. Similarly, the small currents which flow in the heart cause biomagnetic fields that could now be easily measured by a superconducting device called the **SQUID.**

4. When superconducting films are used to interconnect computer chips, chip size could be reduced and speeds would be enhanced because of the small size. Information are transmitted more rapidly and more chips could be contained on a circuit board with far less heat generation.

5. Superconductors are used in the production of very high magnetic fields that are used in particle accelerators.

6. The highly sophisticated technique called magnetic resonance imaging make use of superconductors and rf-radiation to produce images of the body sections, instead of using X-rays.

7. Many other applications are established, such as very sensitive magnetometers, digital applications, study of brain disorders and epilepsy.

In the following part we give a brief account, and highly simplified, about the strange properties of superconductors. The professional student, who will take physics as a complementary subject, will find it very useful particularly after his graduation when he has to deal with modern equipment based on such recent discoveries in physics.

A.1. Superconductivity and the quantum behavior of matter

At extremely low temperatures, materials develop very unusual properties. Many metals and alloys become superconducting, i.e. their resistance to electric currents becomes completely zero. Liquid helium near the absolute zero becomes a superfluid, i.e. its internal friction disappears and it can flow without drag through very fine holes. These strange properties are a manifestation of a high degree of order within the material. At absolute zero the entropy vanishes according to the third law of thermodynamics. Disorder due to thermal agitation disappears. The material acquires a microscopic state that cannot be described by classical mechanics. Superconductivity and superfluidity are macroscopic manifestations of the quantum behavior of matter.

A.1.1. Discovery of superconductivity

Onnes and his school in Holland, were studying the resistivity of metals at low temperatures. Platinum showed some residual resistivity when extrapolated to absolute zero, depended on the purity of the sample. Mercury which could be obtained in a very

pure state by evaporation and condensation (distillation) showed a sharp drop in its resistance at 4.15 K to a zero value. The results obtained are shown in Fig. 6.2.

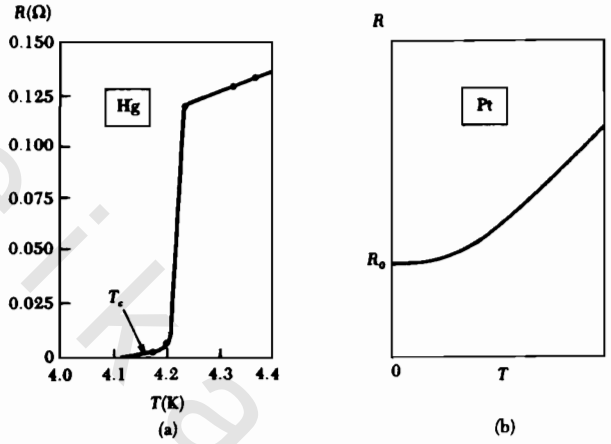


Fig. 6.2. Resistivity of mercury as a function of temperature. Below 4.15 K the resistivity drops to nearly zero, while above this temperature it follows the path of a normal metal similar to the curve for platinum showed in a dashed line.

After the discovery of Onnes, many other metals were found to exhibit zero resistance when the temperature was lowered below a certain characteristic temperature of the material called **the critical temperature, T_C.**

A.1.2. Properties of type I superconductors

The elemental metals that showed superconductivity are called type I superconductors. The critical temperatures of some elemental superconductors are given in the following table. It should be noted that although copper, silver and gold are excellent conductors, yet they do not show superconductivity.

Superconductor	T_C (K)	$B_C(0)$ (Tesla)
Al	1.196	0.0105
Hg	4.153	0.0411
Pb	7.193	0.0803
Sn	3.722	0.0305
Ta	4.470	0.0829
V	5.300	0.1023
W	0.015	0.000115
Zn	0.850	0.0054

Critical temperatures and critical magnetic fields (measured at T = 0 K) of some elemental superconductors.

When the critical temperature of a superconductor is measured in the presence of an applied magnetic field (B), the value of T_C decreases with increasing magnetic field. Intense magnetic fields destroy superconductivity. For instance, at a temperature near the absolute zero, a magnetic field of 0.041 tesla will destroy the superconductivity of mercury. At temperatures near the critical temperature (4.15 K) an even smaller magnetic field is enough to destroy superconductivity. The minimum magnetic field that will quench the superconductivity of a material is called **the critical magnetic field, B_C** . Its strength depends on the temperature. A plot of the critical field strength for mercury as a function of temperature is shown in Fig. 6.3. When the magnetic field exceeds a certain critical value, B_C , the material behaves like a normal conductor with finite resistance. It is found that the critical magnetic field varies with temperature according to the following approximate expression:

$$B_C(T) = B_C(0) [1 - (T/T_C)^2]$$

It could be seen that the value of B_C is maximum at T = 0 K.

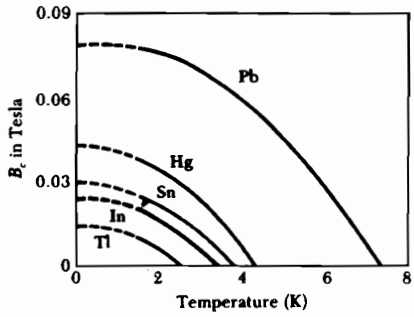


Fig. 6.3. Critical magnetic field as a function of temperature for mercury.

The breakdown of superconductivity imposes serious restrictions on the maximum current that can be carried by a superconductor. The current is associated with a magnetic field, and if this magnetic field is intense enough, it will cause a breakdown of the superconductivity. Such restriction must be taken into consideration in the design of superconducting magnets.

Values of the critical fields for type I superconductors are quite low. For this reason, type I superconductors cannot be used to construct high field magnets, called **superconducting magnets.** There is another class of superconductors, called **type II superconductors**, which is ideally suited for this application.

A.1.3. Persistent currents

Since the resistance of a superconductor is zero, then once a current is set up in the material, it will persist without any applied voltage. If a superconducting loop is placed in the external magnetic field, flux passes through the hole in the loop, (see Fig. 6.4), even though it does not penetrate the interior of the superconductor. After the external field is removed, the flux through the hole in the loop will remain trapped associated with the persistent current that has been generated in the loop. The persistent current will remain flowing indefinitely as long as the material remains superconducting with zero resistance.

Persistent currents induced by changing magnetic fields bring about some spectacular **levitation effects**. If a small bar magnet is

dropped towards a superconducting material, the magnetic field induces persistent currents along the surface of the superconducting material. By Lenz's law, the direction of these currents is such that their magnetic force on the magnet is repulsive. When the magnet is close enough to the superconductor, this magnetic force is so large to be able to support the weight of the magnet, and the magnet looks like floating above the superconductor.

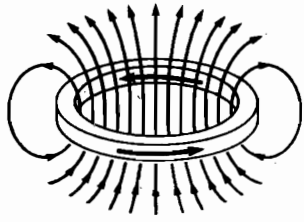


Fig. 6.4. The persistent currents generated in the loop of the superconductor will trap the flux through the hole in the loop.

Levitation effects are used in several applications. One of these is a form of levitation based on a frictionless bearing, Fig. 6.5. Such a bearing may be of any capacity, from large heavy shafts to the smallest gyroscopic systems. In all these bearings the energy dissipation is minimal.

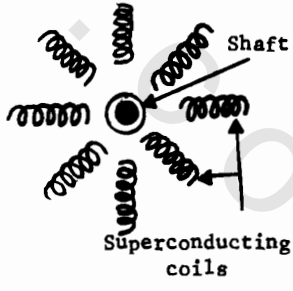


Fig. 6.5. Schematic diagram representation of levitation superconducting bearing.

The bullet train, which has velocities exceeding 300 miles/hour, is based on magnetic suspension and levitation effects. Superconducting coils with a current (a superconducting magnet) surrounds the shaft carrying the train. The persistent currents

induced will levitate the wheels of the train few inches above the rails. The suspension is frictionless. The train sits on a cushion of magnetic fields and slides without friction. A prototype train based on magnetic levitation has already been constructed in Japan using superconducting magnets on the vehicle with liquid helium as the coolant. The moving train levitates above a normal conducting metal track through eddy current repulsion. It is expected that the future vehicles of all sorts will glide above a freeway making use of superconducting magnets.

A.1.4. The Meissner effect

According to Ohm's law, the electric field in a conductor is proportional to the resistance of the conductor. The superconductor has the important property of having zero resistance, accordingly, the electric field in its interior must be zero. According to Faraday's law of induction, the line integral of the electric field, E , around any closed loop is equal to the negative rate of change in the magnetic flux ϕ_m through the loop. Since E is zero everywhere inside the superconductor, the integral over any closed path in the superconductor is zero. Hence, the rate of change of magnetic flux everywhere is zero. Hence, the rate of change of magnetic flux everywhere is zero, which tells us that the magnetic flux in the superconductor cannot change. From this we conclude that if we transport a superconducting cylinder, e.g. into a magnetic field it will push the magnetic lines aside so that none of these penetrate the cylinder. What happens here is that as the superconductor touches the magnetic field, currents are induced on the surface and the magnetic field of these currents produces just the right deformation of the magnetic field lines to prevent their penetration into the cylinder. This behavior is characteristic of a perfect conductor.

The superconductor does not only prevents the magnetic flux from penetrating through, but also it expels any magnetic field lines that are initially inside the material before it becomes superconducting. Fig. 6.6 shows the behavior of a lead cylinder in a magnetic field before and after being a superconductor. When it is an ordinary conductor, magnetic lines of force penetrate it without

hindrance. But, if we cool the lead below its critical temperature, it will expel these field lines and the magnetic field inside it becomes zero.

(a) (b)

Fig. 6.6. Lead cylinder in a magnetic field: (a) when it is in its normal state above T_C , and (b) when it is superconducting.

The expulsion of magnetic flux from a metal during the transition from normal to the superconducting state is called the **Meissner effect**. It means that a superconductor is not only a perfect conductor, but also a perfect **Diamagnet.**

In the next section we give the distinction between different magnetic materials and the properties of para-, ferro- and dia-magnetic materials.

A.2. Magnetic materials

The magnetic properties of materials originate from the spin and orbital motion of electrons in the atom, see Fig. 6.7. The orbital motion may be regarded as a flow of electric current within the atom, and these currents generate magnetic fields. The spinning electrons around their axes could also be considered as flows of electric current that generate magnetic fields. The cooperative action of all the dipoles formed by the moving electrons in spin or orbital motion give rise to the magnetic properties of materials.

An electron moving in an orbit around a nucleus produces a current I given by

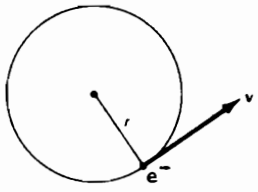


Fig. 6.7. Electron in a circular orbit around the nucleus.

$$I = N e$$

where N is the number of times per second the electron completed one rotation. If v is the electron velocity, then this number is:

$$N = 1/T$$

where T is the orbital period given by:

$$T = 1/N = 2 \pi r / v$$

where r is the radius of the orbit.

Such current will give rise to a magnetic moment:

$$\mu = I \times (\text{area}) = (e v / 2 \pi r) \times \pi r^2 = - e v r / 2$$

In terms of the angular momentum, $L = m_e v r$, the magnetic moment can be expressed as:

$$\mu_{orbit} = - e L / 2 m_e$$

Thus, the magnetic moment of an orbiting electric charge is proportional to its angular momentum. The net magnetic moment of the atom is the sum of the magnetic moments of all its electrons. Besides, the magnetic moment generated by orbital motion of the electrons, we must be take into consideration that generated by the spin motion of the electrons.

An electron may be considered as a small ball of negative charge rotating about an axis at a fixed rate. The spin angular momentum of the electron has a value $(h/4\pi) = 0.53 \times 10^{-34}$ J.s. This kind of

rotational motion again involves circulation of charge and gives the electron a magnetic moment of magnitude:

$$\mu_{spin} = - e S / m_e = 9.27 \times 10^{-24} \ A.m^2$$

S is the spin angular momentum. This magnetic moment is called **the Bohr magneton**. The direction of this magnetic moment is opposite to the direction of the spin angular momentum, S.

The net magnetic moment of the atom is obtained by combination both the orbital and spin moments of all electrons, taking into account the directions of these moments.

In the case of the spin angular momentum, most electrons are paired with opposite spin, so that the pair gives no net spin contribution to the magnetic moment. For atoms and molecules in which pairing is incomplete, the magnetic moment is due to the few unpaired electrons. These atoms have permanent magnetic moments.

A.2.1. Magnetization

Consider a macroscopic object comprising a large collection of molecules. At the macroscopic level we deal with quantities that involve averages over many molecules. The magnetization, M, of the material is defined as the average magnetic dipole moment per unit volume in the medium, and is given as:

$$M = \sum_i m_i / V$$

where m_i represents the magnetic moment of a molecule labeled i in the volume V for which the net average magnetic moment is $\Sigma\ m_i$. The vector sum is over all molecules in this volume. The magnetization is a vector quantity, its SI unit is ampere/meter (A/m).

When a material is introduced in a magnetic field, the field will induce magnetization in the material. The magnetic property of a material is usually measured by its magnetic susceptibility defined as the increase in the magnetization per unit magnetic field strength, i.e.

$$X_m = M / H$$

M being the magnetization and H is the field strength.

A negative magnetic susceptibility indicates that the material is diamagnetic. Whereas, a positive magnetic susceptibility indicates a para- or ferromagnetic material. Paramagnetic materials have low magnetic susceptibility while ferromagnetic materials have very high magnetic susceptibility.

A.2.2. Diamagnetism

Materials whose individual atoms or molecules have paired electrons in their electronic structure, do not have magnetic moments. In the presence of a magnetic field these atoms do have small induced magnetic moments. The direction of these induced moments is opposite to the direction of the external magnetic field so that the magnetization of the material is also opposite the magnetic field direction. Such materials are called **diamagnetic** and they have a negative magnetic susceptibility. They are repelled from the strong parts of the magnetic field. In isotropic diamagnetic material M and H have opposite directions.

When a magnetic field is applied to a diamagnetic material, the magnetic moments of the paired electrons do not cancel. Consider one pair of electrons orbiting the nucleus. Before a magnetic field is applied the magnetic moments of the two electrons are equal and opposite and cancel. As a uniform magnetic field is applied, see Fig. 6.8, an induced electric field (see Faraday's law) changes the orbital

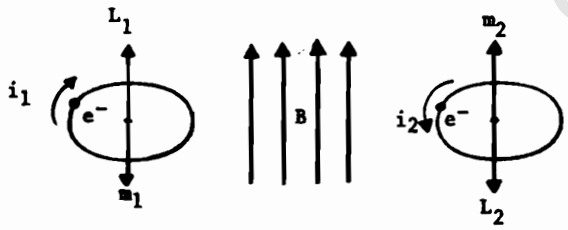


Fig. 6.8. If a magnetic field is applied, the magnetic moments of the paired electrons in the diamagnetic material do not cancel. Since m_1 is greater than m_2 a diamagnetic effect results.

speed of the two orbits. In one orbit the current increases to oppose the changing magnetic flux. Thus, due to the opposite directions of motion of the two electrons, the current in one orbit increases while it decreases in the other. This effect results in an increase in magnetic moment of the first electron while it decreases for the second. The net resultant magnetic moment is the diamagnetic effect.

A.2.3. Paramagnetism

Atoms and molecules with one or more unpaired electrons possesses a permanent magnetic moment. The magnetic moments of the atoms are randomly oriented in the absence of a magnetic field, and the magnetization is thus zero.

When a magnetic field is applied, the molecular magnets will tend to take the field direction and the net magnetization thus increases. The partial alignment of these magnets in the field direction induces magnetization in the same field direction. If the field is removed, the randomness in orientation returns and the magnetization is again zero. Materials having such properties are called **paramagnetic**. They have positive magnetic susceptibilities and are attracted towards the strong parts of the magnetic field. Fig. 6.9 shows the behavior of a dia- and a paramagnetic material in a nonuniform magnetic field.

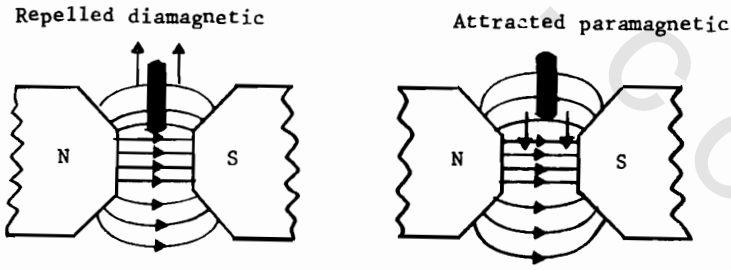


Fig. 6.9. The diamagnetic material is repelled from the strong parts of the field, while the paramagnetic material is attracted to the strong parts.

A.2.4. Determination of magnetic susceptibility

Gouy's method is used for the measurement of magnetic susceptibility, X_m , for a magnetic material. A sensitive balance is used to determine the pull (if paramagnetic) or the push (if diamagnetic) of the material in the form of a thin cylinder whose one end is placed in the strong part of the field, as shown in Fig. 6.10.

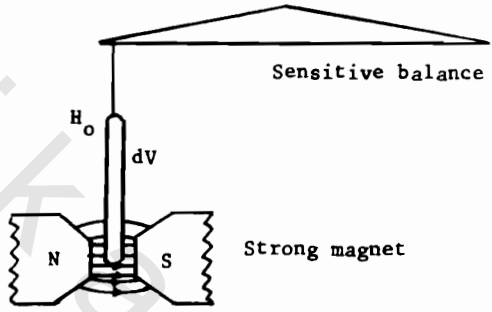


Fig. 6.10. Gouy's method for determination of the magnetic susceptibilities of materials. Liquids could be tested if a cylindrical tube of glass is used as container.

The magnetic field experiences a force (push or pull) on the material that is proportional to the volume, magnetic field intensity, H, and the field gradient, dH/dz. If the material is in the form of a cylinder of uniform cross sectional area, A, then consider an element of volume, dV, the force on this element in the Z-direction is dF, where

$$dF = X_m \, dV \, H \, dH/dz$$

The constant of proportionality is the magnetic susceptibility, X_m .

Since the cross section of the cylinder is uniform, then

$$dV = dx \, . \, dy \, . \, dz = A \, . \, dz$$

Integrating this force over the whole cylinder we get the total force on the material, F. Therefore:

$$F = \int_{H_o}^{H} X_m A \, dz \, H \, dH/dz = \int_{H_o}^{H} X_m A H \, dH = \frac{1}{2} X_m A (H^2 - H_o^2)$$

H is the field intensity at the lower end of the cylinder, and H_o is the field intensity at the other end outside the pole pieces of the strong magnet. H is much higher than H_o , so we can neglect H_o^2 relative to H^2 . The measurement of pull or push by the sensitive balance gives the value of force (F = mg), where (m g) is the increase or decrease in weight of the pan to which the cylinder is attached.

Knowing the area, A, and the field intensity between the pole pieces of the magnet we find the magnetic susceptibility of the material from the equation:

$$m \cdot g = \frac{1}{2} X_m A H^2$$

A.2.5. Ferromagnetism

For both dia- and paramagnetic materials, the magnetization is non-zero only if an applied magnetic field is present. There are some materials who has got magnetization even in the absence of field. These are **ferromagnetic materials**, like iron, nickel and cobalt. These materials contain permanent magnetic dipoles which cooperate by having their magnetic moments aligned together. The tendency towards cooperative alignment of the magnetic dipoles in a ferromagnetic material is diminished under the action of thermal agitation. Thus, the aligning tendency increases with an increase of field intensity and decreases with increasing temperature. These dependencies were first observed by Pierre Curie and are summarized in Curie's law, which relates the magnetization, M, of an isotropic material with the applied magnetic field, H, and the absolute temperature, T, as:

$$M = C H / \mu_o T$$

The constant C, called **Curie constant**, is characteristic of the material and depends on the molecular magnetic moment. The constant μ_o is the **permeability constant.**

A.2.6. Magnetic domains

The magnetization of a ferromagnetic material could be increased by a magnetic field until it reaches saturation. To understand how this happens, we must recognize that all of the magnetic dipoles may not be aligned in a single direction. The ferromagnetic material consists of a large number of regions, with the magnetic dipoles aligned differently in each region. These regions are called **magnetic domains**. In a given domain the magnetic dipoles are aligned in a particular direction, which is the direction of magnetization in this domain. In an adjacent domain the direction of magnetization is different, i.e. magnetic domains have randomly distributed directions of magnetization. The average magnetization for all the sample is zero for the unmagnetized state.

If a magnetic field is applied, then the direction of the magnetization in some domains switches towards closer alignment with the applied field. The size of a domain might increase because of the motion of domain walls. Those domains have their magnetization directions close to the applied field direction. Saturation magnetization is achieved after the material has become one whole domain with its magnetization vector pointing in the same direction as the magnetic field direction, see Fig. 6.11.

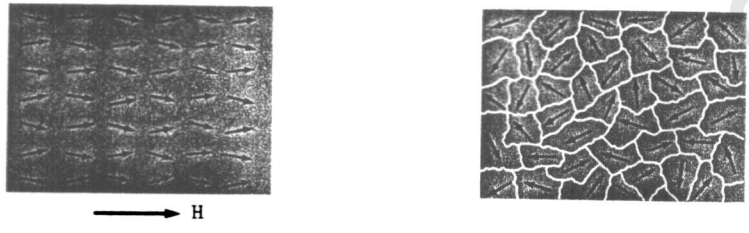


Fig. 6.11. Magnetic domains: unmagnetized and partially magnetized.

Removing the magnetic field might leave the material partially magnetized due to some remnant magnetization, thus forming a permanent magnet.

A.2.7. Hysteresis

The dependence of the state of a system on its past history is called **hysteresis**. The hysteresis of a ferromagnetic material is due to the arrangement of magnetization vectors in the domains under the action of the applied magnetic field. When a piece of ferromagnetic material is introduced in a magnetic field its magnetization increases with the increase of the field intensity until saturation is attained, see Fig. 6.12. If the field intensity is reduced gradually until we reach a zero field, the magnetization does not go to zero but some remenant magnetization remains. Reversing the field direction and increasing its intensity we reach saturation with opposite magnetization vectors. If a complete cycle of operations is made a hysteresis cycle as that shown in Fig. 6.12 is obtained. The coercive field, H_C , is defined as the magnetic field that should be applied in the opposite direction in order to remove completely the remenant magnetism of the sample.

The area of the hysteresis loop gives a measure of the internal energy loss during one complete cycle of magnetization.

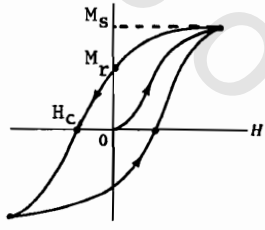


Fig. 6.12. Magnetization, M, of a piece of iron as a function of magnetic field, H, during a cycle of magnetization. M_r is the remnant magnetization. H_C is the coercive field. M_S is the saturation magnetization.

A.3. Type II superconductors

There existed another class of superconductors (known as type II) characterized by two critical magnetic fields $B_C(1)$ and $B_C(2)$. When the applied field is less than the critical field $B_C(1)$ the material is entirely superconducting and there is no flux penetration, just as in the case of type I superconductors. When the applied field exceeds the upper critical field $B_C(2)$ the flux penetrates completely and superconductivity is destroyed. However, for fields lying between the two critical fields, the material is in a mixed state having zero resistance as well as partial flux penetration through filaments called **vortex lines**. These filaments are formed of normal material that run through the sample. This state happens when the magnetic field strength lies between $B_C(1)$ and $B_C(2)$, see Figs. 6.13 and 6.14.

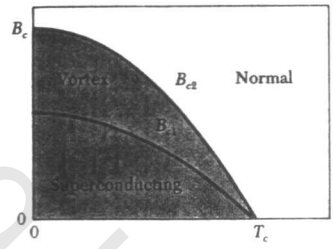


Fig. 6.13. Critical fields as a function of temperature for type II superconductor. Below $B_C(1)$ it behaves as type I superconductor. Above $B_C(2)$ it is a normal conductor. Between these two fields it is in a mixed state.

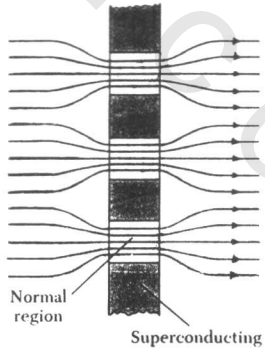
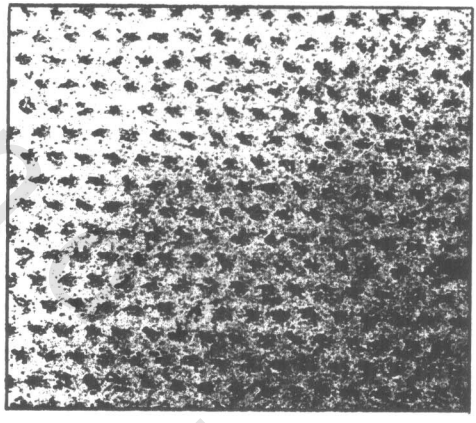


Fig. 6.14. A schematic diagram of type II superconductor in the mixed state. The sample contains vortex lines through which magnetic field lines can pass. The field lines are excluded from the superconducting regions.

A.3.1. Flux quantization

When a type II superconductor is immersed in an intermediate magnetic field to transfer it into a mixed state, the bulk of the material is superconducting, but it is threaded by thin filaments of normal material. The vortex lines are oriented parallel to the external magnetic field and they serve as paths for the magnetic flux lines of the external field. A current circulates around the perimeter of each vortex line. This current circulates around the perimeter of each vortex line. This current shields the bulk of the superconductor from the magnetic field in the filament. The flow of this current has the character of a vortex and that is why the filaments were called **vortex lines.**



End of vortex lines at the surface of a sample of superconducting lead-indium. The vortex lines have been made visible by dusting with powdered iron. The separation between the vortex lines is about 0.005 cm.

It was found that increasing the magnetic field will not cause an increase of the flux associated with each vortex line; instead it will cause an increase in the number of vortex lines threading the superconductor. The stronger the external field, the more densely will the vortex lines be packed. The ends of vortex lines at the surface of a superconducting (type II) material in the mixed state have been made visible by dusting the surface with powdered iron. The vortex lines are packed in the form of heaps having regular pattern on the surface. Knowing the magnetic field intensity and the number of vortex lines per square cm, it was found that the amount of flux associated with each vortex line has a fixed value related to

Plank's constant, h, and the electric charge of the electron, e. The quantum of flux, ϕ_o , is given by:

$$\phi_o = h / 2e = 2.07 \times 10^{-15} \ T.m^2$$

In general the flux, ϕ , is given by

$$\phi = n \phi_o$$

where n is an integer. (n = 1,2,3, ...).

A.3.2. The BCS theory

The conduction mechanism in a normal metal is the charge transfer by electrons through the lattice. The resistivity is due to collisions between the free electrons and the thermally displaced ions of the metal lattice. A superconducting state could never be explained with this classical model because it was not understood why in this state electrons are not scattered by impurities and lattice vibrations.

A microscopic theory of superconductivity was presented by Bardeen, Cooper and Schrieffer in 1957. It could explain the various features of superconductors. We shall describe some of the features of this theory which is now known as the **BCS theory.**

The central feature of this theory is that two electrons in the superconductor are able to form a bound state called: **Cooper pair** which they experience an attractive interaction. Classically speaking, electrons normally repel one another because of their like charges. However, a net attraction could be obtained if the electrons interact with each other via the motion of the crystal lattice as the lattice structure is momentarily deformed by a passing electron. A schematic representation of the basis for the attractive interaction between two electrons via the lattice deformation is shown in Fig. 6.15.

The net effect is that the two electrons (Cooper pair) experience an attractive force between them which arises from an electron-lattice-electron interaction, where the ions of the crystal serves as the mediator of the attractive force. This force is very weak and that is why it appears only at very low temperatures.

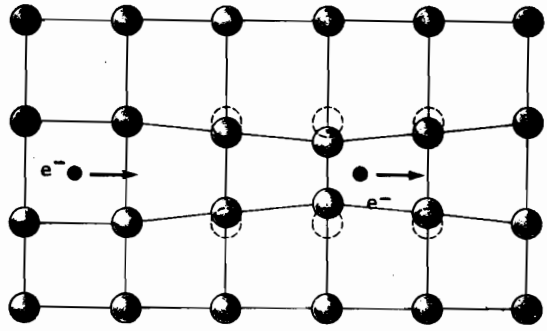


Fig. 6.15. A schematic representation of the basis for the attractive interaction between two electrons via the lattice deformation. The first electron attracts the positive ions which move inwards from their equilibrium positions (open circles). This distorted region of the lattice has a positive charge, and hence a nearby second electron is attracted to it.

The energy gap of the superconductor is defined as the energy required to breakdown the attractive force between the Cooper pair, and this happens at the critical temperature when the superconductor transforms to a normal conductor. The energy gap thus equals kT_C and it is of the order of 0.001 eV.

According to Pauli exclusion principle, the two electrons of the Cooper pair should not have the same quantum numbers, i.e. they should have equal and opposite momenta and spin. The Cooper pair forms a system with zero total momentum and zero spin.

The conduction mechanism in a superconductor is effected by the charge transfer by Cooper pairs. These pairs move through the lattice without resistance because whenever the lattice scatters one of the electrons and changes its momentum, it will also scatter the other electron of the pair and change its momentum by an opposite amount. Consequently, the lattice is not able to scatter Cooper pairs. In the absence of scattering the resistance is zero and the current persists forever.

The BCS theory predicted that the energy gap of a superconductor, namely, the energy needed to breakup one of the

Cooper pairs, at the absolute zero is related to the critical temperature T_C by the relation

$$E_g = 3.53\, k\, T_C$$

A.3.3. The Josephson effect

Tunneling is a phenomenon in quantum mechanics that enables a particle to penetrate through a barrier even though classically it has insufficient energy to go over the barrier. If two metals are separated by an insulator, the insulator normally acts as a barrier to the motion of electrons between the two metals. However, if the insulator is made sufficiently thin, there is a small probability that electrons will tunnel from one metal to the other across the barrier.

If a potential difference is applied between two normal metals separated by an insulator, the current voltage relation is linear and Ohm's law is obeyed. However, if one of the metals is replaced by a superconductor something very unusual occurs. As the potential difference, V, is increased no current is observed until V reaches a threshold value satisfying the relation

$$V_t = E_g / 2e$$

E_g is the energy gap of the superconductor, i.e. the binding energy of the Cooper pair. Above the critical temperature T_C Ohmic behavior is retained between I and V. The non-linear current-voltage relation for electron tunneling through a thin insulator between a superconductor and a normal metal is shown in Fig. 6.16.

Consider next the tunneling in the case of two superconductors separated by a thin insulator. Josephson predicted that the tunneling of Cooper pairs could occur without any resistance and would produce a dc current with zero applied voltage. He also found a second effect, in which an ac current would develop of a dc voltage is applied across the junction. This was called after him, **the Josephson effect,** and the junction of this sort is called **the Josephson junction.**

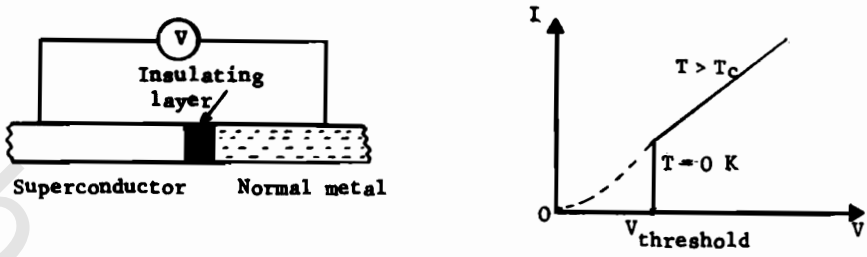


Fig. 6.16. Current-voltage relation of electron tunneling through thin insulator between superconductor and a normal metal at 0 K.

Josephson found that if a dc potential difference ΔV is applied across the junction, the result is an ac current of frequency f, where:

$$f = 2e\Delta V / h$$

The precise measurement of frequency of the oscillating current produced by the steady potential difference applied to Josephson junction has been used for a new determination of the ratio of the fundamental constants e and h, with very high precision. Such measurements also provide the most precise method for the determination of potential differences.

A.3.4. The SQUID

We now consider the effect of magnetic field on Josephson tunneling current. It was found that the maximum critical current in the junction depends on the magnetic flux through the junction. The tunneling current under these conditions is predicted to be periodic in the number of flux quanta through the junction. For typical junctions the field periodicity is about 10^{-4} T.

If a superconducting circuit is constructed with two Josephson junctions in parallel with each other, as shown in Fig. 6.17, one can observe that the total current depends periodically on the flux inside the ring. Since the ring can have an area much greater than a single junction, the magnetic field sensitivity is greatly increased. The device that contains two Josephson junctions in a loop is called a

SQUID, which is the abbreviation of **"Superconducting Quantum Interference Device"**.

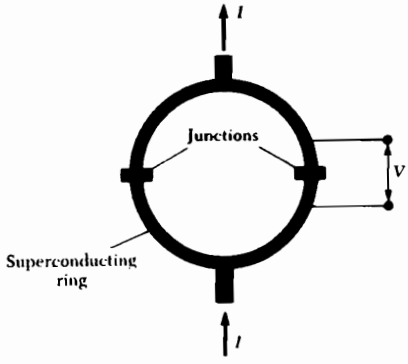


Fig. 6.17. Two Josephson junctions connected in parallel form a SQUID. The net current passing through this device depends on the magnetic flux intercepted by the area spanned by the loop.

The current in the SQUID is zero whenever the flux is half integer multiple of the quantum flux, Φ_0, and the current is maximum whenever the flux is an integer multiple of Φ_0. This sensitive dependence of current on magnetic field makes the SQUID a very useful device for detecting very weak magnetic fields, of the order of 10^{-14} T.

Commercially available SQUIDS are able to detect a change in flux of about 10^{-5} T, i.e. of the order 10^{-20} T.m^2 in a band width of 1 Hz. They are being used to scan "brain waves" corresponding to fields generated by current-carrying neurons, also the currents flowing in the heart. Systems with many SQUIDs are being used to map these biomagnetic fields. It is now hoped that this new tool might be important in locating the source of brain disorders and epilepsy.

A.3.5. High-temperature superconductivity

In the year 1987, it was announced that superconductivity near 92 K was discovered in a mixed phase sample containing yttrium, barium, copper, and oxygen, namely, $YBa_2Cu_3O_{7-\delta}$. A plot of the resistivity versus temperature for this compound is shown in Fig. 6.18. This discovery was very important in the field of high

temperature superconductivity because the transition temperature of this compound is above the boiling point of liquid nitrogen (77 K), a coolant that is readily available, cheap, and simple to handle compared to liquid helium.

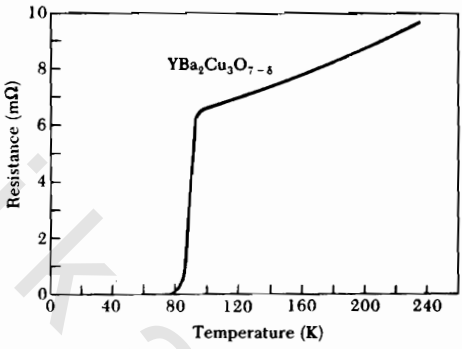


Fig. 6.18. Temperature dependence of the resistance of $YBa_2Cu_2I_{7-\delta}$ showing a critical temperature near 90 K.

Recently, several complex metallic oxides in the form of ceramics have been found with a critical temperature at about 120 K, e.g. Bi-Sr-Ca-Cu-O at 120 K and Tl-Ba-Ca-Cu-O at 125 K.

From a structure study, it was found that there exists a direct relation between the number of copper-oxygen layers in these compounds and the critical temperature T_C . The more these layers are added to the structure before they repeat the higher the critical temperature of the superconductor.

It is now well established that these new copper oxides exhibit the two characteristic properties of superconductors, namely, zero resistance and diamagnetism. The supercurrents were found to be maximally high in the copper-oxygen planes.

The following properties were also found:

1. These high temperature superconductors were of type II with very high upper critical fields.
2. They are anisotropic in nature. Their resistivity is very small in the copper-oxygen planes, and much higher in the perpendicular direction.

3. They have granular or ceramic composition that makes them very brittle and inflexible.
4. There is a direct relation between the superconducting properties and their crystallographic structures that can be classified in terms of the so-called pervoskite crystal structures.
5. Substitution of atoms on the copper oxide layers degrades or destroys the superconductivity.

A.4. Applications of superconductivity

The discovery of high-temperature superconductivity may introduce many important technological advances in future, such as having superconducting devices in every household. There are some difficulties that have to be overcome before such applications become reality. These difficulties are:

1. The high T_C superconductor is a ceramic material. It is brittle and cannot be shaped at our own will in useful shapes.
2. Thin films for small devices such as SQUIDs are needed.
3. Low current densities are measured in bulk ceramic compounds.

It is hoped that these difficulties are overcome in the near future. We give in the following part some practical applications.

The property of zero resistance to dc currents is used for low-loss electrical power transmission. It is known that a significant fraction of electrical power is lost as heat to overcome the resistance of the normal conductors in ordinary devices. If the power transmission lines could be made superconducting, these dc losses could be eliminated and there would be substantial savings in energy costs. Superconducting cables are constructed in the form of very thin filaments of superconducting NbTi alloy. These filaments are embedded in a copper matrix. A bundle of thin filaments has more surface area than a single thick wire of the same cross section as the bundle; since the supercurrents always flow on the surface of the superconductor, the large area of the filaments permits the flow of a large current. As long as these filaments remain superconducting filaments. But if the cooling is somewhat inadequate or in excessive

magnetic fields, the superconductor might become a normal metal and the very high current in the filaments can continue to flow in the parallel wires of copper. This protects the device from the explosive conversion of magnetic energy into heat that would occur when a large current is stopped by a large resistance.

Superconducting generators have the advantage of small size and small weight for a given power output. They find application in nuclear power plants and in marine propulsion where weight is a problem. Electric generators with superconducting coils have an efficiency close to 100%. They have very large power output.

Another application is the magnetic suspension of trains. The phenomenon of magnetic levitation has already been used in Japan for construction of a train with superconducting magnets on the vehicle. The moving train levitates above a normal conducting metal track through eddy current repulsion.

An important application of superconducting magnets is the magnetic resonance imaging (MRI). This technique relies on intense magnetic fields generated by superconducting magnets. It uses the safe rf radiation to produce images of body sections, rather than X-rays.

A.5. Fullerine, a new type of superconductor

The discovery of fullerine, C_{60} , was considered the most striking news of the year 1985 in chemical physics. Chemical physicists at Rice University discovered, after they laser-vaporized a graphite target in jet of helium gas, that 60-atom clusters of carbon remained uniquely stable. They rationalized that such a stable molecule might have assumed the perfectly symmetrical form of a foot-ball. In tribute to the famous architect Buckminster Fuller, whose geodesic dome was taken to resemble the molecule, the Rice group named the spatial C_{60} cluster "Bucky ball". Other research groups observed that many other even-numbered carbon clusters existed in a stable form. Researchers proposed that if these clusters were formed of closed spheres, as they suspected, then the laws of geometry would demand that their structures would possess close similarities, i.e. exactly 12 of their faces would consist of pentagons of carbon atoms, and

depending of the size of the cluster, some several hexagonal rings would connect the 12 pentagons. Researchers proposed that these clusters belonged to a new class of pure carbon called **Fullerine**, typically represented by the Bucky ball molecule, C_{60} , shown in Fig. 6.19.

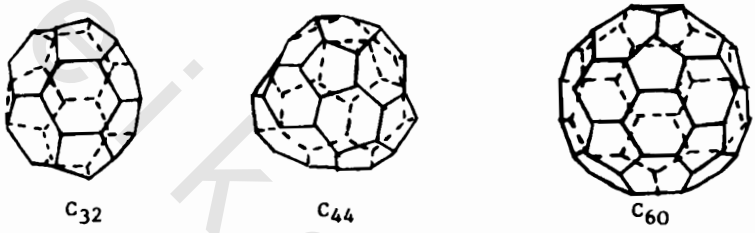


Fig. 6.19. Fullerine molecules.

Short time after its discovery, it was realized that there are other similar carbon clusters, but only with even number of atoms, which are rather stable. The fullerine C_{60} , being the most stable. All these clusters ranged from as small as 32 atoms to as large as 960 atoms forming a network of hexagons and pentagons.

For C_{60} , the diameter of the cage on which carbon atoms are arranged is about 7 Angstroms, while the molecule as a whole is roughly 10 angstroms in diameter. The center of the cage is devoid of charge since the charge density falls off rapidly away from the cage. The atoms are distributed symmetrically such that the strain is minimum.

A.5.1. Structure and bonding of fullerine

It has been now well agreed that fullerines are a distinct class, differing in characteristics from diamond or graphite, the well known crystalline forms of solid carbon. The fullerine represents a third unique bonding structure for carbon, joining planar graphite and tetrahedral diamond. Fleming and his co-workers in the Bell

Laboratories showed that the structure is face centered cubic at 300 K, with the Bucky molecules sitting in the lattice sites. The bonding between the C_{60} molecules in the f.c.c. structure is Van der Walls in character, but covalent bonds exist between the carbon atoms in any one molecule. This means that out of 4 electrons of each carbon atom, three electrons are shared with three neighboring carbon atoms. The remaining electron forms an orbital perpendicular to the surface of the cage. These orbitals of each carbon atom can overlap with each other forming molecular orbitals. The electrons from these orbitals are delocalized over the cage.

It is interesting to make a comparison of the structures of the three allotropic forms of solid carbon. Fig. 6.20 shows such a comparison of atomic arrangement and interatomic distances and density of each form.

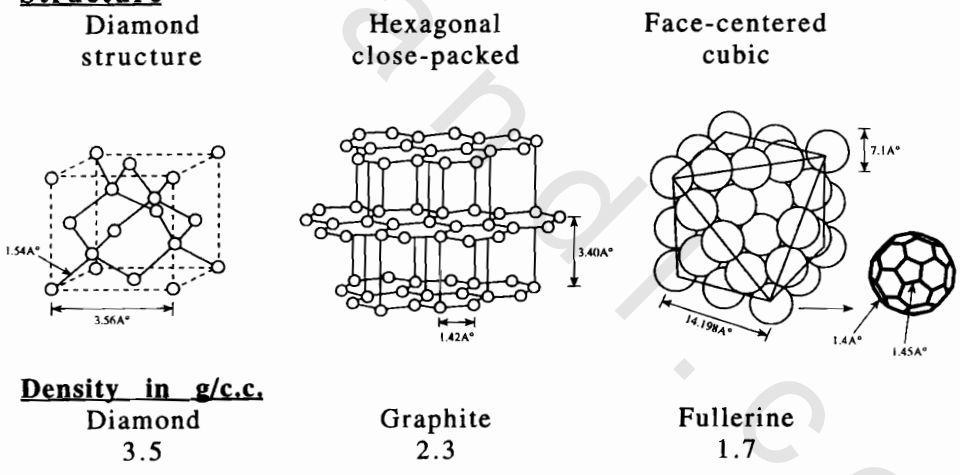


Fig. 6.20. Allotropic forms of carbon.

A.5.2. Production of fullerine

After the spectacular discovery and announcement of fullerine, many scientists directed their efforts to understand more this new form of carbon. However, the exotic experimental setup in which graphite was vaporized in the atmosphere of supersonic helium beam

by high power laser was not accessible to most of the researchers. No progress took place from 1985 until a break through in 1990 was made by Kratschmer in Germany who announced inexpensive method to produce reasonable quantities of C_{60} . Production of a few milligrams of fullerine per day has not remained a difficult task any more.

Fullerine is simply produced in a deposition chamber which can be evacuated to a pressure of about 10^{-6} to 10^{-7} torr. The chamber is cooled with water. Helium gas is introduced in it at a pressure of about 40 to 300 torr. An arc is struck between two graphite electrodes using a power supply which can deliver about 50 volts and 100 amperes, see Fig. 6.21.

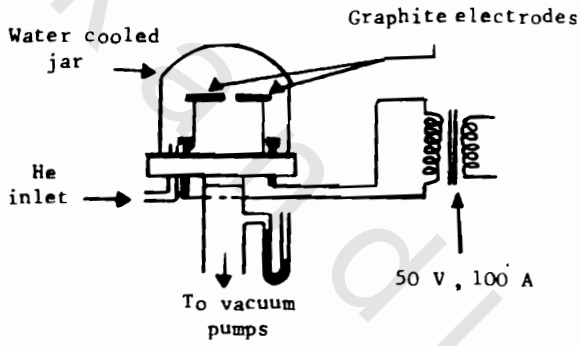


Fig. 6.21. The experimental setup to deposit fullerine.

When the arc strikes, graphite from the electrodes start to evaporate and then be condensed either on the water cooled chamber or on water cooled substrates. After deposition for a few hours, a sufficiently large quantity of black soot can be collected. This black soot is dissolved in benzine or toluene. The solution is then filtered to separate carbon contents which are not fullerines. After heating this solution to evaporate leaving behind a residual powder, we get C_{60} as well as other fullerines. Chromatographic methods need to be employed in order to separate out fullerines of different sizes.

It is now well accepted that fullerines could be formed at high temperatures, even a candle soot produces C_{60} , although in very

small amounts. Graphite heated to about 1500 K in helium atmosphere can produce fullerine. It is speculated that the vaporized carbon atoms first form chains that link together and then become graphite sheets which curl together in the inert atmosphere of helium to form cagelike molecules without dangling bonds. Pentagons are crucial in curling up.

A.5.3. Properties of fullerines

Fullerine has some remarkable properties summarized in the following:

- The linear compressibility of the lattice is very small and is comparable to that of the planes of graphite.
- The volume compressibility is very large, it is many times greater than that of graphite or diamond. This was attributed to the hollow structure of the cage in which foreign atoms can fit in easily. It is a very good host for other atoms.

One of the most remarkable aspects of fullerines is related to their ability to accept charge from donor atoms, such as alkali metals, thus forming alkali metal fullerides. It was found that films of fullerine exposed to lithium, sodium, potassium, ..., etc., vapor sources, changed the films from being insulators to conductors. The alkali metal atoms donated their weakly bound s-electrons to the fullerine so that the first group of empty states would be partially occupied, giving rise to electronic conduction.

A.5.4. Superconducting fullerides

K_xC_{60} was found to be a superconductor with a transition temperature T_C = 18 K. Rubidium fulleride was also found to be superconducting with a critical temperature of 28 K. Investigators in Japan proved that Cs_2RbC_{60} , has a critical temperature of 33 K.

The field of study of fullerine is still very widely open. Particular intriguing results may come when fullerine is used in the field of polymerization, or when substitutions for carbon atoms are effected.

B. NON-LINEAR PHENOMENA IN PHYSICS AND CHAOS

B.1. Historical

In 1831, Faraday observed the existence of harmonic frequencies, f, 2f, 3f, ... as well as subharmonic frequencies, f/2 , f/3 , ... , in forced vibration of a broad immersed in the water of a container; f was the fundamental frequency of the system. Some years later, Rayleigh repeated Faraday's experiment and tried to explain the appearance of subharmonics by what he called parametric resonance. This experiment was the first which attracted the attention to the role of non-linearity.

In frequency transformers, if we feed in signals of frequencies f_1 and f_2 and analyses the output, then it will contain the frequency components only if the system is essentially linear. If frequency combinations such as $f_1 + f_2$, $f_1 + 2f_2$, $2f_1 - f_2$ or more generally, $nf_1 + mf_2$, where n and m are integers, then the system is a non-linear one.

To show this phenomenon, consider the linear case of Ohm's law: V = IR, where V is the voltage, R is the resistance, and I is an alternating current. If I has the form:

$$I = A \cos \omega_1 t + B \cos \omega_2 t$$

$\omega = 2 \pi f$, is the angular frequency, then the voltage V will contain the same ω_1 and ω_2 and no other frequencies will appear. This is the linear case.

On the other hand, if Ohm's law contains a non-linear term, say:

$$V = R_0 I + R_1 I^2$$

then, in addition to ω_1 and ω_2 , the voltage will contain harmonics: $2\omega_1$, $2\omega_2$ and $\omega_1 + \omega_2$, $\omega_1 - \omega_2$; these new frequencies appear as a direct result of the trigonometric relations: $\cos \alpha \cos \beta = 1/2 [\cos (\alpha+\beta) + \cos (\alpha-\beta)]$, etc. One cannot get subharmonics in this way. It was previously reported by Faraday and Rayleigh that harmonic

frequencies appear always whenever a non-linear term exists, no matter how small it may be.

But subharmonics have a threshold. The f/2 component appears only after the amplitude of the driving force reaches a well defined value. The appearance of subharmonics and the threshold was explained one century later in a new class of phenomena called **CHAOS**. It was recognized later that subharmonics usually appear at the start of Chaos. Presently, Chaos acquires a highly non-trivial explanation.

B.2. Chaos as a deterministic phenomenon

In 1963, Lorentz, working in meteorology, presented a model which described thermal convection in the atmosphere. This model was completely described by a completely deterministic system of three differential equations. By completely deterministic we mean that no random force, no external noise, no stochastic elements of any kind, have been introduced in the model.

Nevertheless, for some range of parameters, the system possesses aperiodic solutions which show irregular variations that seem to be undeterministic and unpredictable as a random process.

Since then, several observations in the domain of Chaos were reported. Examples were in the onset of turbulence in dissipative dynamical systems, the study of motion of celestial bodies (a few body problem) and the kinetic theory of gases (many body problem).

In the beginning many physicists took Chaos for randomness. However, once convinced in the existence of Chaos, we began to see it everywhere. Chaotic behavior was discovered in laboratories, in computer experiments, and in natural processes.

B.2.1. What Chaos meant in the past and at present

Chaos is a Greek work in origin, it describes "the unorganized state of primordial matter before the creation of orderly forms in the universe", i.e. it is a "disordered collection of state", a confused mixture.

It is a "state of things in which chance is supreme". Thus, Chaos in the past was associated with randomness and disorder.

At present, and for practical purposes, it is appropriate to use a working definition of Chaos. One is dealing with Chaos, if the following ingredients are all present:

1. The underlying dynamics is deterministic, i.e. no external noise or other random elements are introduced.
2. The individual trajectories show a seemingly erratic (have no certain course, irregular) behavior which depends sensitively on small changes of initial conditions.
3. In contrast to a single trajectory, if we average the global characteristics over a long time, it does not depend on initial conditions.
4. When a parameter is varied, an erratic irregular state is reached via a sequence of sudden changes of the dynamical behavior, usually including the appearance of one or more subharmonics.

B.2.2. The non-linear motion of the pendulum

We consider the motion of the simple pendulum because it is one of the simplest mechanical systems.

In Fig. 6.22, L is the length of the weightless rod carrying the bob of mass m and is mounted on a freely rotating hinge, O. The rod may oscillate or rotate without friction around O, and the motion takes place in the plane of the figure. When the mass m hangs vertically at the point A, it is in a position of minimal potential energy and could stay there for ever. We take this minimum potential energy, V, to be zero.

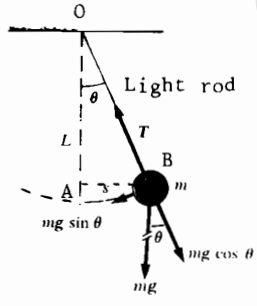


Fig. 6.22. The pendulum.

When the mass is moved away from the equilibrium point A, to a point, say B, and then released, it comes into motion under the force of gravity. The potential energy at B is:

$$V = m g L (1 - \cos \theta)$$

where g is the acceleration due to gravity on the earth and θ is the angle in radians between the two positions of the rod. We see that V = 0 when $\theta = 0$. The mass point, m, can move only along a circle of radius L centered at O, so the actual acceleration of m is $L\, d^2\theta/dt^2$. The displacement is $L\theta$.

According to the second law of Newtonian mechanics, the equation of motion is given by

$$m L (d^2\theta/dt^2) = F = -(\partial V/\partial L\theta)$$

where the force F is calculated as the gradient of the potential energy V. Using the potential energy equation we get the equation of motion for the pendulum:

$$(d^2\theta/dt^2) + \omega^2 \sin \theta = 0$$

where the angular frequency, ω, is given by:

$$\omega = \sqrt{g/L} = 2\pi/T$$

Therefore $$T = 2\pi \sqrt{L/g}$$

The above equation is a non-linear ordinary differential equation, which may be solved rigorously in terms of elliptical integrals and shows no trace of Chaos.

B.2.3. Chaotic motion of the pendulum

The second order of differential equation of the pendulum may be written as a system of two first order equations by introducing the angular velocity, p, of the pendulum, where

$$p = d\theta/dt$$

and so $$dp/dt = -\omega^2 \sin \theta$$

There is a mathematical theorem which says that any ordinary differential equation of second and lower order cannot possess Chaotic motion. But, how can we modify this equation to bring into play Chaotic motion?

Let us look first at the case when the pendulum makes only small-angle oscillations, i.e. when θ is small. We can put $\sin\theta = \theta - \theta^3/6 + \theta^5/120 - \ldots$, and neglecting the higher orders of θ, we arrive at the well-known equation of the simple harmonic motion:

$$(d^2\theta/dt^2) + \omega^2\theta = 0$$

The solution of this equation is:

$$\theta = A \sin(\omega t + \phi)$$

A and ϕ are arbitrary constants determined by the initial conditions. The number of arbitrary constants should equal the number of independent variables in the system, here: p and θ.

If we plot the motion in the phase plane: p as ordinate and θ as abscissa, we get a periodic orbit as shown in Fig. 6.23., as long as θ is small.

Since θ may take values only between $-\pi$ and $+\pi$, then the orbit representing the motion of the pendulum in the phase plane will never get out of the dotted lines in the figure. Actually $\theta = -\pi$ and $\theta = +\pi$ correspond to one and the same angle. So, we can identify two vertical sides making a cylindrical surface of radius π , in the phase plane of the pendulum. The trajectory cannot go outside this cylinder.

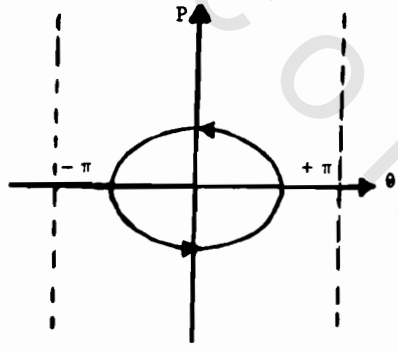


Fig. 6.23. A periodic orbit in the phase plane.

When the mass point performs small-angle oscillations the trajectory makes an ellipse in the phase plane. The point (θ,p) is called the representative point of the system. This point makes the trajectory of the system. In case of small amplitudes, the trajectory is a closed curve that repeats itself each one period T, where:

$$T = 2\pi / \omega$$

This is the simple periodic motion.

B.2.4. The introduction of friction

Frictional forces opposing the motion always exist in real systems. To a good approximation we may represent the frictional forces by a term that is proportional to the velocity.

$$\text{The frictional force} = - k (d\theta/dt)$$

where k is the coefficient of friction and the minus sign shows that friction diminishes the acceleration.

The equation of motion of the pendulum could thus be modified as follows:

$$(d^2\theta/dt^2) + k (d\theta/dt) + \omega^2 \theta = 0$$

The solution of this equation contains an attenuation factor

$$\theta = A e^{-kT} \sin (\omega t + \phi)$$

so the amplitude of oscillation decays with time and finally the pendulum stops at $\theta = 0$. Its trajectory in the phase plane spirals towards the origin, as shown in Fig. 6.24. This is the trajectory of a free pendulum that conserves the energy it has been given in the initial state. A conservative system with friction dissipates its energy and finally arrives at a stand still.

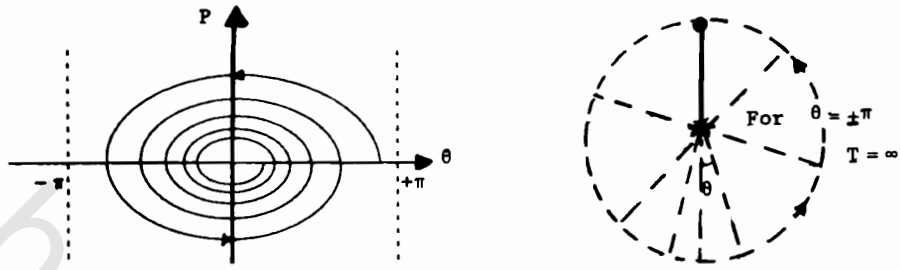


Fig. 6.24. A damped trajectory in the phase plane.

B.2.5. Increasing the amplitude of vibration

When the amplitude of vibration is increased, we can no longer use the linearized equation of motion, i.e. that of the simple harmonic motion. In the phase plane, oscillations with larger and larger amplitudes are represented by larger and larger oval curves. The period T will no longer be a constant. It will grow with amplitude and eventually reaches infinity when the trajectory reaches the point $\pm\pi$. The $T = \infty$ orbit is called a **separatrix,** because it separates the oscillation of the pendulum from its rotation about the hinge. Above the separatrix the angular velocity p always takes on positive values, corresponding to anticlockwise rotation. Below the lower part of the separatrix, p takes on negative values, corresponding to clockwise rotation, see Fig. 6.25.

If θ is not small, and the amplitude is increasingly large, then we cannot put $\sin\theta = \theta$, in the equation of motion. Thus,

$$(d^2\theta/dt^2) + k\,(d\theta/dt) + \omega^2 \sin\theta = 0$$

Now, it is no longer a conservative system. The pendulum may start with say, anticlockwise rotation and then, after making several full turns, change to oscillations with smaller amplitude and eventually stops. However, in all that has been mentioned, there is no Chaotic motion.

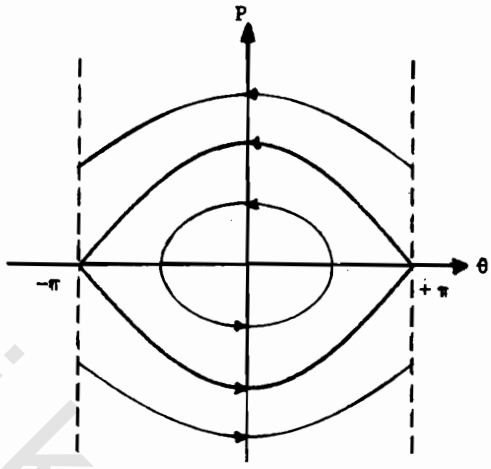


Fig. 6.25. Phase plane of the non-linear pendulum. The heavy lines show the separatrix.

B.2.6. Chaotic motion of the pendulum

In order to allow Chaotic motion, we apply a periodic external force to the pendulum. This could be done in practice by an electromagnet fed by alternating current with some frequency Ω .

$$(d^2\theta/dt^2) + k\,(d\theta/dt) + \omega^2 \sin\theta = A \sin(\Omega t)$$

The right hand side of this equation may be viewed as the output of some linear oscillator. Accordingly, we may say that the above equation describes a system formed of two coupled oscillators: the non-linear damping oscillator (L.H.S.) and the linear oscillator (R.H.S.).

Two extremes could readily be understood:

1. When the driving force is very strong, the pendulum will closely follow the external oscillation with frequency Ω .
2. When the driving force is very weak, the non-linear damping oscillator will dominate.

In the intermediate state various regimes of oscillation may occur and Chaos may appear in some range of parameters A and Ω . Fig. 6.26 shows an example of Chaotic orbit in the phase plane.

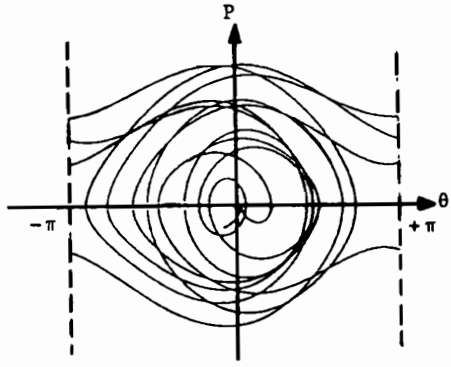


Fig. 6.26. A Chaotic trajectory in the phase plane.

B.2.7. Example of Chaotic motion

A girl or boy playing a swing provides a good example for Chaotic motion. The person on the swing cannot get into motion by himself from a stand still position no matter how skillful he is. Another person should give him a small push as an initial driving force. Then the person on the swing lowers his body when the swing goes high and stands up when passing through the lowest point. Gradually, the system is driven in full swing, and it might perform rotation after passing through the separatrix ($\theta = \pm \pi$), see Fig. 6.27.

Fig. 6.27

We see that during one period of the swing, the person changes the effective length L of the swing twice and finally reaches a kind of

resonance. Counting the times the person stands up, it is twice in one period of swing. Thus, the swing is moving with half of the frequency, i.e. the resonance comes at a subharmonic of the driving frequency. This is the parametric resonance that has been mentioned by Lord Rayleigh one century ago when he tried to explain the observed subharmonics in systems that vibrated forcedly. It is called **parametric resonance** because instead of applying a periodic force directly to the system, a periodic change is introduced in one of the parameters of the system.

If we describe the swing system by changing the frequency term, in the equation of motion, from ω^2 to $\omega^2 (1 + A \sin \Omega t)$, then:

$$(d^2\theta/dt^2) + k (d\theta/dt) + \omega^2 (1 + A \sin \Omega t) \sin \theta = 0$$

Systems with periodic parametric driving are capable of exhibiting Chaotic behavior.

Because such systems have wide technical applications, engineers usually encounter parametric resonance and Chaos in their work. A well-known example is the vibration of a bridge under the action of the regular pace of moving troop of an army.

B.3. Chaos in the population number of a species

A study of the population change in some states of America and Europe showed that the population when left free, goes on doubling itself every twenty-five years, or increased in a geometrical ratio. To put this observation in a mathematical form, let us divide history into consecutive periods of 25 years each and denote the population in the beginning of the n^{th} period by y_n , thus:

$$y_{n+1} = 2 y_n$$

This factor of 2 is because we assume that the population is doubled every period. It is better to put a coefficient 'a' instead of 2 and write

$$y_{n+1} = a y_n$$

This is a linear difference equation, whose solution is given by

$$y_n = a^n y_o$$

where y_o is the known population of the initial period.

It could be seen from this equation that if we take y_o the initial population to be one million, then in less that fifteen 25-year period, y_n , the present population would exceed the present population of the Earth.

It could be seen that if 'a' is greater than 1 ($a > 1$), then sooner or later there will be a population explosion, i.e. y_n inevitably tends to infinity, provided the initial population y_o was not zero. But, when 'a' is less than 1 ($a < 1$), the population eventually vanishes.

Although the above law is too simple to count for human population, yet it could be modified to describe the population number of certain kinds of species, such as some seasonally breeding insects.

Suppose we observe the y_n population in the n^{th} summer. We consider only one isolated species and there is no interaction with other species. Assume further that there is no overlap of different generations, i.e. one generation dies entirely after each insect lays a eggs and the next spring all eggs hatch. The law

$$y_n = a^n y_o$$

holds true approximately when y_n is small and there is plenty of food and space.

New phenomenon comes into play when y_n gets large enough. Insects will fight and kill each other for limited food; some contagious epidemic disease may sweep through the population, etc. Fighting or touching requires the contact of at least two insects. The total number of such events is proportional to y_n^2. Taking into account this suppressing factor we can modify the above equation to read:

$$y_{n+1} = a\, y_n - b\, y_n^2$$

One of the two parameters a and b in the above equation could be eliminated by taking $a\, y_n$ as a new variable and b/a as a new parameter. We can further normalize the population, taking the maximal population to be 1.

We are then led to the following abstract population model

$$x_{n+1} = 4 \lambda\, x_n\, (1 - x_n)$$

The population now varies between 0 and 1.

It is more convenient to write the logistic map (above equation) in a different forms:

$$x_{n+1} = (1 - \mu\, x_n^2)$$

Thus x_{n+1} is a function of μ and x_n , i.e. $f(\mu, x_n)$. It is non-linear function depending on the parameter μ.

Usually we start from an initial instant $n = 0$, and put the initial values x_o in the right hand side of the equation with a given μ , then calculate the output x_1 . Repeat to get x_2 , x_3 , ...

Depending on the parameter μ the x's may exhibit different behaviors. As example: take $\mu = 0.1$ and $x_o = 0$, and calculate: $x_o = 0$, $x_1 = 1$, $x_2 = 0.9$, $x_3 = 0.919$, $x_4 = 0.9155439$, ... , $x_{14} = 0.9160797831$, $x_{15} = 0.9160797831$.

From x_{14} onwards we see that the value of x_n will not change, i.e. we have reached a fixed point.

In terms of insect population, this means that after 13 periods (here 13 summers) the number of insects settles down at a fixed value; the species will survive forever at this population level.

B.3.1. What happens if μ is large, say 0.9 ?

Proceeding with the calculation leads eventually to the alteration of two numbers: $x_{45} = 0.9858870385$; $x_{46} = 0.1252240726$; $x_{47} = 0.9858870385$; $x_{48} = 0.1252240726$; ...

There is a point where we get a period-doubling phenomenon. Period doubling is a well known observation from experience. In the language of insect population we may say that if during this summer there are a lot of insects, then in the next summer there will be fewer - this fact is known by many farmers from experience. Similar behavior has been known for the harvest of some kind of fruits.

Another example for period doubling: in the suburb of a big city, traffic jam happened on one weekend on a certain road. Due to the spreading of this information among people, then there will be no traffic jam the next weekend on the same road. Thus, although urban life has a one-week period, traffic jams may happen at a two-week period.

A necessary condition for period doubling to happen requires that there will be two opposing factors:

1. an encouraging factor, such as the breeding of insects, the good news about a free road, and
2. a discouraging factor, such as the fighting of insects to get food, the bad news on traffic jam.

The parameter μ reflects the interplay of these two factors. Moreover, the presence of these two factors can lead to a variety of more complicated behavior besides period doubling. This could be seen clearly by drawing a bifurcation diagram on a computer screen.

B.3.2. What is bifurcation ?

Bifurcation is a mathematical term for a sudden change of the number of solutions in an equation. As we have seen, at a small value of $\mu = 0.1$, there is one fixed point, while for large $\mu = 0.9$, there is a two-cylce, i.e. an alteration of two points in the iterating process. So we say that a bifurcation happened at a some

intermediate value of μ . The value of μ where the bifurcation takes place could be calculated easily, Fig. 6.28.

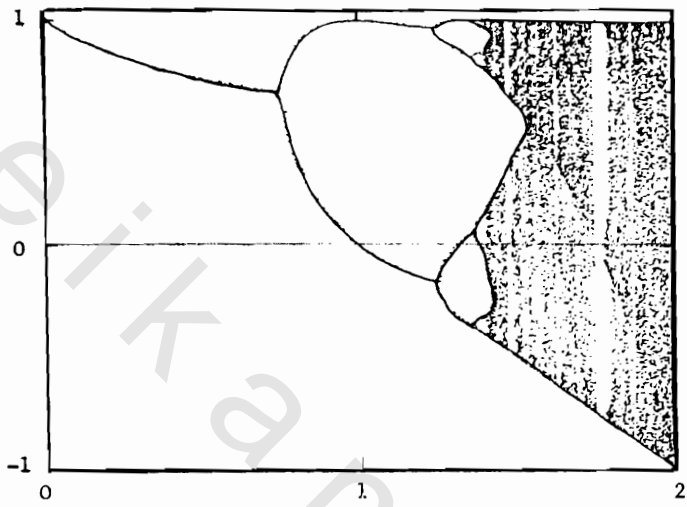


Fig. 6.28. A bifurcation diagram for the quadratic map.

B.3.3. Chaos and turbulence

The study of turbulence was made by Reynolds in 1880. He studied the flow of liquids in pipes and found that the transition from stream line motion to turbulent motion is governed by the following factors:

1. A characteristic length, e.g. the size of the container L .
2. Characteristic velocity, e.g. the average speed v.
3. The coefficient of viscosity η .

He introduced a characteristic number, which now carries his name, the Reynolds' number:

$$Re = (L \times v) / \eta$$

Reynolds observed that increasing the velocity of flow in a pipe until Reynolds number reached 2300 the flow became turbulent.

The understanding of the onset mechanism of turbulence and the description of fluid properties such as heat conductivity or viscosity, in a turbulent state are problems of great practical importance. Atmospheric turbulence encountered by a jet plane may cause serious trouble . Turbulent flow in blood vessels, say, that behind an artificial cardiac valve in the heart, may bring about fatal results.

After Reynolds experiments, the problem of understanding turbulence has remained a challenge for science for more than one hundred years.

The difficulty of the turbulence problem partially originates from the necessity to take into account too many length scales in the fluid, ranging from molecular size to the size of the container or the airplane. If turbulence manifests itself merely as a process of disintegration of motion, i.e. the energy transfer from large eddies to smaller and smaller eddies, and at last, into heat at the molecular level, then a suitable statistical description by taking some kind of average would basically solve the problem. However, in the development of turbulence they may emerge large scale coherent structures such as vortex formation in a street behind an obstacle, see Fig. 6.29. How can we describe the erratic motion at so many disparate (incapable of being compared) length scales, and at the same time to allow for the ordered structures (such as the rows of vortices) at large scales, this problem remains unsolved.

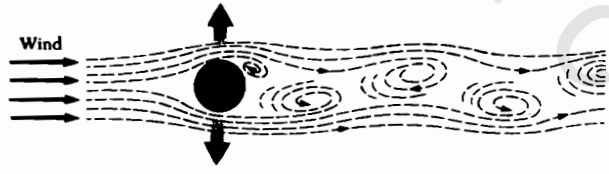


Fig. 6.29. Row of vortices behind an obstacle in a moving fluid.

It should be made clear that studies on chaos succeeded only with systems having few degrees of freedom. To understand turbulence it is necessary to treat the interaction among infinite spatial degrees of freedom.

C. CHAOS AND FRACTAL ANALYSIS

C.1. The fractal, a new dimension in physics

The fractal is a special kind of geometrical object with a self-similar structure. Self-similarity means that a part of the object, if properly stretched or compressed, looks much like the whole object. Such objects are described by a fractal dimension.

Let us first recall the conventional dimension of ordinary objects. Consider a square in a plane and let us increase its linear size to L times its original size in each direction. We get a square which is L^2 times as large. The same increase of linear sizes, when applied to a cube, leads to a cube which is L^3 times as large.

In general, if we have a geometric object and we make the object N times larger by increasing its linear sizes to L times their originals. then:

$$N = L^d$$

where d is the dimension of the object.

We know that a point object has dimension 0, a linear object has dimension 1, a planar object has dimension 2, etc. We are familiar with these integer dimensions.

Taking the logarithm of both sides of the above equation, we get:

$$d = \log N / \log L$$

Taking this equation as a new definition of the dimension, d, releases us from the restriction of d being an integer. So, d can have fractional values as well, hence the name **fractal dimension.**

It should be emphasized that a prerequisite for the notion of fractal to be applicable, there must be some kind of scale invariance, i.e. the phenomenon under study seems unchanged when the scale of observation, e.g. the resolution of the microscope, has been changed by many orders of magnitude. For mathematical models the

invariance may exist over an infinite range of scales, see Fig. 6.30. In real world one can have approximate scale invariance over several orders of magnitude, as shown in Fig. 6.31. Only when there is enough evidence that this scale invariance does exist, one can use fractal geometry.

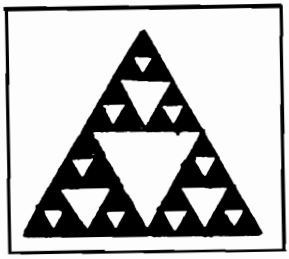


Fig. 6.30. Sierpiniski gasket fractal after four stages of iteration. Notice that holes exist in all scales.

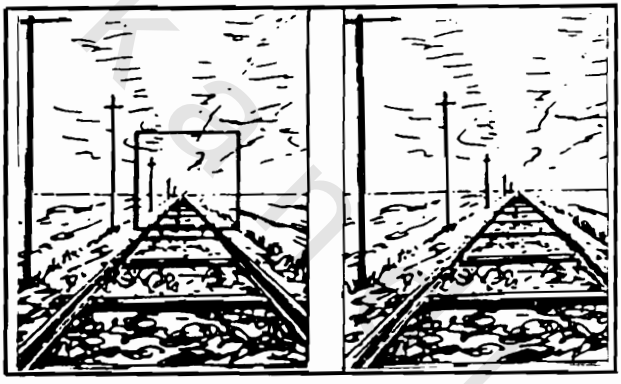


Fig. 6.31. Scale invariance in real world. Photograph scenery from the back of a train. The photograph looked the same at all stages of enlargement.

C.1.1. Calculation of the fractal dimension of Sierpiniski gasket

Sierpiniski gasket is a non-random fractal which is defined operationally as an aggregation model obtained by simple iterating growth rule, like a child assembling a castle from building blocks. The basic unit is a triangular shaped tile which we take of unit mass (M = 1) and of unit edge length (L = 1). If we join three tiles together to form the structure of Fig. 6.30, we get an object of mass M = 3 and edge L = 2 in the first stage of aggregation. In the next stage of aggregation, we have L = 4 and M = 9, and so on.

If we define the density ($\rho = M/L^2$), then for stage 1 the density drops from $\rho = 1$ to $\rho = 3/4$. By further growing the body to stage 2 where $M = 9$ and $L = 4$ we get $\rho = 9/16$, and so on. It could be seen that ρ decreases monotonically with L without limit in a predictable fashion, i.e. following a simple power law:

$$\rho = A \cdot L^{\alpha}$$

where A is a constant that is not of intrinsic interest and α is a constant depending on the rule that we follow when we iterate.

From the double logarithmic relation between ρ and L we get a straight line, the slope of which yields the fractal dimension, d, defines as:

$$M = A L^{d}$$

but we have: $$\rho = M / L^{2}$$

from which the density is given by:

$$\rho = A \cdot L^{d-2}$$

For Sierpiniski gasket, the slope of the double log plot of ρ vs. L is given by:

$$\alpha = \text{slope} = \frac{\log 1 - \log (3/4)}{\log 1 - \log 2} = \frac{\log 3}{\log 2} - 2$$

Consequently: $$d = \frac{\log 3}{\log 2} = 1.58$$

Thus Sierpiniski gasket has a fractal dimension of 1.58 .

"The Cantor set"

Another example of a geometric object with a fractal dimension is the Cantor set. This set is obtained by dividing the unit interval (0,1) into three equal segments, disregarding the middle third, and repeating the operation for the remaining intervals, deflation and inflation to infinity, see Fig. 6.32.

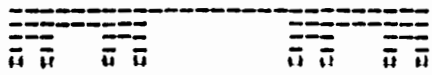


Fig. 6.32

In order to calculate the fractal dimension of the Cantor set, take either the right or the left part of Fig. 6.32 as the original unit and increase the linear size by a factor L = 3 ; we get N = 2 of the original unit. Therefore, the Cantor set has a dimension:

$$d = \frac{\log 2}{\log 3} = 0.6309$$

The definition of fractal dimension

$$d = \frac{\log N}{\log L}$$

works well, whenever there is a way to count the numbers N and L. However, it is not easy in some cases to calculate the fractal dimension of an object. Sophisticated methods are sometimes required.

"The Mandelbrot set"

The most famous and extremely beautiful icon of Chaos, the Mandelbrot set, named after the founder of the theory of fractals Benoit Mandelbrot, demonstrates clearly the highly complicated self similar geometrical structure (see Fig. 6.33). If one looks at a small region near the boundary and magnify it, we see ever tinier complicated geometric structures - spirals, blobs, seahorses, fans, trees, crystals, and so on. This small-scale intricacy - beautiful but

unpredictable and uncomputable - goes on forever. The boundary of Mandelbrot set is a fractal.

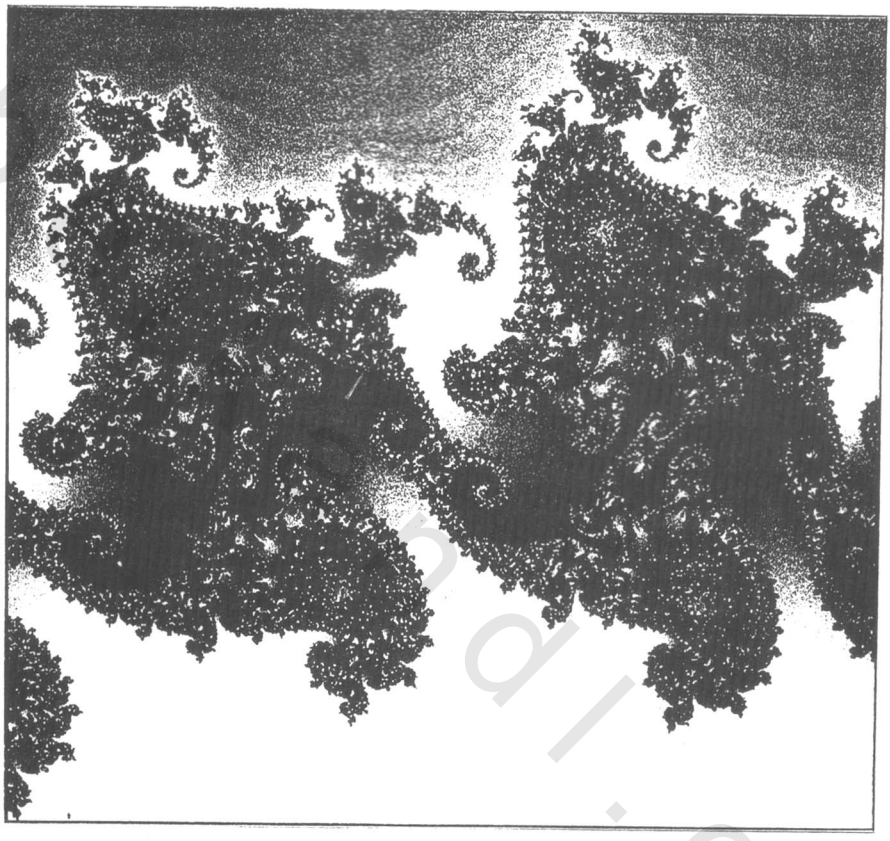


Fig. 6.33

C.1.2. Classes of geometric objects

In general, there are three classes of geometrical objects:

1. Ordinary geometrical objects where the topological dimensions (0,1,2,3) coincides with the fractal dimensions, hence there is no need for the latter.
2. Regular, infinitely nested, self-similar objects, such as the Cantor set and Sierpiniski gasket and their multi-extensions, the fractal dimension of which exceeds their topological dimension.

3. Irregular objects with self-similarity that manifests itself in a certain statistical distribution. The fractal dimensions of such objects happens to be even larger than that of the corresponding topological dimension. A typical example of that is the trajectory of a molecule in a Brownian motion, which is a continuous curve of integer "fractal dimension" 2.

C.1.3. Random fractals

Real systems in nature do not resemble those geometrically regular and self similar objects, which are called non-random geometric fractals, such as the Cantor set and the Sierpiniski gasket. Nature exhibits numerous examples of objects which, strictly speaking, are not fractals but which have the remarkable feature that, if we form a statistical average of some property such as the density, we find a quantity that decreases linearly with length scale when plotted on a double logarithmic paper. Such objects are termed **random fractals.** The first vivid description of random fractals was given by Perrin using the trajectory of a molecule in a Brownian motion. The trail is a random fractal (see Fig. 6.34).

Perrin was studying the random motion at successive moments of a tiny Brownian particle observed under a microscope. The fractal nature of the random walk of the particle could be seen directly from the photograph of the motion. If we take any part of the image and enlarge it several times we find clearly self-similarity and the same appearance as the original pattern. Fresh irregularities appear each time we increase the magnification.

The trail of a random walk with a constraint that the trail does not intersect itself was obtained by computer simulation (Fig. 6.35). It showed holes of all sizes, small and big. The trail of this constrained random walk is a model of a polymer chain. A polymer is a string of smaller molecules called monomers. A monomer itself is not a fractal, but a long string of monomers is a random fractal.

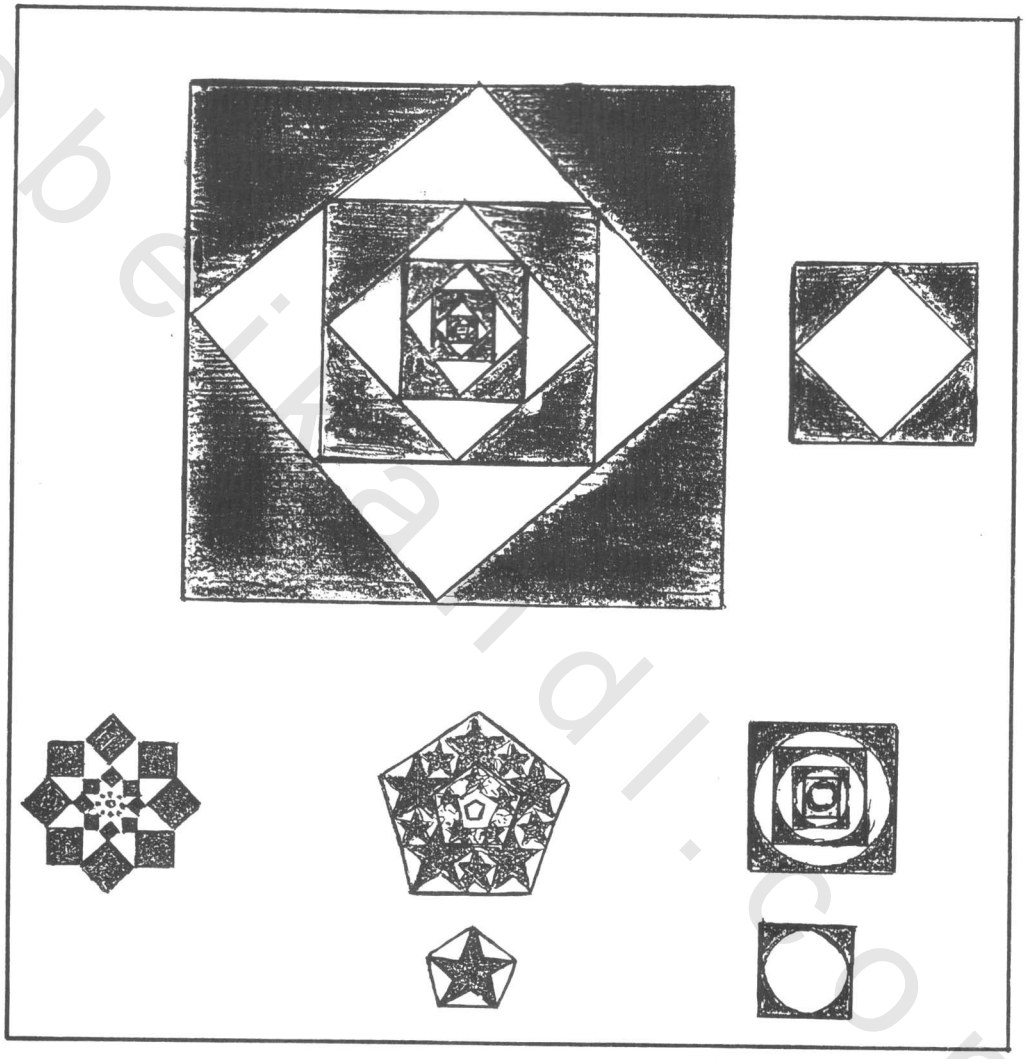


Fig. 6.34. Some examples of non-random fractals, infinitely nested and self-similar.

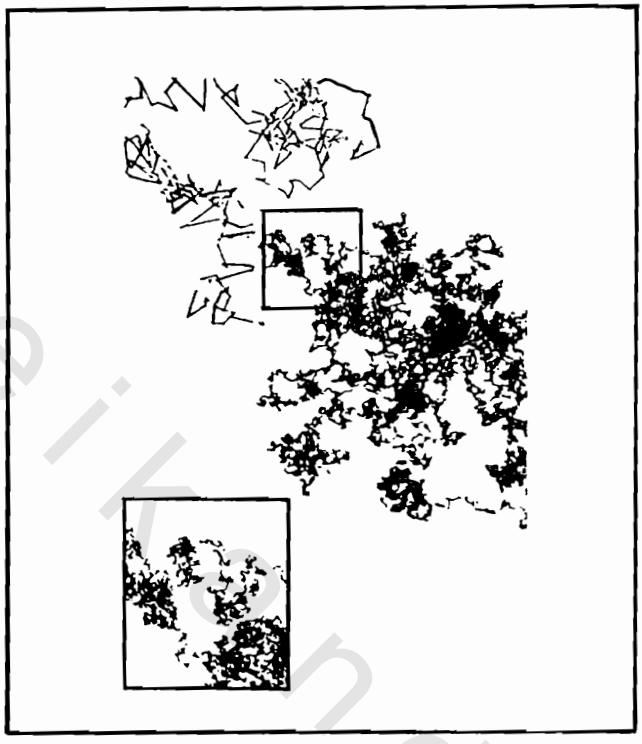


Fig. 6.35. The trail of a random walk is a random fractal. At the top the actual random motion of a tiny Brownian particle observed under the microscope. The fractal nature of a random walk is seen from the middle image, generated on a high speed computer. The magnification of a small part of the walk has the same appearance as the original pattern.

We conclude that any object for which randomness is the basic factor determining its structure will turn out to be fractal over some range of length scales, for the same reason that the random walk is a fractal. Uptill now more than one thousand fractal objects were discovered in nature. These are mainly formed by diffusion limited aggregation in open systems.

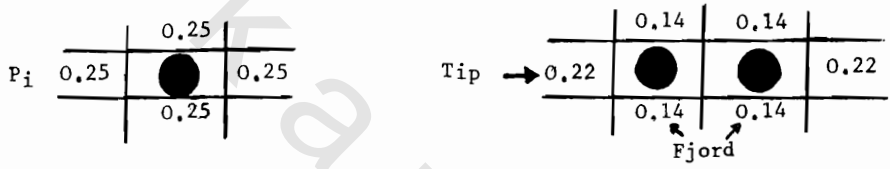
C.1.4. Diffusion limited aggregation, DLA

Diffusion limited aggregation in physical systems and the formation of fractals in nature appear in physical and chemical

phenomena such as electrochemical deposition, solidification and dendritic formation, breakdown phenomena such as dielectric breakdown, chemical dissolution and rapid crystallization of lava from volcanos.

The rule defining DLA is simple.

Consider an atom, represented by the black dot in Fig. 6.36. This atom has four perimeter sites, called growth sites which have equal probability P_i to accept other atoms going around in a random walk motion. So, we write: $P_i = 1/4$ $(i = 1, \ldots, 4)$.



(a) $t = 1$ **(b)** $t = 2$

Fig. 6.36. **(a)** Square lattice DLA at a time $t = 1$, showing the four growth sites, each with growth probability $P_i = 1/4$. **(b)** DLA at time $t = 2$, with six growth sites, and their corresponding growth probabilities written on the tips and Fjords.

At the time $t = 2$, the cluster has a mass $M = 2$. There are six growth sites (see Fig. 7.14b), but the growth probability P_i is no longer the same at the different sites. . At the tips the growth probability is maximum and amounts to about 0.22 while each of the growth sites on the sides (Fjords) has growth probability about $P_i = 0.14$.

Since a site on the tip is 50% more likely to grow than a site on the sides, the next site that receives an atom is more likely to be at the tip. It is like capitalism, in that the rich gets richer.

If the DLA growth rule is simply iterated, then we obtain a large cluster characterized by a range of growth probabilities that spans several orders of magnitude, from tips to Fjords. Fig. 6.37 shows large DLA cluster on a square lattice, showing dendritic growth

formation. It is clear that the "last to arrive" particles are never found to be adjacent to the "first to arrive" particles. Thus the growth probability, P_i, for the growth sites on the tips must be much larger than the growth probability, P_i, for the sites in the Fjords.

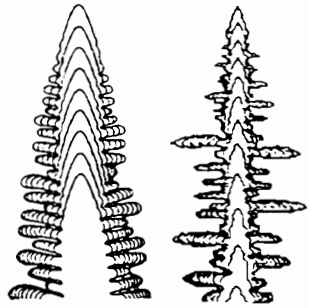


Fig. 6.37. Large DLA cluster showing dendritic formation on a square lattice.

Diffusion limited aggregation, DLA, was found applicable in biological phenomena such as the growth of bacterial colonies, neuronal outgrowth. In the last example, it could be understood why evolution chose diffusion limited aggregation, DLA, as the morphology for the nerve cell. A fractal object is the most efficient way to obtain a great deal of intercell "connectivity" with a minimum of a cell volume. Fig. 6.38 shows a typical retinal neuron and its fractal analysis.

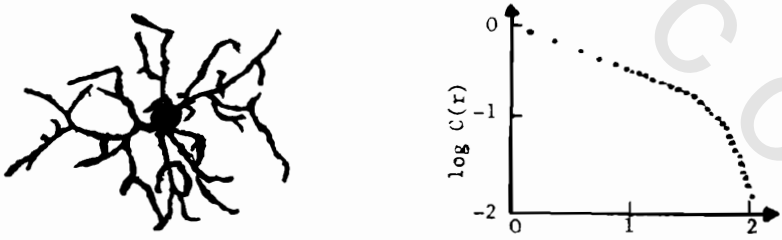


Fig. 6.38

C.2. Biological applications of fractals

DNA nucleotide sequence is considered as a chain of n-steps with short range correlations. The question is:

- Do newly formed genes have correlations with initial ones? Do they have memory?
- Is there long distance correlation between nucleotides along the DNA chain?

Fractal analysis of the DNA sequence showed that it is a fractal object with a remarkable long-range power law correlation. There exists "scale invariant property of DNA".

DNA nucleotide sequences were analyzed using models such as a chain of n-steps incorporating the possibility of short-range nucleotide correlations. It was always asked whether newly formed genes have correlations with the initial ones. Fractal analysis gave a direct answer to that; DNA sequences do have a remarkable long-range power law correlation, which meant that the DNA sequence forms a fractal object with self similarity. The scale-invariant property of DNA was done by the study of strands of DNA (see Fig. 6.39), which are found in the nuclei of cells. The DNA molecule is made by stringing together a large number of nitrogenous based

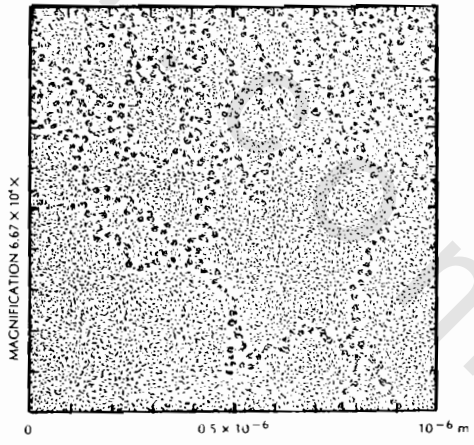


Fig. 6.39. Stands of DNA, or deoxyribonucleic acid. At intervals, the strands of DNA are wrapped around larger protein molecules that form lumps looking like the beads of a necklace.

molecules on a backbone of sugar and phosphate molecules. The base molecules are of four kinds, the same in all living organisms, but the nucleotide sequence changes from one organism to another. This sequence contains all the genetic instructions governing the metabolism, growth, and reproduction of the cell.

The importance of fractal analysis is that it provides a new method for studying long-range correlations and memory in genes.

C.3. Concluding remarks

It is now inspiring that remarkably complex objects in nature can be quantitatively characterized by a single fractal dimension, d, and can be described by various models of extremely simple rules.

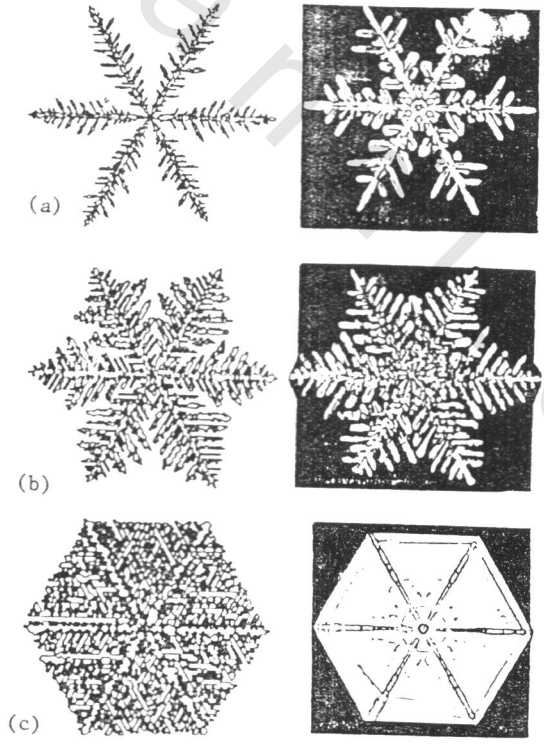


Fig. 6.40. Comparison between photographs of real snow crystals and some typical DLA computer simulations with 4,000 particles.

No two natural fractal objects that we are likely to ever see are identical, although many fractals, such as diffusion limited aggregation, DLA, have a genetic form and common character that no one can miss. For example, snow-flakes, which are fractal objects, are the same yet every snow-flake has a recognizable genetic form, Fig. 6.40.

The new thing that we learned from fractal analysis is the fact that geometrical models - with no Boltzmann factors - are sufficient to describe features of real statistical mechanical systems. If we understand DLA, then we can understand variants of DLA. Thus, DLA may be a paradigm (model) for all kinetic growth models.

D. CHAOS AND NATURE PHILOSOPHICAL COMMENTS

D.1. Application of Chaos

The study of Chaos brought to light an important form of motion that exists everywhere in nature but has been overlooked for centuries.

At the most primitive level, in simple mechanical or electrical systems, Chaos usually belongs to the type of motion that one manages to avoid. In this domain there appeared recently the possibility of controlling Chaos.

In higher state of matter and motion, Chaos is of primary importance. The notion of Chaos may help us bridge the long-standing gap between the two opposite systems of description for the totality of nature, namely, the deterministic and the probabilistic descriptions, with a view toward deepening our understanding of such phisolophical categories as chance and necessity. The following examples might explain more the idea.

In the superconducting Josephson junction, superconducting tunneling through a thin layer of isolator that separates two superconductors may be used to build low-noise parametric amplifiers. It was observed that a Josephson junction put in a microwave cavity may exhibit anomalous noise together with the increase in gain. The experiment was carried out at temperature as low as 4 K, but the noise may have an equivalent temperature as high as 50,000 K. This could not be explained by any mechanism known before Chaotic motion was discovered. The system proved to enter the Chaotic regime and the noise is generated by the dynamics itself.

There are many other examples such as:

1. The ferromagnetic parametric amplifier that was used until late 1950s and later passed from use due to its high noise level. The reason might have been in Chaos too, the second subharmonic was observed experimentally.

2. The escape of particle beams in high energy accelerators.
3. The leakage of the confining magnetic fields in a device of controllable thermonuclear fusion.
4. The harmful back flow of cycling water in a nuclear power reactor.
5. The instabilities in an optical bistable device.

Dangerous accidents might occur to an articulated mooring tower used for loading oil tankers near deep offshore oil installations due to Chaotic oscillations developed by the subharmonics created under the periodic driving of the steady surge of ocean waves and the non-linear slackening of the mooring line, see Fig. 6.41.

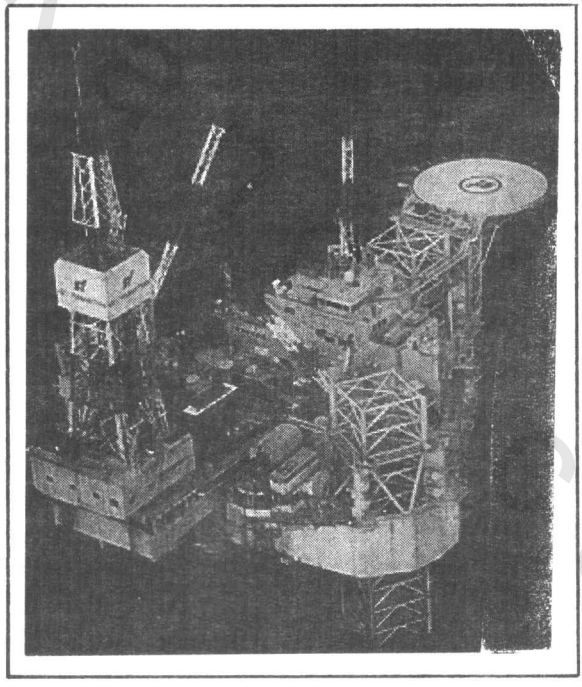


Fig. 6.41

All above examples come under the type of Chaos that should be avoided.

D.2. Natural phenomena to be explained by Chaos

There are many complicated dynamical processes in geophysics. The magnetic field of the Earth has changed randomly its orientation many times during the last several millions of years.

Southern oscillations of the ocean temperature which may reach an amplitude as high as 4 degrees affects the global weather in an erratic way.

Chaotic regimes have been used to improve the combustion efficiency of gaseous or powdered fuel, or to increase the effect of plasma heating.

The understanding of Chaos for hydrodynamics and aerodynamics are crucial for aircraft and spacecraft design.

In life phenomena Chaotic behavior certainly provides further inspirations.

Various biological rhythms are neither completely periodic nor purely random. They have tendency to look into natural periods such as seasons, day and night, etc., and at the same time they are capable of preserving their own autonomous features. Many types of biological rhythms may be simulated by coupled non-linear oscillators or oscillating chemical reactions.

Heart beat was simulated by non-linear electric circuits.

Recent physiological experiments revealed a possible connection between Chaos and cardiac arrhythmia, atrioventricular blocks, electrocardiogram (ECG) and probably, ventricular fibrillations.

The electroencephalographic (EEG) signals in epiliptic seizures show clear patterns of periodicity, while the brain waves of a normal person look much like random signals. The measurement of dimension has shown that the EEG waves of normal brain activity are not random noise, but may have been originated from dynamical processes on attractors of not very high dimensions (2 to 5).

Although for the time being we are far from a genuine understanding of brain dynamics, the experimental and model study of neuron networks and brain activity is certainly becoming a subject of concern in physics.

D.3. Philosophical comments

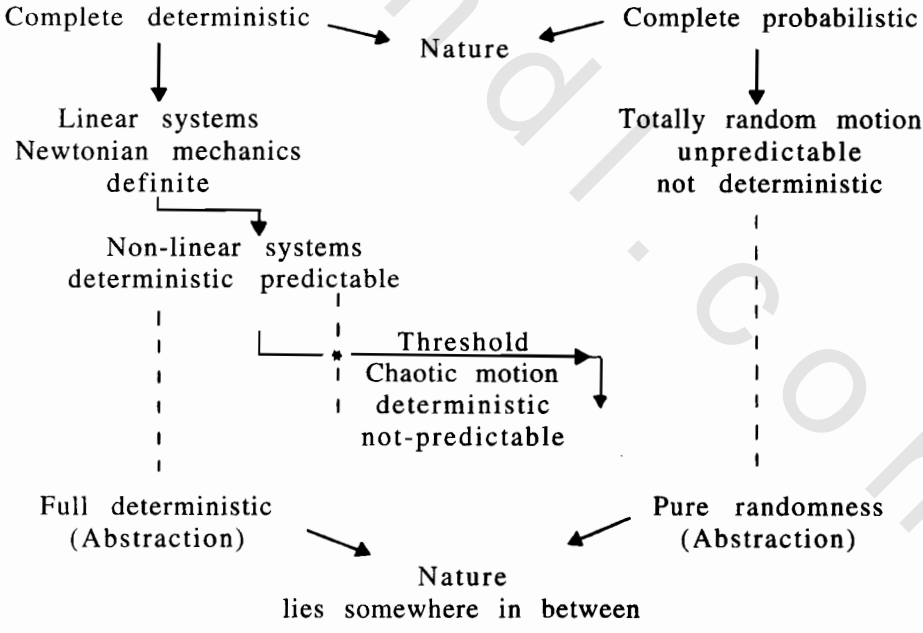
Nature is a unified integrity. However, our natural science has two systems of description:

1. deterministic, and
2. probabilistic .

Scientific tradition since the time of Newton has held the deterministic system in high esteem, with the probabilistic description as a kind of involuntary complement.

Yet, as we approach closer to higher forms of matter, its structure and motion require more statistical arguments.

Full determinism and pure randomness are both abstraction. Nature lies somewhere in between. In this sense the notion of Chaos may help to reach a better understanding of nature.



E. ELEMENTARY PARTICLES

Since long time ago scientists searched for the elementary brick forming matter. Early in this century, the atom was discovered and was thought at that time that it was an indivisible part of matter. Afterwards, Rutherford presented his nuclear model of the atom stating that "the atom is formed of a massive nucleus having positive charge surrounded by electrons moving in orbits around it". Protons and neutrons formed the nucleus, strongly bound together by strong forces of attraction despite Coulomb repulsion between the similar charges. Thus, until the thirties of this century, the main bricks of matter were the protons, the neutrons and the electrons.

In order to attempt breaking to pieces these elementary bricks of matter, huge accelerators were built. Since then, a great number of elementary particles were discovered as a result of bombarding these particles with each other after giving them huge momenta. Conservation laws of momentum and mass were found to prevail. This implied that electrons and protons have themselves infra-structure formed of much smaller particles and of various types.

At present, there is a standard model that explains hundreds of subatomic particles and their properties by postulating six basic constituents called quarks and another six called leptons from which all matter is made.

It is now well accepted that there are four fundamental forces: electromagnetic, gravitational, weak, and strong.

Students are well familiar with electromagnetic interactions. Similar electrical charges repel each other and different charges attract. Inside the nucleus strong forces of attraction overcome the much smaller forces of Coulomb repulsion between protons. Gravity forces are also too weak.

In radioactive decay, we now know that the alpha ray is actually a helium nucleus emitted in a spontaneous nuclear fission, in a strong interaction process. The gamma ray is an energetic photon emitted in a transition involving electrical interactions. The beta ray, however, is inexplicable on the basis of strong or electrical interactions. It was

found that the beta's were electrons coming from a transition in which a neutron changes into a proton, emitting the beta and an antineutrino. To explain this process requires a new type of interaction, which occurs relatively slowly compared to the emission of comparable energy gamma rays in electrical interactions. It is called the weak interaction.

The three interactions discussed above - strong, electromagnetic, and weak - are now described by a theory called: **the standard model.**

E.1. The fundamental forces in nature

All particles in nature are subject to four fundamental forces. These are:

1. **The strong force** is very short-ranged force and is responsible for the binding of neutrons and protons into nuclei. This force represents the glue that holds the nucleons together. These forces are only operative at distances less than 10^{-14} m, which is about the size of the nucleus.

2. **The electromagnetic force** is 0.01 times the strength of the strong force. It is responsible for binding atoms and molecules. It is a long-range force obeying the inverse square law of the separation between the interacting particles.

3. **The weak force** is a short-range nuclear force that tends to produce instability in certain nuclei. It is responsible for radioactive decay processes. Its strength is 10^{-9} times that of strong force.

4. **The gravitational force** is a long-ranged force that has a strength 10^{-38} times that of the strong force. This force holds the planets and galaxies and stars together. It is very weak and has negligible effect on elementary particles.

Photons: are the field particles coming out of electromagnetic interaction.

Gluons: are the field particles that mediate strong forces.

Bosons (W and Z): are the particles mediating weak forces.

Gravitons: are quanta of gravitational field mediating the gravitational forces.

E.2. Particles and antiparticles

In 1947, Dirac presented a theory based on quantum mechanics, implying that for every particle there should be an antiparticle having the same mass but of opposite charge. For example, the electron's antiparticle is called **the positron** having a mass of 0.511 MeV, and a positive charge of 1.6×10^{-19} C. Usually we shall designate an antiparticle with a bar over the symbol for the particle.

The positron, e^+, was discovered by Anderson in 1932 in the Wilon cloud chamber. The most common process in producing positrons is the process of pair production, in which a gamma photon of at least 1.02 MeV energy, transforms to an electron and a positron. If extra energy is available, the two particles share it together and appear as kinetic energy.

Particles like proton and antiproton, neutron and antineutron, nutrino and antinutrino, etc. have already been discovered recently.

E.3. Pions and muons

In the same way as the exchange of electrons in a covalent chemical bond of two atoms, Yukawa explained the strong force by proposing a new particle that is exchanged between nucleons in the nucleus and is responsible for this strong force. He predicted that the mass of this particle is about 200 times the electron mass, and he called it a **meson.** In 1937, Anderson discovered experimentally the existence of the π-meson, or the **pion**, and the lighter meson (μ) now called **muon.**

The pion comes in three varieties corresponding to their charge states: π^+, π^- and have masses of 139.6 MeV/c^2 , while π^o has a mass of 135.0 MeV/c^2.

Pions and muons are very unstable particles, they decay according to the following sequence:

$$\pi^- \longrightarrow \mu^- + \upsilon^-$$

Then:

$$\mu^- \longrightarrow e + \upsilon + \bar{\upsilon}$$

υ and $\bar{\upsilon}$ are nutrino and antinutrino.

In the same way as virtual photon mediates the electromagnetic force between two interacting electrons (see Fig. 6.42), a proton and a neutron interact in the nucleus by pion exchange via the strong force.

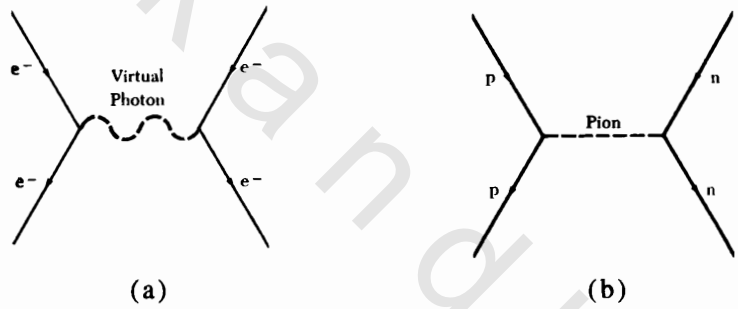


Fig. 6.42. Feyman diagrams showing: (a) A photon mediating the electromagnetic force between two interacting electrons. (b) A proton interacting with a neutron via the strong force, in this case, the pion mediates the strong force.

The graviton, which is the mediator of the gravitational force, has yet to be observed.

In 1983, at CERN, the force carriers $W^{\pm}$ and Z^o boson particles, were discovered using a proton-antiproton collider. In this huge accelerator, protons and antiprotons that have momentum of 270 GeV/c undergo head-on collisions with each other, and so the $W^{\pm}$ and Z^o particles were produced and identified, by their decay products.

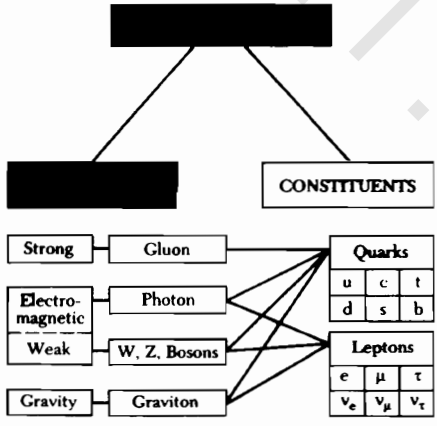
E.4. The quark model

In 1963, it was announced that all elementary particles, (baryons and mesons) have infrastructure. The fundamental constituents are called **quarks** designated by the symbols u, d, s. These were given arbitrary names up, down, and strange. The most unusual property of quarks is that they have fractional electronic charges. The u, d, s quarks have charges of +2e/3 , -e/3 and -e/3 , respectively. They all have spin $1/2$. Associated with each quark there is an antiquark of opposite charge.

The composition of all known elementary particles could be specified by three simple rules:

1. Mesons consist of one quark and one antiquark.
2. Baryons consist of three quarks.
3. Antibaryons consist of three antiquarks.

At present, the building blocks of matter are believed to be formed of six quarks and six leptons (together with their antiparticles). Some of the properties of these particles are given in the following table.



E.5. The standard model theory

The standard model theory describes all observed particle processes. It explains the structure of matter and the interactions responsible for all processes, down to a scale in which protons and neutrons are themselves composite particles made up of constituents called quarks. All of the conservation laws of physics are built into the standard model. It is these laws, along with the dynamics of the interactions, that explain particle lifetimes and decay patterns. The beauty of the standard model lies in the fact that hundreds of elementary particles and processes can be explained on the basis of a few types of quarks and leptons and their interactions.

The following explanation of the standard model does not include the extensive mathematical structure that allows physicists not just to name and describe particles but also to predict which particles can exist and which cannot, to calculate the rates of variety of processes, and to make quantitative predictions about the outcome of experiments.

E.5.1. Fermions and bosons

We distinguish between fermions and bosons in order to understand the behavior of quantum mechanical particles. The following two tables and Fig. 6.43 give information about matter

Fermions spin = $\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots$

Leptons (s = $\frac{1}{2}$, ℓ & $\bar{\ell}$)

Electric charge	Flavor	Mass GeV/c^2	Flavor	Mass GeV/c^2	Flavor	Mass GeV/c^2
0	υ_e	$< 2.0 \times 10^{-8}$	υ_μ	$< 2.5 \times 10^{-4}$	υ_τ	$< 3.5 \times 10^{-20}$
1	e	5.1×10^{-4}	μ	0.106	τ	1.784

ℓ = lepton; $\bar{\ell}$ = antilepton.

υ_e = electron nutrino; υ_μ = muon nutrino; υ_τ = tau nutrino;

e = electron; μ = muon; τ = tau.

Quarks ($s = \frac{1}{2}$, q & $\bar{q}$)

Electric charge	Flavor	≈ Mass GeV/c^2	Flavor	Mass GeV/c^2	Flavor	Mass GeV/c^2
$\frac{2}{3}$	u (up)	4×10^{-3}	c (charm)	1.5	t (top*)	> 41
$-\frac{1}{3}$	d (down)	7×10^{-3}	s (strange)	0.15	b (bottom)	4.7

q = quark; $\bar{q}$ = antiquark
* Not yet observed.

Bosons — Force carriers spin = 0, 1, 2, ...

Unified Electro-weak spin ≈ 1	γ photon	W-	W+	Z^o
Electric charge	0	-1	-1	0
Mass (GeV/c^2)	0	81	81	92

Strong or color spin -1	g gluon
Electric charge	0
Mass (GeV/c^2)	0

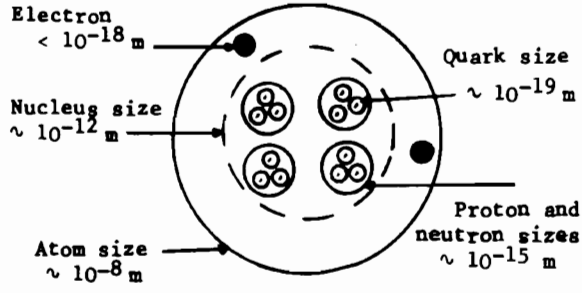


Fig. 6.43. Structure of atom and dimensions.

constituents and the force carriers. The figure shows the structure of the atom and the order of magnitude of dimension of each particle.

In the standard model, a particle experiences an interactions if and only if it carries a charge associated with that interaction. The electric charges for all particles are given on the table. There are weak charges, which are associated with quark and lepton flavor. The charges of the strong interaction are called color charges (or sometimes just colors) and are carried by quarks and by gluons. Particles that are are composites do feel some residual effects of an interaction for which their constituents carry a charge, even though overall the composite may be neutral.

Angular momentum is essential to understanding the behavior of particles. Particle spin is not intuitive, especially as particles do not spin on their axes, but simply act as if they were points. Analogy with the solar system helps in showing the student how an object can have both a spin and an orbital angular momentum. Spin must be included in accounting for conservation of angular momentum in particle processes.

The quantum unit of angular momentum is

$$\hbar = h / 2\pi = 6.58 \times 10^{-25} \text{ GeV.s} = 1.05 \times 10^{-31} \text{ s}$$

Some particles carry half units of their intrinsic angular momentum. Any particle with spin that is an odd number of half units of $\hbar$ is a fermion. Any particle with an integer number of units of $\hbar$ is a boson. The Pauli principle states that: "two fermions cannot occupy the same state at the same time".

All matter is made from quarks and leptons. The electron is the most familiar example of a lepton. Leptons have no color charge, which means that they have no strong interactions. They are particles that can be observed in isolation.

Quarks have color charge. Hadrons may be either: (a) fermionic, made from three quarks and are called **baryons**, or bosonic, made of a quark and an antiquark, called **mesons.**

All hadrons have residual strong interactions due to their quark constituents. All hadrons consisting of three quarks have half-integer spin and are called **baryons**. For example, the proton is baryon,

having the quark content uud. These are color charge neutral combinations made from one quark of each of the three possible quark colors.

Mesons are hadrons consisting of a quark and an antiquark. Mesons can have any integer (0, 1, 2, ...) in units of $\hbar$, and thus they are bosons. The color charges of the quark and antiquark must be combined to form a color-neutral state.

Most hadrons are composites of quarks and gluons. In principle, some particles could be made only from gluons. There is as yet no experimental evidence for such objects. All color-charged particles are confined by the strong interaction. They can never escape to be observed as free quarks.

F. COSMOLOGY PHYSICS

Cosmology physics became a serious branch of physics only after the year 1960. Astronomical observations confirmed the idea that the universe was expanding from a "Big Bang". The study of the rates at which galaxies are receding from us, coupled with improved determinations of their distances, implied that the universe was at least 10 billion years old. This finding was consistent with other observations such as the age of the Earth and of old star clusters.

In 1965 came the discovery of the cosmic background radiation which remained after the first creation of the universe. This hot electromagnetic black body radiation that dominated the universe since its creation, now cooled to 3 degrees Kelvin by the subsequent expansion of the universe. Important conformation of the hot Big Bang picture came from the study of the amount of helium that would be synthetized from hydrogen by thermonuclear fusion in the very early stages of the creation of the universe. This amount proved to be approximately 25% by weight, in agreement with the abundance of helium observed in stars and in interstellar space. The laws of elementary particle physics play a major role in giving an idea about the evolution of the universe.

F.1. The expansion of the universe

Before we give a theoretical analysis of the universe, we should make a basic hypothesis: the laws of physics that we know are the same in all parts of the universe. Newton's law of gravity will be the main tool of investigation although it is not entirely adequate for a description of the universe because at large cosmic distances, relativistic effects become important.

In a clear night sky, one could easily observe the many stars forming our galaxy, the Milky Way, which contains about 10^{11} stars arranged in an irregular dislike region of diameter 10^5 light years approximately. There are many other galaxies beyond ours. They are all in motion. The recession velocities of galaxies can be

determined by the red-shift method. This method uses the fact that the Doppler effect is applied, and the light from a receding source will look redder to us than the light from a stationary source. The change in wavelength is directly related to the velocity. The increase in wavelength $\Delta\lambda$ will thus be

$$\Delta\lambda / \lambda = - v / c$$

v being the velocity and c is the velocity of light, see chapter one.

To find the recession velocity of a certain celestial body, we need only to measure the color of the light received from atoms there and compare it with the color emitted by similar atoms in our laboratories on Earth.

All velocities measured of distant galaxies were found to be directed away from us, i.e. all celestial bodies are moving in recession relative to our galaxy, see Fig. 6.44.

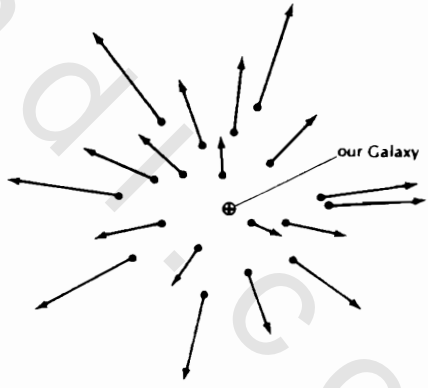


Fig. 6.44. The recession motion of different galaxies showing the expansion of the universe.

F.1.1. Hubble's law

Hubble discovered that the motion of recession obeys a very simple rule: "The velocity of each galaxy is directly proportional to its distance". This law, known as Hubble law, indicates that nearby galaxies move slowly and distant galaxies move fast. Mathematically, Hubble's law is written as:

$$v = H_o r$$

where H_o is Hubble's constant. If the distance r is expressed in light years, the numerical value of H_o is given by: $H_o = 1.7 \times 10^4$ ($m.s^{-1}$/million light years).

In order to understand the idea of expanding universe, consider a bomb that explodes into fragments in midair. The fragments will go in all directions. Different fragments may have different velocities, and after some time, at a given instant, they will reach different distances.

After a time, t, the velocity of a fragment is related to its distance by:

$$v = r / t$$

Thus we see that the fragments with highest velocities are at the greatest distances. Hubble's law, thus suggests that the galaxies were set in motion some time ago through a primordial cosmic explosion. It is only the distance between the galaxies that increase but the galaxies themselves do not expand.

F.1.2. The age of the universe and the Big Bang

The explosion that started the expansion of the universe is called the Big Bang. We can estimate how long ago this did happen by using the above two equations from which we find that the inverse of Hubble's constant must coincide with the expansion time

$$t = 1 / H_o = \frac{1}{1.7 \times 10^4} \left(\frac{\text{million light years}}{m.s^{-1}}\right)$$

$$= \frac{1}{1.7 \times 10^4} \times (9.5 \times 10^{21}) \ (m/m.s^{-1})$$

$$= 5.6 \times 10^{17} \text{ s} = 1.8 \times 10^{10} \text{ years}$$

The above naive calculation predicts the age of the universe to be about 18 billion years. Since the universe started some finite time ago, only the light from those parts of it that are sufficiently near can

have reached us. The speed of light is 3×10^8 $m.s^{-1}$. In a time t, since the Big Bang, light has traveled a distance:

$$c\,t = 1.8 \times 10^{10} \text{ light year}$$

This distance is the radius of the observable universe, since everything within this radius we can see; but anything beyond we cannot see because the light has not yet had enough time to reach us.

F.1.3. Black holes

A black hole is formed when a star exhausted the thermonuclear fuel necessary to produce the heat and pressure that support it against gravity. The star begins to collapse until the radius of the star approaches a critical value called **Schwardzchild radius.** After the collapse nothing is left from the star except an extremely intense gravitational field. A black hole is black because even light photons cannot escape the overwhelming grip of its gravitational field. No particle can ever emerge out of a black hole. Since we get our information about the universe around us from the light reaching us, and since light is pulled down into a black hole which acts as one-way membrane, therefore black holes are considered as our horizon of information. No one can ever know what is behind them.

In order to get the critical radius that changes a star to a black hole, consider the escape velocity from a celestial body, v, in terms of its mass, M, and radius, R:

$$v = (2\,G\,M / R)^{1/2}$$

If we take v = c, the velocity of light, then the radius within which the mass M must be constrained for escape to be impossible is

$$R = 2\,G\,M / c^2$$

This gives the critical radius called **the Schwartzchild radius.** For a body like Earth, the equivalent black hole will be a sphere of radius less than one cm. A black hole of mass equal to that of the sun will have a diameter of about 3 kilometers.

Newton's laws of motion are not adequate to describe the motion of bodies in the interior of a black hole. Instead, Einstein's theory of general relativity should be used. According to the **general relativity**, the space and time near any gravitating body are curved and becoming extreme inside the black hole. We try to explain this concept very briefly in the following section.

F.2. The curvature of space and time

The special theory of relativity introduced the concept of space and time forming a four-dimensional space-time. The general theory requires yet another change in our view of space-time. Instead of a flat space-time, it is curved or warped in the vicinity of a mass. The curvature is in the four-dimensional space-time, and it is not easy to visualize it by the three-dimensional man.

The distinction between a flat space and a curved space can be illustrated in two dimensions. The surface of a plane is flat, whereas the surface of a sphere is curved. These two dimensional surfaces have very different geometrics. For example, the shortest distance between two points in a plane is a straight line. But on the surface of a sphere, the shortest path between two points is along a great circle.

Consider a two-dimensional creature, an ant for example, constrained to move on the surface of a spherical Earth. It knows left from right, and forward from backward, but it does not know the concept of up and down. Going straight forward from a point A to another point B, the ant might think that it is travelling in a straight line. But viewed in three dimensions, the path is curved.

General relativity treats gravitation as a curvature of space-time in four dimensions. The curvature is determined by the presence of mass. Gravity is thus explained in terms of geometry. Fig. 6.45 shows schematically a two dimensional surface with a "Sun" at the center, and with a black hole. For a black hole there is a circle of no return, and at the kink at the bottom, the curvature of the surface is infinite.

Now, consider the motion of Earth around the Sun. Because of its large mass, the Sun distorts space-time in its vicinity. The Earth moves along the shortest path between two points in the curved

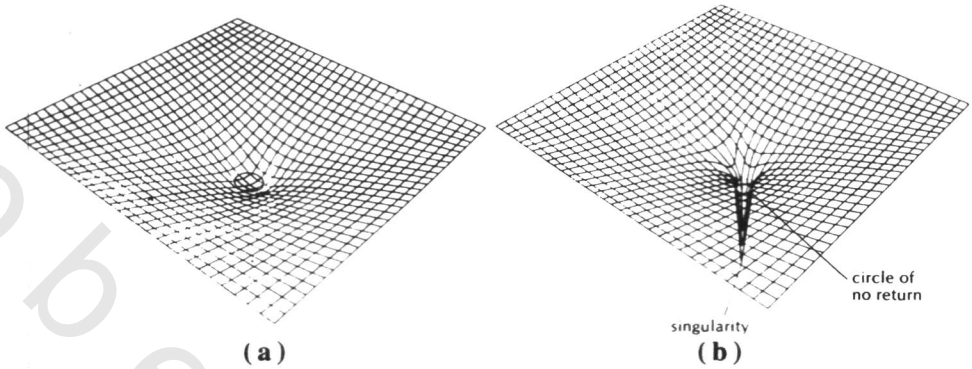


Fig. 6.45. **(a)** Two dimensional surface with a Sun at the center. **(b)** Two dimensional surface with a black hole at the center.

space-time. In this view, no force acts on the Earth; and the distortion of space-time is itself the gravity. Of course, we can only see things from our "flat-space" perspective, so we see the path taken by the Earth as an ellipse.

The effects of the curvature of space-time are particularly important near massive stars or black holes. Cosmological theories say that inside a black hole, space and time are - in a sense - interchanged. The distance from the circle of no return (Fig. 8.4) to the central singularity, is not a distance in space but, rather an interval of time, i.e. the singularity is not a point in space, but a point in time.

F.2.1. Curvature of space and the astronaut

A clock placed near a gravitating body will run slow. This gravitational time dilation is one aspect of the distortion of space and time in the vicinity of a gravitating body. Near the Earth the distortion is small, and the time dilation is small. But near a black hole the distortion and time dilation are very large. At the circle of no return, the time dilation is infinite, and a clock placed there will appear to have stopped.

If we imagine an astronaut in a space capsule parked at the edge of a black hole, all his life functions (pulss. rate, voluntary and

involuntary muscular contraction) will almost be at a stand still - to the scientists at "control room" positioned far away, outside the strong gravitational field. The astronaut will appear to them as frozen. Yet the astronaut will not perceive himself as having slowed down; he would merely perceive the surrounding world as having speeded up.

F.2.2. Detection of black holes

Black holes could not be seen at all. However, we could detect their presence through the intense X-rays generated when a body is crushed by going through it. As the body falls toward the black hole, it is accelerated by the strong gravitational field and heated by compression. The material might reach a temperature of 100 million Celsius. At such temperatures the violent collisions between the particles of the material release X-rays of high intensity that could be observed on Earth, see Fig. 6.46.

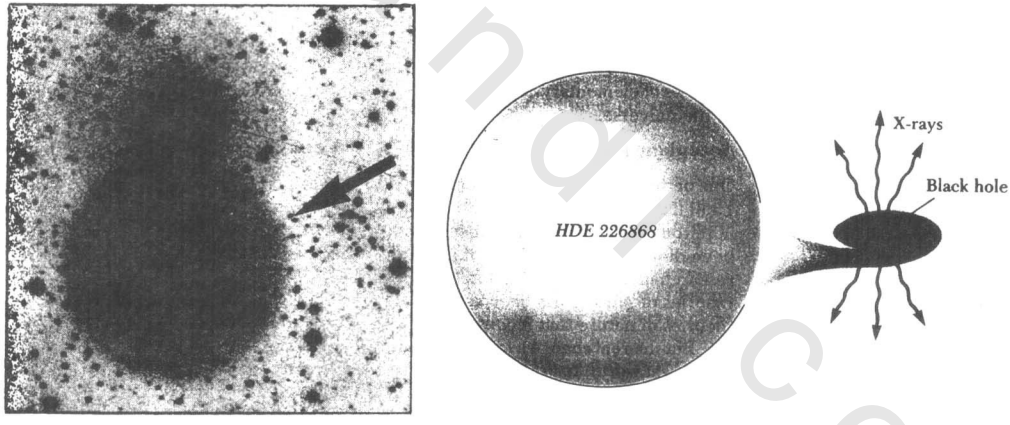


Fig. 6.46. Cygnus X-1 , the first black hole to be discovered.

F.3. The future of the universe

In this section we consider the expansion of the universe and the recession velocities of all galaxies and celestial bodies. The question now is whether it will continue expanding indefinitely, or the gravitational attraction of these galaxies will finally overcome this

recession motion thus bringing about contraction instead of expansion. It has been found that the recession velocities decrease with time thus tending to decelerate the expansion. On the long run this very small deceleration might stop the motion completely, and the galaxies will begin to fall back toward each other. Finally, the galaxies will collide with one another and the universe will collapse in a terminal cosmic implosion.

In order to discuss the assumptions: indefinite expansion or future contraction of the universe, we have to consider the average density of the universe. We assume a uniform distribution of the galaxies throughout the universe, forming "a gas of galaxies". The expansion of the universe is then equivalent to the expansion of this gas. Consider a spherical region centered on our galaxy, as shown in Fig. 6.47. Assume that this region is free from all bodies and the rest of the universe has spherical symmetry about this region. Gravity in the region will be zero. The motion of any galaxy in this region will not be affected by the rest of the universe, only the galaxies in this region will exert gravitational forces on each other.

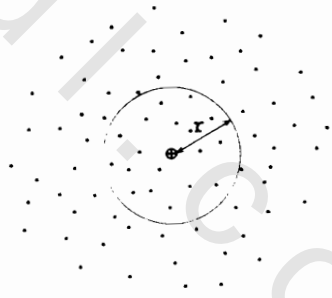


Fig. 6.47. Spherical region of our universe.

Consider one galaxy at the surface of the spherical region at a radial distance r from our galaxy at the center. The gravitational force that the mass in the spherical region exerts on that galaxy will produce an acceleration, a, given by

$$a = - G M / r^2$$

where M is the mass of material in the spherical region which is supposed to be concentrated at the center of the sphere.

This is the same equation as that of a projectile sent from the Earth. The radial motion continually decelerates, but whether it ever stops and reverses direction depends on the initial velocity. If it is larger than the escape velocity, the radial distance continues to increase forever; if the initial velocity is smaller than the escape velocity, the radial motion stops after sometime then reverses its direction.

Regarding the present instant as the initial instant and the present radius, r, as the starting radius, then the initial velocity is given by Hubble's law:

$$v_o = H_o r_o$$

and the escape velocity is

$$v_{esc} = (2 G M / r_o)^{1/2}$$

In these equations the subscript (o) refers to the present instant. The mass M refers to the average density of mass in the universe, d_o :

$$M = 4 \pi r_o^3 d_o /3$$

Thus, $$v_{esc} = (8 \pi G r_o^2 d_o /3)^{1/2}$$

In order that the universe expands continuously: $v_o \geq v_{esc}$, which is equivalent to:

$$H_o r_o \geq (8 \pi G r_o^2 d_o /3)^{1/2}$$

or: $$d_o \leq 3 H_o^2 / 8 \pi G$$

Likewise, the condition for an ultimately contracting universe is

$$d_o > 3 H_o^2 / 8 \pi G$$

Inserting the numerical values of H_o and G in the above equations we get the following conditions:

1. For a permanent expansion $d_o \leq 5.7 \times 10^{-27}$ kg/m^3 .
2. For an ultimate concentration $d_o > 5.7 \times 10^{-27}$ kg/m^3 .

According to this analysis we can predict the future of the universe if we measure the average density of mass in the universe and find out whether it is larger or smaller than the critical value 5.7 x 10^{-27} kg/m^3 . Unfortunately, too many uncertainties exist in such measurement and so it is hard to tell definitely whether the universe will expand permanently or not.