(1) Equilibrium phase transition

A balanced phase transition refers to a transition of a balanced microstructure that can be obtained in a slow heating or cooling rate. The equilibrium phase transitions of metal materials in solid state are mainly as follows:

a) Allotropic transformations and polymorphic transformation

When temperature and pressure change, the process of converting pure metal from one crystal structure to another is called allotropic transformation. It is called a polymorphic transformation when it occurs in a solid solution. For example, the transition from ferrite to austenite or austenite to ferrite when steel is heated or cooled belongs to polymorphic transformation.

b) Equilibrium dissolution precipitation

In the case of slow cooling, the process of precipitating the excess phase from the supersaturated solid solution is called equilibrium dissolution precipitation. As is reflected by the schematic A–B binary alloy phase in Fig. 9.1, when the alloy of b component is heated to T_1, the β phase will completely dissolve into α phase to form a single solid solution. If the temperature is cooled slowly from T_1, β phase will be continuously precipitated along the solution curve MN, which is called equilibrium dissolution precipitation. The characteristic is that the parent α phase does not disappear, but as the new β phase precipitates, the component and volume fracture of the parent phase change constantly. The structure and composition of the new phase are different from that of the parent phase, and the composition of the new phase is generally changed.

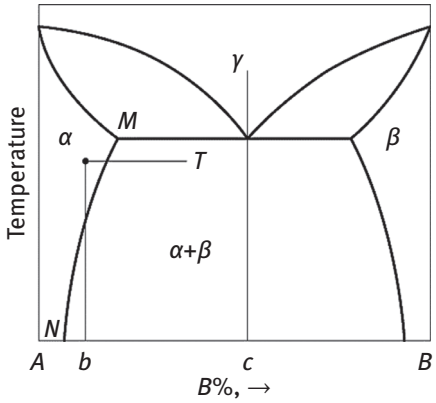


Fig. 9.1: The binary equilibrium phase diagram containing dissolution precipitation.

c) Eutectoid transformation

When the alloy is cooled, the transformation from one solid phase to two different solid phases is known as eutectoid transformation. As shown in Fig. 9.1, when the alloy of the c component is cooled slowly from the y state, a eutectoid transition occurs when the temperature is lower than the critical temperature: $y \rightarrow \alpha + \beta$. The

eutectoid phase is similar to the eutectic reaction in eutectic crystallization, and the structure and component of the two phases are different from those of the parent phase. $\alpha + \beta \rightarrow y$ can also occurs during heating, which is called inverse eutectoid phase change. For example, the transition of austenite(y) to pearlite($\alpha + Fe_3C$) during cooling ($y \rightarrow \alpha + Fe_3C$) and the transition from pearlite to austenite during heating ($\alpha + Fe_3C \rightarrow y$) belong to eutectoid and inverse eutectoid transition.

d) Spinodal decomposition

Some alloys have homogenous single-phase solid solution at high temperature, but when cooled to a range of certain temperatures they can be decomposed into two micro zones with the same structure but different composition, like $\alpha \rightarrow \alpha_1 + \alpha_2$. This change is known as spinodal decomposition. The characteristic of spinodal decomposition is that there is no obvious interface and component mutation between the two microzones formed at the initial stage of change, but the homogeneous solid solution becomes inhomogeneous through uphill diffusion finally.

e) Order transformation

In the solid solution (including the solid solution based on mesophase), the relative position of the atoms in the crystal lattice is changed from disorder to order (refers to long-range order), which is called order transformation. This order transformation can occur in many alloy systems, such as Cu–Zn, Cu–Au, Mn–Ni, Fe–Ni, and Ti–Ni.

(2) Nonequilibrium phase transition

If heating or cooling rate is fast, the equilibrium phase transitions above will be suppressed. Solid materials may undergo transformation that cannot be reflected in the equilibrium diagram and obtain an metastable structure, which is called the nonequilibrium phase transition. The nonequilibrium phase transition is mainly listed as the following:

a) Pseudoeutectoid phase transition

Figure 9.2 is the bottom left part of the Fe–C alloy equilibrium diagram. When the austenite (y) is cooled slowly below the *GES* line from the high temperature, the ferrite (α) or cementite (Fe_3C) will be precipitated, while the carbon content of the austenite is close to the *S* point. When the carbon content reaches the *S* point, it will be transformed into pearlite ($\alpha + Fe_3C$) through eutectoid transformation. But if the cooling rate is faster, those changes cannot be carried out. The austenite of noneutectoid component will be supercooled to the temperature below the extension of GS and ES lines (the shadow area in Fig. 9.2), and ferrite and cementite will be precipitated simultaneously. The transition process and transition product are similar to eutectoid phase transition, but the ratio of ferrite to cementite (or the average

composition of the product) is not constant which varies according to the carbon content of austenite. This is called pseudoeutectoid phase transition.

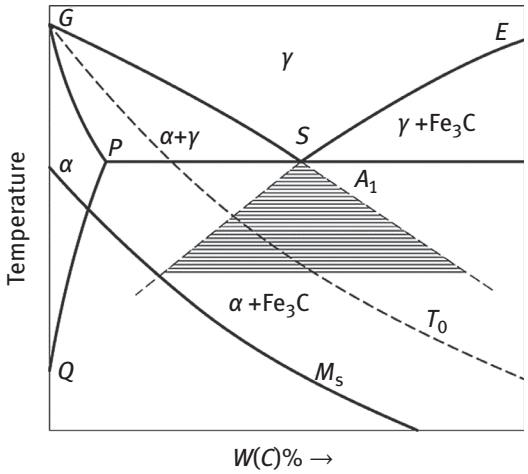


Fig. 9.2: A part of the equilibrium phase diagram of Fe–C alloy.

b) Martensite phase transition

Also, taking Fe–C alloy as an example, if the cooling rate is further improved, the pseudoeutectoid phase transition cannot be carried out, and the austenite will be supercooled to lower temperature. The austenite can only be transformed into the α lattice by shearing from the y lattice in the absence of atomic diffusion and composition change, because iron atoms and carbon atoms are not able or easy to diffuse at low temperature. This is called martensite phase transition, and the transition product is called martensite (in order to distinguish the α phase formed in the equilibrium phase change, it is called α' phase) which has the same composition as the parent's austenite. In Fig. 9.2, T_0 is the temperature at which the α' phase and y phase of the same composition have equal free energy. The free energy of the α' phase is lower than that of the y phase below T_0, and the y phase will change to the α' phase, that is, the martensitic transition occurs. But in fact, for various reasons, martensitic transformation in steel occurs at the M_S point (onset temperature of martensitic transition), which is about 250 °C cooler than T_0 instead of occurring near T_0.

In addition to Fe–C alloy, martensitic transformations can also occur in many other alloys and ceramics.

Not only in the cooling, martensitic phase transformation can also occur when heating, customarily known as reverse martensitic phase transformation.

c) Bainite phase transition

In steel, for example, when the austenite is cooled to the temperature range between the pearlite transition and the martensite transition, the iron atoms are no longer diffused as the result of lower temperature, but the carbon atoms still have a certain diffusion capacity. Thus, there is a unique nonequilibrium phase transition, where the carbon atoms diffuse while the iron atoms do not. This phase change is called bainite phase transition (or known as medium-temperature transformation). The transformation product is also a mixture of α phase and carbide, however, the carbide content and morphology of α phase and the morphology and distribution of carbides are different from that of the pearlite, which is known as the bainite.

d) Nonequilibrium dissolution precipitation

The equilibrium diagram of the alloy is shown in Fig. 9.1, if the alloy of the b component is rapidly cooling at the T_1 temperature, the β phase cannot be precipitated in the cooling process, and then the supersaturated α solid solution will be obtained when the temperature decreases to room temperature. If the solute atoms still have a certain diffusion capacity at room temperature or below the temperature of the solid solubility curve MN, the supersaturated solid solution α may decompose and gradually precipitate new phase at the above isothermal temperature. However, in the initial stage of precipitation, the composition and structure of the new phase are different from the equilibrium dissolution precipitation. This process is called nonequilibrium dissolution precipitation (or aging).

2. Classification by atomic migration

According to the migration of atoms in phase change process, the metal solid-state transition can be divided into diffusion transformation and nondiffusion transformation.

(1) Diffusion transformation

During phase transition, the phase interface moves through atomic short-range or long-range diffusion, which is called diffusion transformation, also known as noncooperative phase change. The diffusion transition occurs only when the temperature is high enough and the atomic activity is strong enough. The higher the temperature is, the stronger the atomic activity is, and the farther the diffusion distance is. Allotropic transformation, polymorphic transformation, dissolution transformation, eutectoid phase transition, spinodal decomposition, and order transformation all belong to the diffusion transformation.

The characteristics of the diffusion transformation are:

i) There is atom diffusion in the process of phase transition, and the phase transition rate is controlled by the atom diffusion velocity.
ii) The composition of the new phase and the parent phase are generally different.
iii) There is no macroscopic shape change except the volume change caused by the different specific volume between the new phase and the parent phase.

(2) Nondiffusion transformation

During the phase transition, the atoms do not undergo diffusion, and the motion of all the atoms involved in the transition is a coherent phase transition called the nondiffusion phase transition, also known as the cooperative phase transition. In the nondiffusion transformation, the atoms only migrate regularly so that the crystal lattice is reorganized. During migration, the relative moving distances of neighboring atoms do not exceed one atom spacing while the relative positions of adjacent atoms remain unchanged. Martensite phase transition and allotropic transformation of some pure metals at low temperatures are nondiffusion transition, and the atoms are unable (or difficult) to diffuse during the process.

The characteristics of the nondiffusion transformation are:

i) Due to the macroscopic shape change caused by the uniform shear, a relief phenomenon occurs on the surface of the polished sample.
ii) The chemical composition of the new phase and the parent phase are the same because the phase transition does not need atomic diffusion.
iii) There is a certain degree of crystallographic orientation relationship between the new phase and the parent phase.
iv) When some materials undergo a nondiffusion transition, the travelling speed of phase interface is very fast which is close to the sound velocity.

3. Classification by phase transformation

According to the method of phase change, metal solid-state phase change can be divided into nuclear phase change and nonnuclear phase change.

(1) Nuclear phase transition

Nuclear phase transition is carried out by nucleation and growth. The new phase nuclei can be formed uniformly in the parent phase, and can also be preferentially formed in some favorable position in the matrix. The phase transition process is completed by the constant growth of the new phase after nucleation. There are phase interfaces between the new phase and the parent phase. Most of the metal solid-state phase transitions belong to the nuclear phase transition.

(2) Nonnuclear phase transition

Nonnuclear phase transition has no nucleation stage. The nonnuclear phase transition starts at the beginning of the composition fluctuation and forms a high concentration region and a low concentration region through this process. However, there is no obvious boundary between the two groups, the composition is continuously increased from the low concentration to the high concentration. Then, the concentration difference is gradually increased by the uphill diffusion, and finally a single-phase solid solution will be decomposed into two phases associated with a coherent interface, in which the compositions are different while the lattice structures are the same. Such as the spinodal decomposition in alloy is the nonnuclear phase transition.

In summary, although there are many types of the solid-state phase transition in metals, the changes occurred in the phase transition process are essentially divided into three aspects: structure, composition, and degree of ordering. Some phase transitions have only one change, while others have two or more changes simultaneously. The same metal can undergo different phase transition under different conditions, so as to obtain different structures and properties. For example, the hardness of eutectoid steel with pearlite structure after equilibrium transformation is about HRC 23; if it is converted into martensite by cooling rapidly, the hardness can be over HRC 60. The tensile strength of Al–4%Cu alloy with equilibrium microstructure is only 150 MPa; the tensile strength can be up to 350 MPa if the alloy undergoes nonequilibrium dissolution precipitation. From here, we can see that by changing the heating and cooling condition, a certain kind of transformation can be occurred to obtain a certain kind of microstructure, which can improve the performance of the materials to a great extent.

9.1.2 The main characteristics of phase transformation in solids

Most solid-state phase transitions (except spinodal decomposition) are accomplished through nucleation and growth. Therefore, the liquid crystal theory of metal and its basic concepts still apply to solid-state phase change in principle. However, because the phase transition is carried out under the specific conditions of the solid state, the atoms of the solid crystal are regularly arranged and have many crystal defects. Therefore, the solid-state phase transition has many characteristics which are different from the liquid crystalline process of metals.

1. Phase interface

When a solid phase changes, both old and new phases are solid. Depending on the crystallographic matching between the old and the new atoms phases, the interface can be divided into coherent boundary, semicoherent boundary, and noncoherent

boundary, as shown in Fig. 9.3. The interface structure between the new phase and the old phase has a great influence on the nucleation and growth process of the solid-state phase change and the morphology of the structure after change.

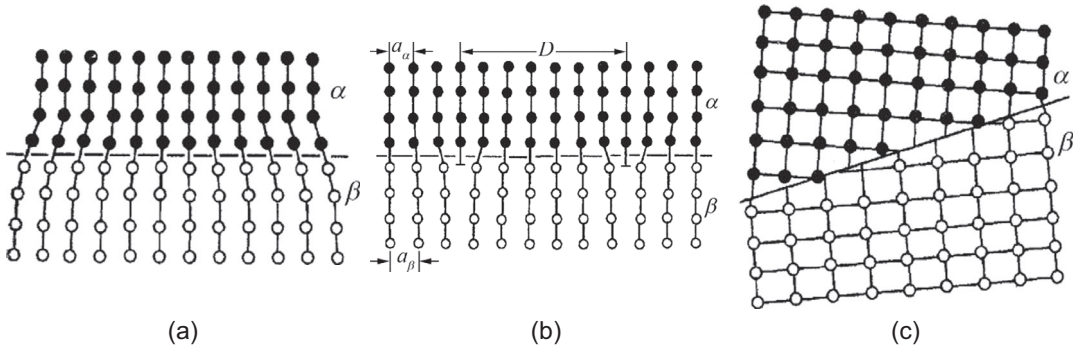


Fig. 9.3: Schematic diagram of the boundary structure in solid phase transformation: (a) coherent boundary, (b) semicoherent boundary, and (c) noncoherent boundary.

(1) Coherent boundary

If the crystal structure is the same and the lattice constant is equal, or the crystal structure and the lattice constant of the two phases are different, there is a set of specific crystallographic planes which can make the perfect matching between the two atoms. At this point, the position of the atoms in the interface happens to be the common position of the two-phase lattices, and the atoms in the interface are shared by the two lattices. This interface is called a coherent interface (Fig. 9.3(a)). The elastic strain energy and the interfacial energy are close to zero under the ideal coherent boundary (such as the twin boundary). In fact, there are always some differences between the two lattices, such as the lattice types and the lattice parameters, thus, when the two phases interface is perfectly coherent, elastic strain will be generated near the phase interface. When the coherent relationship between two phases depends on the positive strain, it is called the first type of coherent lattice while the shear strain is used to maintain it, it is called the second type of coherent lattice, and both sides of the grain boundary have a certain distortion of lattice, shown in Fig. 9.4. Figure 9.4(a) is the first type of the coherent lattice, which is compressed on the one side of the grain boundary and stretched on the other side. Figure 9.4(b) is the second type of the coherent lattice, and there is lattice bending near the grain boundary.

Generally speaking, the coherent boundary is characterized by small interfacial energy and large elastic strain energy due to the distortion near the interface. Coherent interface must rely on elastic energy to maintain. The elastic strain of the coherent interface can be increased when the new phase grows up, and when it exceeds the yield limit of the parent phase, the plastic deformation will occur; as a result, the common lattice relationship will be ruined.

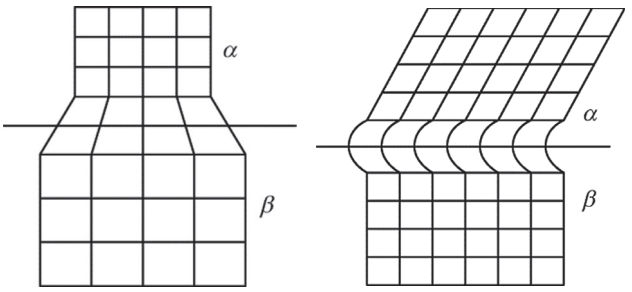


Fig. 9.4: The first type (a) and second type (b) of coherent boundary.

(2) Semicoherent boundary

The magnitude of elastic strain energy in the coherent boundary depends on the relative difference of the atomic distance between the two atoms (referred to as degree of mismatch). If a_α and a_β, respectively, represent the atomic spacing of the two phases along the crystal direction paralleling to the interface, the difference between the spacing of the two atoms in this direction is indicated by $\Delta a = |a_\alpha - a_\beta|$, thus the degree of the mismatch δ is

$$\delta = \frac{|a_\alpha - a_\beta|}{a_\alpha} = \frac{\Delta a}{a_\alpha} \tag{9.1}$$

Obviously, the greater the mismatch δ is, the greater the elastic strain energy is; when δ increases to a certain degree, it is difficult to maintain total coherent relationship. Therefore, some edge dislocations will be generated at the interface to compensate the huge difference between the atomic spacing, so that the elastic strain energy of the interface can be decreased. At this point, the two atoms in the interface become partially mismatch (Fig. 9.3(b)), so called the semicoherent interface. It can be seen that the mismatch of one-dimensional lattice can be compensated by a set of edge dislocations without producing a long-range strain field. The spacing D of this set of dislocations shall be

$$D = \frac{a_\beta}{\delta} \tag{9.2}$$

The interface is almost perfectly matched except for the core part of dislocation. The structure of the core part of dislocation is seriously distorted, and the lattice plane is discontinuous.

(3) Incoherent boundary

When the arrangement of the two atoms in the boundaries is very different, that is when the degree of mismatch δ is large, the mismatch relationship between the two phases is no longer maintained. This interface is called incoherent boundary (Fig. 9.3(c)). The structure of the incoherent boundary is similar to

that of the large-angle grain boundary, which is composed of very thin transition layers arranged by atoms irregularly.

It is generally believed that two phases can form a perfectly coherent boundary when the mismatch is less than 0.05. When the mismatch is greater than 0.25, it is easy to form a incoherent boundary. When the misfit is between 0.05 and 0.25, it is easy to form a semicoherent boundary.

When the solid-state phase transition is occurred, the interfacial energy of two phases is related to the structure and the composition of the interface. The irregular arrangement of atoms on the two phase interface will lead to the increase of the interfacial energy and the absorption of solution atoms meanwhile. The lattice strain energy can be caused by the lattice distortion due to the presence of solution atoms in the lattice, and the interface strain energy can be reduced when the solute atoms are distributed at the interface. Therefore, the solute atoms tend to segregate at the interface to reduce the total energy.

2. Orientation relationship and habit plane

In many cases, there is a certain orientation relationship between the new phase and the parent phase in solid-state phase change of metals, and the new phase is often formed on a certain lattice plane of the parent phase, which is called the habit plane represented by the crystal index of the parent phase. The existence of the habit plane means that the atomic arrangements of the new phase and the parent phase are similar in this plane, and the matching is better, which helps to reduce the interface energy. It also means that there must be a certain orientation relationship between the new phase and the parent phase. Since the orientation relationship between the two-phase crystals relative to their habit plane is determined, respectively, the orientation relationship between the two phases is determined, and the result is the existence of a parallel relationship between the grain surface and the crystal orientation in two phases. For example, when the transition from austenite to martensite occurs in the steel, the close-packed plane $\{111\}_{\gamma}$ of the austenite is parallel to the close-packed plane $\{110\}_{\alpha'}$ of the martensite; the close-packed orientation $<110>_{\gamma}$ of the austenite is parallel to the close-packed orientation $<111>_{\alpha'}$, which is known as K-S relationship, and is remembered as

$$\{111\}_{\gamma}//\{110\}_{\alpha'};\ <110>_{\gamma}//<111>_{\alpha'}$$

Generally speaking, there must be some orientation relationship between the new phase and the parent phase while the two-phase interface is coherent or semicoherent boundary; if there is no certain orientation relationship, the two-phase interface must be incoherent boundary. On the other hand, they may not have a coherent or semicoherent boundary though there is a certain relationship between the two phases. This is due to the destruction of the coherent and semicoherent boundary with the growth of the new phase.

3. Elastic strain energy

When the solid-state phase transition occurs in the metal, the volume may change due to the different specific volume between the new phase and the parent phase. The new phase cannot expand and shrink freely as a result of the constrains from the surrounding mother phase, so the elastic strain and stress must be generated between the new phase and the surrounding mother phase, which will add an additional elastic strain energy to the system. Studies have shown that the elastic strain energy due to phase change in the complete crystal is not only related to the different specific volume and the modulus of elasticity between the new phase and the mother phase, but also to the shape of the new phase. Assuming that the new phases of the different shapes are treated as rotational ellipsoids, the ratio c/a reflects the specific shape of the rotational ellipsoid, a represents the equatorial diameter of the rotational ellipsoid while c represents the distance between the poles of the rotating axis. When $c < a$ it represents a disk; when $c = a$ it is a sphere; and when $c > a$ it is a bar (needle). Figure 9.5 shows the influence of geometry of the new phase on the strain energy generated due to the different specific volume (relative value), which can be seen that the strain energy is the maximum when the new phase is spherical, and that is the smallest in the disk shape while that is centered in the bar (needle) shape.

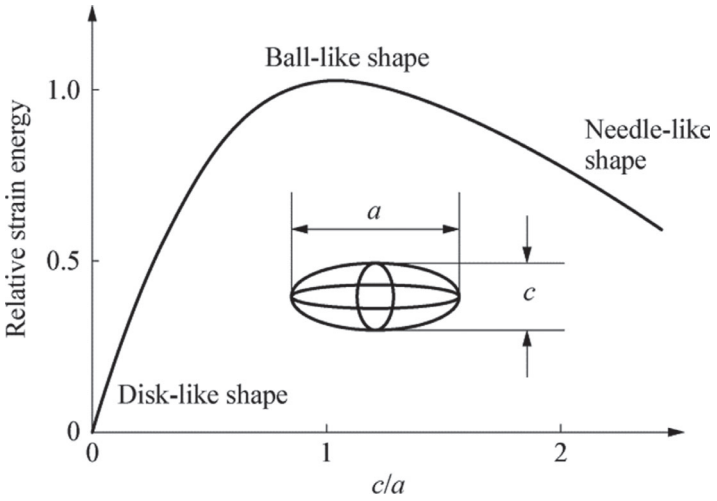


Fig. 9.5: The relationship between the shape of new phase and relative strain energy.

The elastic stain energy is produced by the mismatch between the two-phase interface, in addition to the different specific volume. The elastic strain energy due to mismatch is the largest in the coherent boundary, the semicoherent boundary is next (the elastic strain energy decreases due to the generation of interfacial dislocations), and the incoherent boundary is zero.

From the above, the phase transition resistance includes both the interfacial energy and the elastic strain energy. The interface types between the new phase and the mother phase have different effects on the interfacial energy and the elastic strain energy. The interfacial energy can be reduced when the interface is coherent, but the elastic stain energy can be increased. When the interface is incoherent, the elastic strain energy of disk shape is the lowest while the interfacial energy is higher. However, the spherical shape has the lowest interfacial energy, but the highest elastic strain energy. Whether the interfacial energy or the elastic strain energy plays the leading role during the phase transition depends on the specific conditions. If the undercooling is very deep, the radius of critical nucleus is small and the unit volume of new phase interface is very big, the leading role is played by the interfacial energy because the huge of energy increases nucleation energy to become the main resistance of the phase transition. The interface is easy to be coherent to reduce the interfacial energy, and the reduction of the interfacial energy exceeds the increase of the elastic strain energy caused by the coherent boundary, which can reduce the total nucleation energy and easy to nucleation. In the case of very low degree of undercooling, the radius of critical nucleus is big, the incoherent interface is easy to form because the interfacial energy is no longer dominant. At this situation, the elastic strain energy plays the leading role on account of huge difference between the specific volume of two phases, and the new phase in shape of disk will form to reduce the elastic strain energy; if the volume difference between two phases is small and the elastic strain energy has little effect, a spherical phase is formed to decrease the interfacial energy.

4. The formation of the transition phase

According to the phase transition thermodynamics, the phase transition is caused by the negative free energy between the new phase and the mother phase, and it is also aimed to change from the unstable mother phase with high energy to the stable new phase with the lowest energy. However, when crystal structures of the stable new phase and the parent phase are different, only a incoherent boundary of high energy can be formed between the two phases. The radius of critical nucleus of the new phase is small, and the unit volume of the new phase interface is very big. Furthermore, the interfacial energy which can prevent nucleation and the nucleation energy of the incoherent interface are both large, which helps the phase transition difficult to happen. In this case, instead of being directly transformed into the stable phase with the lowest free energy from the parent phase, the metastable transition phase is formed firstly, whose crystal structure or composition is close to the mother phase while the free energy is lower than the mother phase. In order to reduce the nucleation energy and make it easy to happen, the transition phase generally has a coherent or semicoherent interface with lower interfacial energy.

Although the transition phase can exist under certain condition, it has the tendency to change continuously to the equilibrium phase because the free energy of the transition phase is still higher than that of the equilibrium phase, and this tendency will increase with the increase of temperature. If the transition phase obtained by proper heat treatment is used in the room temperature, the tendency of changing into equilibrium phase is often slow to be negligible.

5. The influence of crystal defects

There are many crystal defects such as grain boundary, sub-grain boundary, vacancy, and dislocation in solid crystals, thus the lattice distortion occurs in the surrounding which is stored the distortion energy. Generally speaking, the nucleation of the new phase in solid-state phase transition is always preferentially formed at the crystal defects. This is because crystal defects are the largest region of energy fluctuations, structure fluctuations, and component fluctuations. When the nuclei are formed in these areas, the activation energy of atom diffusion is low, the diffusion speed is fast, and the phase transition stress is easily relaxed. For example, the grain boundaries which has a lower nucleation energy is the origin of the nonspontaneous nucleation, thus it catalyzes the phase transition.

Dislocation also has a significant catalytic effect on phase change. It is generally believed that the dislocation line will disappear when nucleation is on that dislocation, then the distortion energy in the dislocation center can be released and the free energy of the system will be decreased. The energy released can be used to overcome the formation of the new phase interface and the energy of the transition strain, thus accelerating the phase change. There is another way to catalyze the phase transition. The dislocation adheres to the new phase interface and forms a part of the dislocations in the semicoherent boundary instead of disappearing. As a result, the free energy of the system will also be reduced. In conclusion, from the energy point of view, the nucleation energy of homogeneous nucleation is the largest, the vacancy nucleation is the second, the dislocation nucleation is the third, and the grain boundary inhomogeneous nucleation is the smallest.

6. The diffusion of atoms

In many cases, the solid-state phase transition must be carried out through the diffusion of some components because of the different composition between the new phase and the mother phase, then the diffusion becomes the governing factor in the phase transition. Thus, the diffusion rate of the atom has a significant influence on the solid-state phase transition because the diffusion rate of the atom in solid metal is much lower than that in liquid metal. A solid-state phase transition controlled by atomic diffusion can produce a large degree of undercooling. The phase transition drive force increases with the increase of undercooling, and velocity

of the phase change also increases. However, when the degree of undercooling increases to a certain degree, the phase change decreases with the increase of the degree of undercooling because the capacity of the atomic diffusion is declining. If the degree of undercooling is further increased, the phase transition in diffusion type can be inhibited, and the nondiffusion phase transition will occur to form a metastable phase at a low temperature. For example, when steel is rapidly cooling from austenite, the diffusion phase transition can be suppressed, while in the low temperature, the nondiffused martensite phase change occurs in the sheer mode to generate the metastable martensite.

9.2 Thermodynamics of solid phase transformation

9.2.1 Thermodynamic condition of solid phase transformation

1. Driving forces of phase transformation

The thermodynamics of phase transition points out that the stability of the system state is determined by the free energy, and the system is the most stable when the free energy reaches the lowest. All systems have a spontaneous tendency to reduce the free energy to achieve a steady state. If the system has the conditions for lowering the system free energy, it will spontaneously change from a high-energy state to a low-energy state. This change is called spontaneous transformation. The solid phase transformation is also spontaneous transition. So, only when the free energy of new phase is lower than the old phase, the phase transformation can occur. The free energy difference of the new and the old phases and the lower free energy of the new phase are the driving forces for the spontaneous transformation of the old phase into the new phase. The above is the thermodynamic condition of phase change. Thus, for a solid phase transformation to occur, it is necessary to create a negative free energy difference between the new and the old phases. Otherwise, the phase change is impossible.

Free energy G is a feature parameter of the system. Set H as enthalpy, S as entropy, and T is absolute temperature, and

$$G = H - TS \tag{9.3}$$

The free energy of any phase is a function of temperature. By changing the temperature, it is possible to obtain thermodynamic conditions of phase transformation. To investigate the characteristics of this relation, the first and second derivative of G with respect to T can be done:

$$dG = dH - TdS - SdT \tag{9.4}$$

For reversible processes, the general equations of the first and second laws of thermodynamics can be written as,

$$TdS = dH + dW$$

In solid materials, there are only small volume changes caused by phase change, so these volume changes can be neglected. Suppose that W is expansion work, so in the constant volume process, the volume (V) is constant, $dW = 0$, and $TdS = dH$. Put it into eq. (9.4), we obtain $dG = -SdT$, and the first derivative is

$$\left(\frac{\partial G}{\partial T}\right)_V = -S \tag{9.5}$$

Since S is always positive, $(\partial G/\partial T)_V$ should always be negative, that is, G always decreases as T increases. And the second derivative is

$$\left(\frac{\partial^2 G}{\partial T^2}\right)_V = -\left(\frac{\partial S}{\partial T}\right)_V \tag{9.6}$$

Also, since entropy S always increases with T increasing, so $(\partial S/\partial T)_V$ is positive and $(\partial^2 G/\partial T^2)_V$ is negative. This means that the characteristic curve of free energy G and temperature T is always concave downward. Figure 9.6 shows the change of the free energy G with temperature T in the α and γ phases in a material. Both of the free energy decrease with the increase in temperature, but because of the different of the entropy value between two phases and the degree of entropy varying with the temperature, the free energy curve of the two phases may intersect at one point T_0. *At* T_0, $G_\alpha = G_\gamma$, the two phases are in equilibrium and T_0 is called the theoretical transition temperature. Since the system tends to minimize the free energy, when the temperature is lower than T_0, G_α is lower than G_γ, and phase γ should change into phase α;

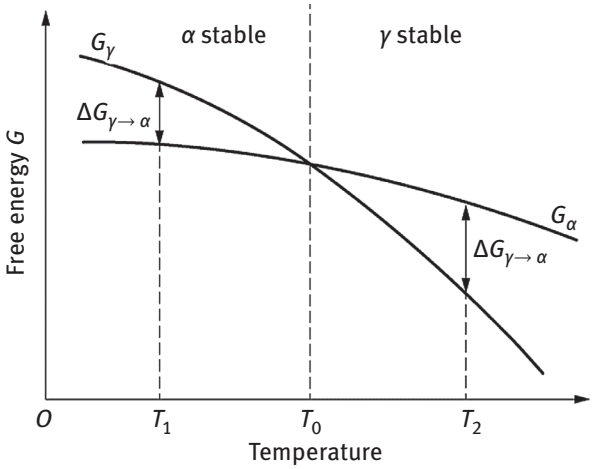


Fig. 9.6: The relationship between free energy of each phase and temperature.

conversely, when temperature is higher than T_0, the phase α changes to phase γ. However, this change does not happen in T_0. Only by cooling or heating, get the necessary undercooling ($\Delta T = T_0 - T_1$) or superheating ($\Delta T = T_2 - T_0$) to obtain the sufficient free energy difference for phase change ($\Delta G_{\gamma\to\alpha}$ or $\Delta G_{\alpha\to\gamma}$). In other word, the phase change $\alpha \to \gamma$ or $\gamma \to \alpha$ can occur only when the energy condition of the thermodynamics of phase transformation is satisfied. Obviously, as the undercooling or superheat increases, the driving force of phase transformation also increases, which is beneficial to the phase transformation. And the phase transformation is always in the direction of decreasing the free energy.

2. Phase transformation barrier

To make the system change from the old to the new phase, in addition to the driving force of phase transformation, the phase transformation barrier also need to be overcome. The phase transformation barrier refers to the interatomic attraction of reorganization the crystal lattice overcome during phase transformation. In Fig. 9.7, the state I represents the unstable old phase γ, and the free energy is higher; the state II represents the stable new phase α, the free energy is lower. According to thermodynamic conditions, the free energy of phase α is lower than that of phase γ. So, there is a free energy difference $\Delta G_{\gamma\to\alpha} = G_\alpha - G_\gamma$, and $\Delta G_{\gamma\to\alpha} < 0$. Phase γ has a spontaneous tendency to change to phase α. But to make the phase change possible, not only the free energy difference $\Delta G_{\gamma\to\alpha}$ is needed, also need the additional energy Δg to overcome the phase transformation barrier caused by interatomic attraction.

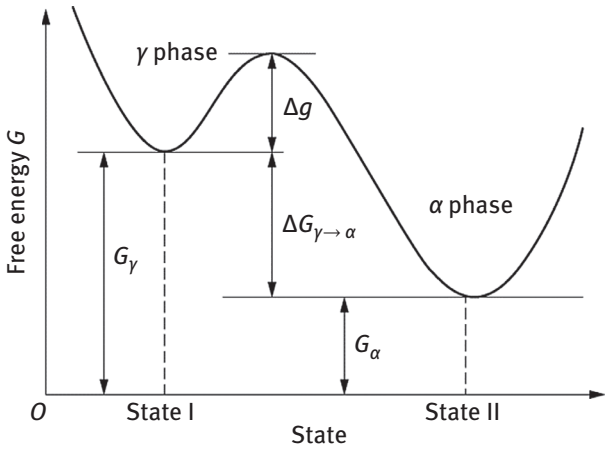


Fig. 9.7: Schematic diagram of the phase transformation barrier.

In crystals, atoms can acquire this additional energy in two ways. One is the heterogeneity of atomic thermal vibration. It makes some atoms to have very high thermal vibrational energy to overcome the interatomic attraction and leave the equilibrium position to obtain additional energy. The two is mechanical stress. For example,

elastic deformation or plastic deformation can destroy the regularity of the arrangement of atoms in crystals, cause an internal stress in crystal and force some atoms out of equilibrium position to get the additional energy.

The barrier can be approximately expressed by activation energy Q. The activation energy is the energy required to move the atoms in crystal away from the equilibrium positions to another new equilibrium or nonequilibrium position. Obviously, the greater the activation energy, the higher the phase transformation barrier. The activation energy is related to temperature. The higher the temperature is, the smaller the activation energy is. This is because the increase of the distance between atoms, the decrease of attraction. Therefore, the higher the temperature, the easier the phase transformation. However, in more cases, the barrier is represented by the self-diffusion coefficient D of the crystal atom, and the self-diffusion coefficient D decreases exponentially with the decrease of temperature.

$$D = D_0 \exp\left(-\frac{Q}{RT}\right) \tag{9.7}$$

In the formula, D_0 is the coefficient (frequency factor), R is the gas constant, T is the absolute temperature, and Q is the activation energy. Thus, the greater the self-diffusion coefficient, the stronger the ability to overcome the barrier, and the easier the phase transformation.

9.2.2 Nucleation of solid phase transformation

Most of the solid phase transformation is done through the process of nucleation and growth. The nucleation process is often firstly forming the essential component and structure of the new phase in some small areas of the parent phase, called the nuclear embryo. If the size of a nuclear embryo exceeds a critical value, it will be stable and grow spontaneously then become crystal nucleus of new phase. If the crystal nucleus is uniformly distributed not preferentially in the parent phase, it is called homogeneous nucleation; if the crystal nucleus is preferentially distributed unevenly in some regions of the parent phase, it is called heterogeneous nucleation.

Researches point out that the solid phase transformation is similar to the crystallization process of liquid metal, which rarely occurs homogeneous nucleation. And the new phase mainly heterogeneously nucleates at the grain boundaries of the parent phase, stacking faults, dislocations, and other crystal defects. To facilitate the analysis, first discuss the condition of homogeneous nucleation.

1. Homogeneous nucleation

Compared with the crystallization process of liquid metal, the driving force of homogeneous nucleation in solid phase transformation is also the difference of the

free energy between the old and new phases. But in addition to the interfacial energy, the resistance to nucleation also has the elastic strain energy. The interfacial energy of the crystal nucleus is proportional to the surface area of the crystal nucleus, and the elastic strain energy is proportional to the quality of crystal nucleus. According to the classical nucleation theory, during the homogeneous nucleation in solid phase transformation, the total change of the free energy ΔG of the system is equal to:

$$\Delta G = -V \cdot \Delta G_V + S\sigma + V\varepsilon \tag{9.8}$$

In the formula, V is the volume of new phase; ΔG_V is the free energy difference per unit volume between the new phase and the parent phase; S is surface area of new phase; σ is the interfacial energy per unit area between the new phase and the parent phase; ε is elastic strain energy per unit volume of new phase. In formula (9.8), $V \cdot \Delta G_V$ is volume free energy difference, that is, the driving force of phase change. And $S\sigma$ is the interfacial energy, $V\varepsilon$ is elastic strain energy, both of which are phase transformation resistance. It is easily seen that only when $V \cdot \Delta G_V > S\sigma + V\varepsilon$, the formula can be negative, $\Delta G < 0$, and the new phase nucleation is possible. It can only be achieved at a certain undercooling when a new phase nucleus larger than the critical size is formed in the high-energy micro region.

If the crystal nucleus of the new phase is spherical, formula (9.8) can be written as

$$\Delta G = -\frac{4}{3}\pi r^3 \Delta G_V + 4\pi r^2 \sigma + \frac{4}{3}\pi r^3 \varepsilon \tag{9.9}$$

For $d\Delta G/dr = 0$, the critical nucleation radius of the new phase r_c is

$$r_c = \frac{2\sigma}{\Delta G_V - \varepsilon} \tag{9.10}$$

The nucleation energy for the critical nucleation is

$$W = \Delta G_{\max} = \frac{16\pi\sigma^3}{3(\Delta G_V - \varepsilon)^2} \tag{9.11}$$

From formulas (9.10) and (9.11), it is found that the critical nucleation radius r_c increases and the nucleation energy W increases as the surface energy σ and elastic strain energy ε increase. Therefore, the coherent new phase nucleus having a low interfacial energy and high elastic strain energy tends to form disk or flake shape; the incoherent new phase nucleus having a high interfacial energy and low elastic strain energy is easy to form equiaxed shape. But, if the interfacial energy of new phase nucleus is highly anisotropic, the latter may be lamellar or acicular.

The critical nucleation radius and nucleation energy are both functions of the free energy difference, so they also vary with undercooling (superheat). With the

increase of undercooling (superheat), the critical nucleation radius and nucleation energy decrease, the probability of new phase nucleation increases, and the number of new phase crystal nucleus increases, that is, phase transformation is easy to occur. The same as the additional energy needed to overcome the phase transition barrier, the energy required for nucleation energy comes from two aspects: one is the energy fluctuations in the parent phase; the other is the internal stress caused by deformation and other factors.

Similar to the crystallization of liquid metals, the nucleation rate I of homogeneous nucleation in solid phase transformation can be expressed by the following formula:

$$I = nv \exp\left(-\frac{Q+W}{kT}\right) \tag{9.12}$$

In the formula, n is the number of atoms in a unit volume parent phase, v is atomic vibration frequency, Q is atomic diffusion activation energy, k is the Boltzmann constant, and T is temperature of phase transformation. The diffusion activation energy of solid metal atoms is relatively large, and the nucleation rate of metal solid phase transition is much lower. At the same time, a large number of crystal defects in solid materials can provide energy to promote nucleation. Therefore, heterogeneous nucleation becomes the main nucleation mode of the solid phase transformation.

2. Heterogeneous nucleation

All kinds of crystal defects existing in the parent phase can be used as nucleation sites, and the energy stored in the crystal defects can reduce nucleation energy and make nucleation easy. When the new phase nucleus is formed at the crystal defects of the parent phase, the total change of the system free energy is

$$\Delta G = -V \cdot \Delta G_V + S\sigma + V\varepsilon - \Delta G_d \tag{9.13}$$

Compared with formula (9.8), the last item is added, which is the energy caused by the disappearance or reduction of crystal defects. Therefore, the crystal defects can promote nucleation. The effects of crystal defects in nucleation are explained below.

(1) Nucleation at grain boundary

In the polycrystal, the boundary between two adjacent grains is called the interface; the common junction line of the three grains is called the boundary edge; the common point of the four grains is called the boundary corner. Interfaces, boundary edge and boundary corner are not plane, line and point in geometric sense, and they all occupy a certain volume. If δ represents the boundary thickness, and L represents the mean diameter of the grain. It can be approximately estimated that the volume fraction of the interface, boundary edge, and boundary corner in the

polycrystal is (δ/L), $(\delta/L)^2$, and $(\delta/L)^3$, respectively. The interface, boundary edges, and boundary corner, all can provide the stored distortion energy to promote nucleation. At the interface nucleation, there is only one interface for the crystal nucleus to swallow, at the boundary edge nucleation, there are three interfaces, and the boundary corner is six. So, from the energy point of view, the boundary corner provides the largest energy, the boundary edge is the second, and the interface is minimum. However, from the volume fraction of the three nucleation position, the interface occupies the most and the boundary corner is the smallest. Considering the two factors comprehensively, in different interface position the heterogeneous nucleation rate N can be expressed as

$$N = n\nu\left(\frac{\delta}{L}\right)^{3-i} \exp\left(-\frac{Q}{kT}\right) \exp\left(-\frac{A_i W}{kT}\right) \tag{9.14}$$

In the formula, $i = 0, 1, 2, 3$, respectively, represent the boundary corner nucleation, boundary edge nucleation, interface nucleation and homogeneous nucleation. A_i is the ratio of the nucleation energy of nucleation at different positions in grain boundaries to the nucleation energy of homogeneous nucleation. $A_0 < A_1 < A_2 < 1$, $A_3 = 1$.

In order to reduce the surface area of nucleation and the interfacial energy, the interface is always spherical in incoherent nucleation. The shape of incoherent crystal nucleus of the interface, boundary edge and boundary corner are biconvex lens, curved triangular prism with both ends of the tip and spherical tetrahedron, respectively, as shown in Fig. 9.8. While the coherent and semicoherent interfaces are usually planes. It has been said before that there is a certain the crystallographic relationships between the new phase and the parent phase on the interfaces. When gain boundary nucleation at large angles, it cannot have the crystallographic relationships with grains on both sides of the grain boundary. Therefore, the new phase crystal nucleus can only be coherent or semicoherent with parent phase crystal of one side, and the other side is incoherent. It results in the change in the shape of crystal nucleus, one side is a spherical cap shape, and the other side is a plane, as shown in Fig. 9.9.

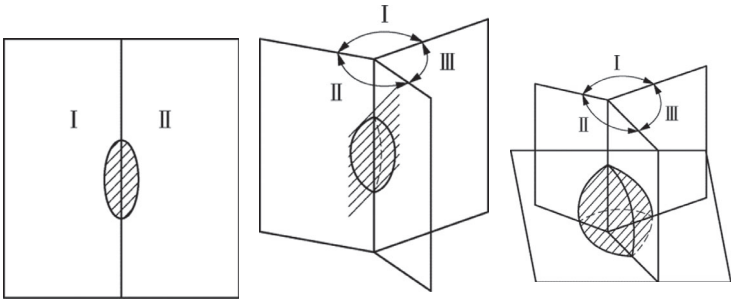


Fig. 9.8: The shape of the incoherent nucleus on boundaries: (a) Interface nucleation, (b) edge nucleation, and (c) corner nucleation.

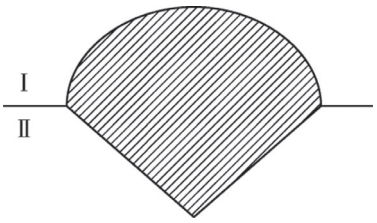


Fig. 9.9: Interface nucleus with coherent boundary on one side.

When α is the parent phase and β is the new phase, in grain boundary nucleation the total change of the free energy of the system is expressed as

$$\Delta G = -V \cdot \Delta G_V + S_{\alpha\beta}\sigma_{\alpha\beta} + V\varepsilon - S_{\alpha\alpha}\sigma_{\alpha\alpha} \tag{9.15}$$

In formula (9.19), $S_{\alpha\beta}$ stands for the surface area of phase β; $\sigma_{\alpha\beta}$ is the interfacial energy per unit area of phase α and β; $S_{\alpha\alpha}$ is α phase grain boundary area swallowed by β; $\sigma_{\alpha\alpha}$ is the interfacial energy per unit area of phase α. Rewrite formula (9.15):

$$\Delta G = -V \cdot \Delta G_V + V\varepsilon + S_{\alpha\beta}\sigma_{\alpha\beta}\left(1 - \frac{S_{\alpha\alpha}}{S_{\alpha\beta}} \cdot \frac{\sigma_{\alpha\alpha}}{\sigma_{\alpha\beta}}\right) \tag{9.16}$$

Set $\chi = \frac{\sigma_{\alpha\alpha}}{\sigma_{\alpha\beta}}$, and the nucleation energy of grain boundary nucleation can be derived:

$$W = \Delta G_{\max} = \frac{16}{3} \cdot \frac{\pi\sigma_{\alpha\beta}{}^3\left(1 - \frac{S_{\alpha\alpha}}{S_{\alpha\beta}} \cdot \chi\right)^3}{(\Delta G_V - \varepsilon)^2} \tag{9.17}$$

For the interfacial nucleation, from the balance of the interfacial tension (Fig. 9.10a), the relationship between the interfacial energy is shown as follows:

$$2\sigma_{\alpha\beta}\cos\theta = \sigma_{\alpha\alpha}$$

$$\chi = \frac{\sigma_{\alpha\alpha}}{\sigma_{\alpha\beta}} = 2\cos\theta \tag{9.18}$$

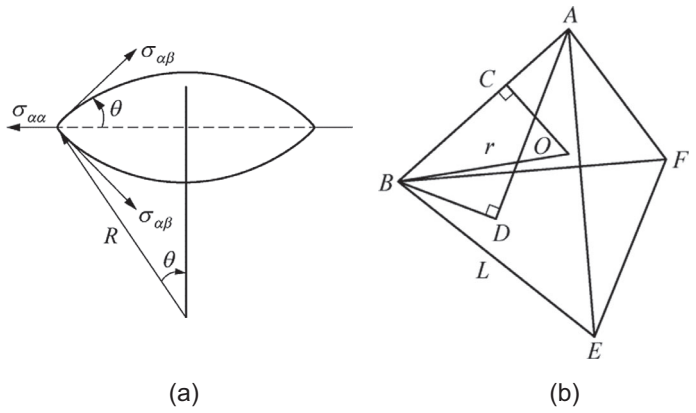
If the crystal nucleus is double spherical cap shape, R is the radius of curvature

$$S_{\alpha\alpha} = \pi R^2 \sin^2\theta = \pi R^2(1 - \cos^2\theta)$$

$$S_{\alpha\beta} = 4\pi R^2(1 - \cos\theta)$$

$$\frac{S_{\alpha\alpha}}{S_{\alpha\beta}} = \frac{1}{4}(1 + \cos\theta) = \frac{1}{4}\left(1 + \frac{1}{2}\chi\right) \tag{9.19}$$

Depending on the formula (9.17), when $1 - \frac{S_{\alpha\alpha}}{S_{\alpha\beta}} = 01 - \frac{S_{\alpha\alpha}}{S_{\alpha\beta}} = 0$, $W = 0$, formula (9.19) can be rewritten as



(a) (b)

Fig. 9.10: The surface area and the absorbed boundary area of the interface and corner nucleus: (a) interface nucleation and (b) corner nucleation.

$$\frac{1}{2}\chi^2 + \chi - 4 = 0 \tag{9.20}$$

The solution of the quadratic equation $\chi = 2,\ \chi = -4$. Thus, in the interface nucleation, as long as $\chi = \frac{\sigma_{\alpha\alpha}}{\sigma_{\alpha\beta}} \geq 2$, nucleation will no longer require additional energy.

In the boundary corner nucleation, the crystal nucleus can be approximately regarded as a tetrahedron (Fig. 9.10b) for the convenience. Set the length of a tetrahedron as L, the distance from the center of the tetrahedron O to the vertex is r. In the Fig. (9.10b), the extension line of the OA intersects the BEF plane at point D. And the point D should be the center of the triangle BEF, $BD \perp AD$. Make the vertical line OC of line AB, and as ΔAOC–ΔABD, so,

$$\frac{OC}{BD} = \frac{AC}{AD}$$

and

$$\because AC = \frac{1}{2}L, BD = \frac{2}{3}\cdot\frac{\sqrt{3}}{2}L = \frac{1}{\sqrt{3}L}, AD = \sqrt{L^2 - \frac{1}{3}L^2} = \sqrt{\frac{2}{3}}L$$

$$\therefore OC = \frac{AC \cdot BD}{AD} = \frac{1}{2\sqrt{2}}L$$

The triangle OAB is one of the six swallowed interfaces, the area of which is $S_1 = \frac{1}{2}L \cdot OC = \frac{1}{4\sqrt{2}}L^2$. So the total area swallowed is $S_{\alpha\alpha} = 6S_1 = \frac{3}{2\sqrt{2}}L^2$. The surface area of tetrahedral crystal nucleus is $S_{\alpha\beta} = 4 \cdot \frac{1}{2} \cdot \frac{\sqrt{3}}{2}L^2 = \sqrt{3}L^2$. Therefore,

$$\frac{S_{\alpha\alpha}}{S_{\alpha\beta}} = \frac{\sqrt{3}}{2\sqrt{2}} \tag{9.21}$$

Put this into the formula (9.17), when $1 - \frac{S_{\alpha\alpha}}{S_{\alpha\beta}} = 0$ and $W = 0$, $1 - \frac{\sqrt{3}}{2\sqrt{2}}\chi = 0$, So

$$\chi = \frac{2\sqrt{2}}{\sqrt{3}} \tag{9.22}$$

That is, when $\chi = \frac{\sigma_{\alpha\alpha}}{\sigma_{\alpha\beta}} \geq \frac{2\sqrt{2}}{\sqrt{3}}$, the boundary corner nucleation does not have energy barrier.

For boundary edge nucleation, calculation results show that when $\chi = \frac{\sigma_{\alpha\alpha}}{\sigma_{\alpha\beta}} \geq \sqrt{3}$, there is no energy barrier.

The above analysis indicates that boundary corner nucleation has the smallest energy barrier. However, whether the boundary corner can become the preferred nucleation site, it depends on undercooling and values of $\sigma_{\alpha\alpha}/\sigma_{\alpha\beta}$. When undercooling is rather large, the nucleation driving force increases, and the nucleation energy decreases. There is little energy barrier in either position. At this time, the interface nucleation with a larger volume fraction has the greater contribution to the nucleation. And when $\chi = \frac{\sigma_{\alpha\alpha}}{\sigma_{\alpha\beta}} \geq 2$, these positions all have no energy barrier. Of course, the interface nucleation contributes most.

(2) Dislocation nucleation

There are three forms of dislocation nucleation.

The first form: the new phase nucleates at dislocation line and this dislocation line disappears. The released distortion energy reduces the nucleation energy and promotes nucleation. If the new phase nucleus around the dislocation is considered approximately as a cylinder with a radius of r, the distortion energy of the unit length due to the disappearance of the dislocation line is

$$A \ln \frac{r}{r_0} = A(\ln r - \ln r_0) = A \ln r \tag{9.23}$$

For edge dislocation: $A = Gb^2/4\pi(1-\upsilon)$; for screw dislocation: $A = Gb^2/4\pi$. Here, r_0 is the radius of the central hole of the dislocation supposed, G is the shear modulus, B is Burgers vector, and υ is Poisson's ratio. It is obvious that the distortion energy of dislocations is related to Burgers vector B. The larger the value of Burgers vector, the greater the role of dislocations in nucleation. At this time, the change of the free energy of a unit length crystal nucleus should be

$$\Delta G = -A \ln r - \pi r^2(\Delta G_V - \varepsilon) + 2\pi r\sigma \tag{9.24}$$

And the critical radius of crystal nucleus r_c can be deduced from eq. (9.24):

$$r_c = \frac{2\pi\sigma \pm \sqrt{4\pi^2\sigma^2 - 8\pi A(\Delta G_V - \varepsilon)}}{4\pi(\Delta G_V - \varepsilon)} \tag{9.25}$$

When ΔG_V and A is too large, $4\pi^2\sigma^2 < 8\pi A(\Delta G_V - \varepsilon)$, the r_c has no real root. In this situation, dislocation nucleation has no energy barrier.

The second form: the dislocation line does not disappear and is attached to the new phase interface. Then it becomes the dislocation part in the semicoherent interface, compensating the mismatch and reducing the interfacial energy. So, the nuclear energy of the new phase decreases.

The third form: when the composition of the new phase is different from that of the matrix, the solute atoms segregate on dislocation line (form atmosphere). It is favorable for the formation of precipitate nucleation, and therefore promotes the phase transition.

It is estimated that when the driving force of phase transformation is very small, and the interfacial energy between the new phase and the parent phase is about $2\times10^{-5}\,\mathrm{J/cm^2}$, the nucleation rate of homogeneous nucleation is only $10^{-70}/(\mathrm{cm^3\cdot s})$. If the density of dislocation is $10^8/\mathrm{cm}$, the nucleation rate of heterogeneous nucleation caused by dislocation can reach $10^8/(\mathrm{cm^3\cdot s})$. Therefore, when there is a higher density of dislocation in the crystal, it is difficult for the solid phase transformation to be homogeneous nucleation.

(3) Vacancy nucleation

Vacancies promote nucleation by influencing diffusion or by providing their energy to the driving force of phase transformation. For example, in the case of desolvation decomposition of supersaturated solid solution, when the solid solution cools rapidly from the high temperature, the supersaturated solute atoms are retained in the solid solution. At the same time, a large number of supersaturated vacancies are also retained. On the one hand, they promote the diffusion of solute atoms. On the other hand, they also promotes nucleation of heterogeneous nucleation as the nucleation site of precipitation phase, and distribute the precipitate in the whole matrix. But near the grain boundary there is always the precipitation free zone, in which no precipitate is found. This is because the supersaturated vacancies near the grain boundary diffuse to the grain boundary and disappear, so there is no inhomogeneous nucleation here. More vacancies remain away from the grain boundaries, and precipitates are easy to nucleate and grow here.

9.2.3 Nucleus growth of solid phase transformation

1. Growth mechanism of new phase

The growth of the new phase crystal nucleus is essentially the migration of the interface towards the parent phase. Different types of the solid phase transformation have different growth mechanism. For eutectoid phase transformation and dissolution transformation and so on, due to the different composition between new phase and the parent phase, the growth of the new phase crystal nucleus must depend on the long-range diffusion of solute atoms in the parent phase. Until the composition

near the interface meets the new requirements, the new phase crystal nucleus can grow up. When this type of phase transformation occurs, there must be a mass transfer process. On the contrary, for allotropic transformation and martensitic transformation and so on, the composition of the new phase and the parent phase are the same, and crystal growth does not need the mass transfer process. In the growth of the new phase crystal nucleus, the atoms near the interface only do the short-range diffusion or even do not diffuse at all.

If the new phase nucleus and parent phase have a certain crystallographic relationships, this relationships will remain during growing. The growth mechanism of new phase is also related to the interface structure of the crystal nucleus (like coherent, semicoherent, or incoherent interface). In fact, the situation that the new phase nucleus is completely matched with the parent phase and forms the perfectly coherent interfaces is very rare. Usually, it forms the semicoherent or incoherent interfaces. These two interfaces have different migration mechanism.

(1) Migration of the semicoherent interface

Because the semicoherent interface has a lower interfacial energy, the interface tends to remain plane during growing process. Taking the martensitic transformation as an example, the growth of crystal nucleus is accomplished by the shear of atoms in one side of the parent phase on the semicoherent interface. Its characteristic is that a large number of atoms move regularly in one direction with the distance less than one interatomic spacing, and maintain the original crystallographic relationships, as shown in Fig. 9.11. This type of growth of crystal nucleus is called synergistic growth or displacement growth. Because the migration of atoms during phase transformation is less than one interatomic spacing, it is also called diffusionless phase transformation. This synergistic growth using a homogeneous shear manner results in tilting on the surface of the specimen during polishing, as Fig. 9.12.

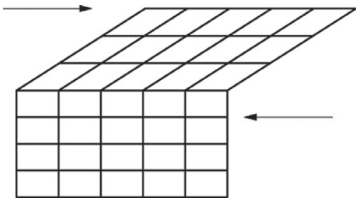


Fig. 9.11: Displacement growth caused by shear deformation.

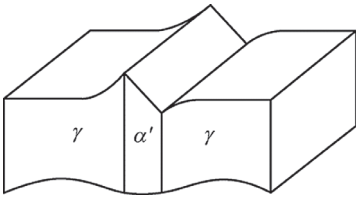


Fig. 9.12: Schematic diagram of surface tilting caused by martensite transformation.

In addition to the above shear mechanism, it can also use the motion of interfacial dislocation on the semicoherent interface, make the interfaces migrate along the normal direction, thus promote the growth of new phase crystal nucleus. The possible structure of the semicoherent interface including interfacial dislocations is shown in Fig. 9.13. Figure 9.13(a) is a flat interface. The interface dislocations are on the same plane, and the Burgers vectors of edge dislocations are parallel to the interface. Now, if the interface migrates along the normal direction, the interface dislocation must climb to move with the interface. It is difficult to achieve without the external force or enough high temperature, so it hinders the migration of the interface and interferes the growth of crystal nucleus. However, if as shown in Fig. 9.13(b), the interface dislocations are distributed on the stepped interface, the Burgers vector of edge dislocation and interface are at a certain angle. Thus, the slip motion of the dislocation can make the steps migrate laterally across the interface, causing the interface to move toward the normal direction and promoting the growth of new phase as shown in Fig. 9.14. This type of growth of crystal nucleus is called stepped growth.

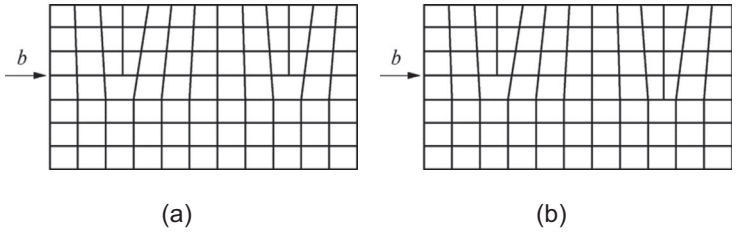


Fig. 9.13: The possible structure of the semicoherent interface: (a) straight interface and (b) step interface.

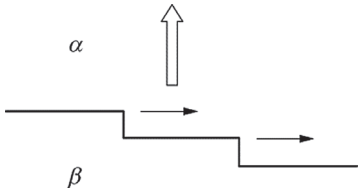


Fig. 9.14: Schematic diagram of the nucleus growth by step mode.

(2) Migration of incoherent interface

In many cases, the interface between the new phase crystal nucleus and the parent phase is incoherent interface. The arrangement of atoms at the interface is disordered, forming a thin transition layer with irregular arrangement, and the possible structure is shown in Fig. 9.15(a). The movement of the atoms on the interface is not cooperative, without a certain order, and the relative displacement of the distance is different and the adjacent relationships may also change. This interface can accept or output atoms in any position. As the parent atoms move toward the new phase, the interface itself moves along its normal direction, thus make the new phase grow up gradually. However, some people think that the micro region in the

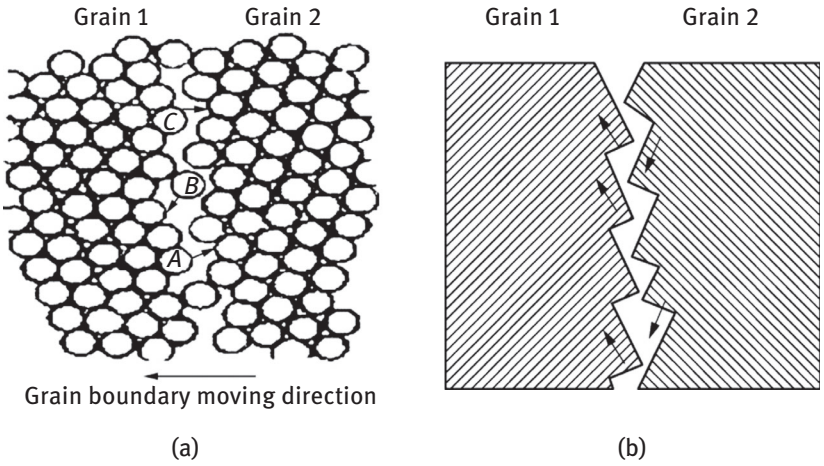


Fig. 9.15: The possible structure of the incoherent interface: (a) transition thin layer of irregular arrangement of atoms and (b) step shape incoherent interface.

incoherent interface may also have a step-like structure, as shown in Fig. 9.15(b). The plane of this step is the most close-packed plane of atoms, and the height of step is approximately equal to an atomic layer. The atoms transfer from the end of the parent phase step to the step of new phase, so that the step of new phase occurs the lateral movement. Therefore, the interface develops toward the normal direction and the new phase grow up. Because the migration of this incoherent interface is through interfacial diffusion, this phase transformation is also called diffusion-type phase transformation.

It should be noted that the solid phase transformation does not necessarily belong to either a pure diffusion type or a diffusion free type. For example, the bainitic transformation in steel has both the characteristics of diffusion-type transformation and diffusion free phase transformation. In other words, it is not only consistent with the migration mechanism of the semicoherent interface, but also has the diffusion behavior of solute atoms.

2. Nuclei growth speed

New phase growth speed is determined by the migration speed of phase interface. For the interface migration realized by lattice shear mechanism, atomic diffusion is not needed and the growth activation energy is zero, thus the growth speed is usually very high. If the interface migration relies on atomic diffusion, the new phase growth speed is relatively lower since diffusion needs time. Diffusion can be divided into short-range diffusion and long-range diffusion. When atoms only diffuse on the short-range level, new phase growth will not lead to composition change. On the contrary, new phase growth relying on atomic long-range diffusion is accompanied by composition change.

(1) New phase growth speed when diffusion is on short-range level without composition change

In this case, new phase growth speed is controlled by interface diffusion (short-range diffusion). Assuming the parent phase is γ, the new phase is α, and the phase transformation happens during the cooling process. Atoms in the parent phase vibrate around their equilibrium position at a vibration frequency of ν_0. Because of the existence of energy fluctuation, some atoms get additional energy surpassing phase transformation energy barrier ΔG and leave their equilibrium position migrating into the new phase. The probability for vibrating atoms surpassing energy barrier is $\exp((-\Delta g)/kT)$, thus the frequency for atoms in the γ phase to enter into the α phase can be expressed as follows:

$$\nu_{\gamma\to\alpha} = \nu_0 \exp\left(\frac{-\Delta g}{kT}\right) \tag{9.26}$$

In the same way, atoms in the α phase could also move across the phase interface into the γ phase, but these atoms need more energy to surpass the energy barrier $(\Delta g + \Delta G_{\alpha\to\gamma})$, and their vibration frequency is

$$\nu_{\alpha\to\gamma} = \nu_0 \exp\left[\frac{-\left(\Delta g + \Delta G_{\alpha\to\gamma}\right)}{kT}\right] \tag{9.27}$$

Apparently, interface migration speed or new phase growth speed is proportional to the net frequency for atoms migrating from the parent phase to the new phase $\Delta\nu = \nu_{\gamma\to\alpha} - \nu_{\alpha\to\gamma}$, that is:

$$\nu \propto \exp\left(\frac{-\Delta g}{kT}\right)\left[1 - \exp\left(\frac{-\Delta G_{\alpha\to\gamma}}{kT}\right)\right] \tag{9.28}$$

Two cases are discussed in the following:

a) The supercooling is very small. In this case, $\Delta G_{\alpha\to\gamma} \to 0$. Since $e^x \approx 1 + x$ (when $|x|$ is very small), formula (9.28) can be simplified as

$$\nu \propto \frac{\Delta G_{\alpha\to\gamma}}{kT} \exp\left(\frac{-\Delta g}{kT}\right) \tag{9.29}$$

This shows that when the supercooling is very small, the new phase growth speed is proportional to the free energy difference between the parent phase and the new phase, indicating that the new phase growth speed increases when temperature drops.

b) The supercooling is very large. In this case, $\Delta G_{\alpha\to\gamma} \ll kT$, $\exp\left(\left(-\Delta G_{\alpha\to\gamma}\right)/kT\right) \to 0$, and formula (9.28) can be simplified as

$$\nu \propto \exp\left(\frac{-\Delta g}{kT}\right) \tag{9.30}$$

This shows that when the supercooling is very large, the new phase growth speed decreases exponentially when temperature drops.

In summary, during the whole phase transformation process, new phase growth speed exhibits increase first and then decrease when the temperature drops, as shown in Fig. 9.16.

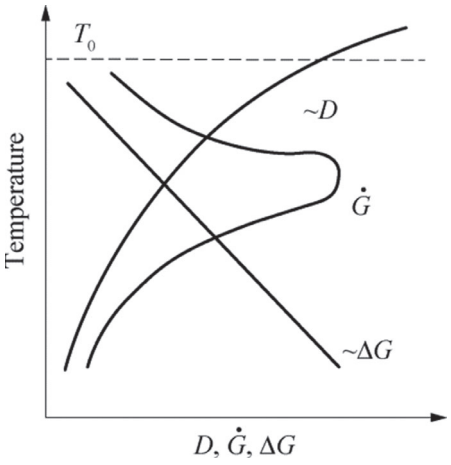


Fig. 9.16: Relationship between the nuclei growth speed and temperature.

(2) New phase growth speed when diffusion is on long-range level with composition change

When the new phase growth is realized by atomic diffusion, redistribution of solute atoms will be inevitably caused, and the growth speed is controlled by long-range atomic diffusion. When phase transformation takes place at a certain temperature, two phases have fixed compositions at the interface, which can be derived from phase diagrams. There are only two possible situations, solute atom concentration in the new phase is higher or lower than that of the parent phase, as shown in Fig. 9.17. There exists concentration gradient between the interface (at the parent phase side) and the parent phase interior, and diffusion of the solute atoms from high-concentration region to low-concentration region will inevitably happen. The result of the diffusion is that the solute concentration near the interface (at the parent phase side) will increase or decrease, broking the phase equilibrium at that temperature. Only if the interface moves towards the parent phase (new phase grows), the new phase in Fig. 9.17(a) excludes solute atoms, or the new phase in Fig. 9.17(b) absorbs solute atoms, the phase interface can be sustained until the concentration gradient inside the phase is absent and a dynamic equilibrium is achieved. When the temperature drops, the above process will repeat until the whole phase transformation is over.

The phase interface migration speed can be calculated according to the law of diffusion. Assuming the unit area phase interface moves by dx in the time of dt, the amount of solute atoms absorbed or excluded for the volume increase of the new

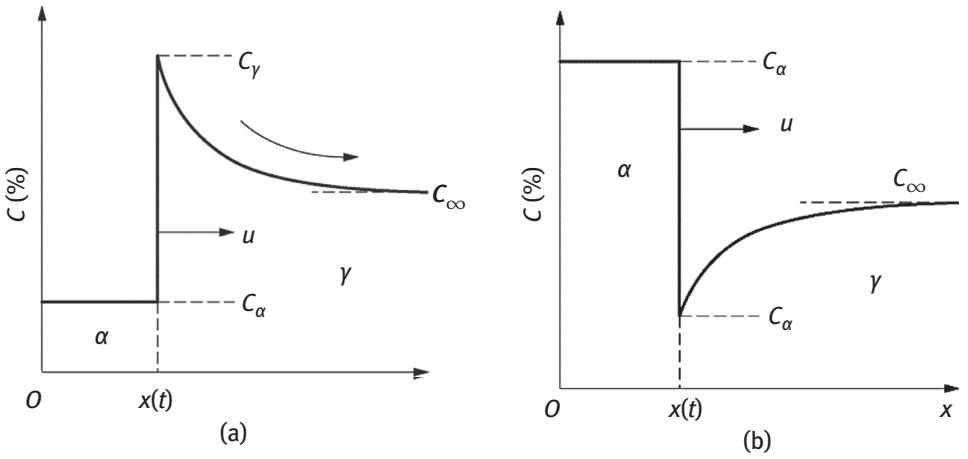


Fig. 9.17: Concentration distribution of solute atoms during the nuclei growth process. (a) The concentration solute atoms in the new phase is lower than that in the parent phase and (b) the concentration solute atoms in the new phase is higher than that in the parent phase.

phase is $|\Delta G_{\alpha/\gamma}|dx$, where $\Delta G_{\alpha/\gamma}$ is the concentration difference at the interface of the two phases. The absorption and exclusion of solute atoms for the new phase are both diffusion process inside the parent phase. Assuming the diffusion process is steady-state diffusion, the amount of atoms absorbed or excluded in unit area and unit time is $D(\partial C_r/\partial x)_{x_0}dt$, and there exists:

$$|\Delta C_{\alpha/\gamma}|dx = D(\partial C_r/\partial x)_{x_0}dt \tag{9.31}$$

Then the new phase growth speed can be expressed as

$$v = dx/dt = D/|\Delta C_{\alpha/\gamma}| \times (\partial C_r/\partial x)_{x_0} \tag{9.32}$$

This shows that the new phase growth speed v is proportional to diffusion coefficient D, also proportional to the concentration gradient of the parent phase near the interface $\partial C_\gamma/\partial x$, and inversely proportional to equilibrium concentration difference of the two phases on the interface $\Delta C_{\alpha/\gamma}$.

(3) Relationship between temperature and the new phase growth speed of the diffusion-type phase transformation

New phase growth speed of the diffusion-type phase transformation is controlled by driving force and diffusion coefficient. For the cooling transformation process, the relationship between driving force as well as diffusion coefficient and the transformation temperature is contradictory. The greater the degree of undercooling, the greater the driving force, and the harder the atomic diffusion. As a result, there is a maximum value of the phase transformation speed during the cooling process. For the heating transformation process, the relationship between driving force as well

as diffusion coefficient and the transformation temperature is consistent. The greater superheating, the greater the driving force, and the stronger the atomic diffusion capacity. As a result, the phase transformation rate increases monotonically with temperature.

9.3 Phase transformation macro kinetic equation

Solid-state phase transformation speed is determined by nucleation rate and growth speed of the new phase, both of which change with temperature and time. As a result, it is very hard to get a precise formula of the phase transformation speed. Assuming that nucleation rate and growth speed do not change with time at a certain temperature, Johnson–Mehl derived the relationship between the volume fraction of the new phase f and the time τ, which is the famous Johnson–Mehl formula:

$$f = 1 - \exp\left(-\frac{\pi}{3} N G^3 \tau^4\right) \tag{9.33}$$

where N is the nucleation rate; G is the linear growth speed of nucleus.

Johnson–Mehl formula is suitable for diffusion-type phase transformation with constant nucleation rate and linear growth speed.

In fact, nucleation rate and linear growth speed change with time, and Avrami empirical formula can be used in this case:

$$f = 1 - \exp(-k\tau^n) \tag{9.34}$$

where k is the constant, which is determined by phase transformation temperature, parent phase composition, and grain size; n is the constant, which is determined by the type of phase transformation.

Most of experimental data of solid-state phase transformation is in good agreement with Avrami empirical formula (9.34).

9.4 Phase transformation kinetic curve

Aiming at different values of N and G in formula (9.33) (i.e., different phase transformation temperatures), the relation curve between the new phase volume fraction and time can be derived. These curves all exhibit an "s" shape, as shown in Fig. 9.18(a). Their characteristic is that phase transformation speed is relative low at the beginning and ending but highest in the middle. In real production, it is more convenient to characterize phase transformation by temperature–time coordinate. Connecting the beginning points and ending points at different temperatures in Fig. 9.18(a) by smooth curves respectively, Fig. 9.18(b) is obtained. Because this figure exhibits "C" shape, it is called C-curve, or time–temperature–transformation curve.

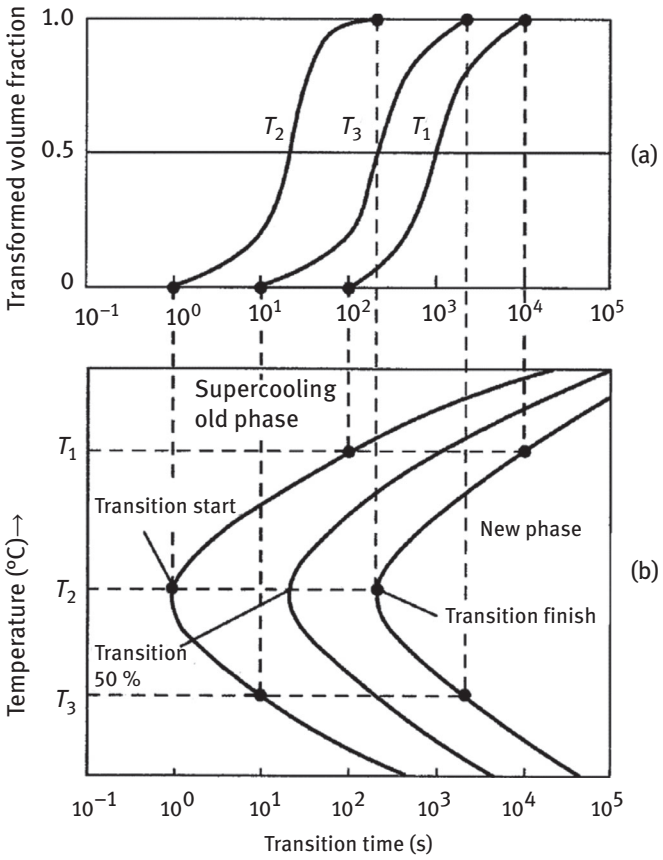


Fig. 9.18: Kinetic curves of phase transformation (a) and the corresponding isothermal transformation diagram (b).

In the left of transformation start line is the parent phase supercooling region. Although the thermodynamic phase transformation condition is satisfied, phase transformation does not happen in this region since a nurture period is needed. The temperature corresponding to the shortest nurture period is called "the nose temperature." When supercooling is too large or too small, the nurture period will be elongated and the supercooled parent phase will be stabilized, which is caused by the inconsistence between relations of driving force as well as atomic diffusion capacity with the temperature. In the right of transformation end line is the new phase region. The horizontal distance between start and end lines represents the transformation quantity of the new phase at different temperatures.

Kinetic curves for isothermal (cooling) diffusion-type phase transformation all exhibit "C" shape. C-curves of metallic materials are all obtained by experiments.

Chapter 10
Functional properties of materials

Solid materials are generally divided into two categories in terms of performance: structural materials and functional materials. Strength and toughness are the primary properties of structural materials. Whereas a particular function, such as electrical property, thermal property, magnetic property, or optical property, is a primary property of the functional materials. The performance of functional materials depends on the electron structure and electron motion of atoms (revolving, scattering, inspiring, jumping, and so on), which is different from structural materials. The performance of structural materials depends on the bonding between atoms (e.g., metallic bonds, ionic bonds, covalent bonds, and hydrogen bonds) and microstructure (e.g., crystal structure, grain size, phase morphology, dislocation substructure, and second-phase characteristics), but is irrelevant to the electron movement. This chapter reviews the physical basis of the functional properties. Emphasis is put on the description of the behaviors and mechanisms of the functional materials with electricity, heat, magnetism, light behavior, and other functional properties.

10.1 Introduction to physical basis of functional materials

10.1.1 Energy band

Energy band theory is the main theoretical basis for studying the electron motions in solids. It is the starting point of energy band theory that the shared electrons are no longer restricted to individual atoms, but move in the whole solid. The electrons are on the different discrete energy levels for the single atom. When a large number of atoms form crystals, energy levels of every atom are separated because of the overlap of electron clouds. In each level, the electron energy is approximately continuous because of very small energy difference among the electrons in the same energy level. The energy that is continuously distributed in the energy level forms an energy band. The relationship of the energy level and the energy band is shown in Fig. 10.1. The energy band of the lower energy level is narrower, while that of higher energy level is wider. The lowest energy band corresponds to the innermost layer electrons which have the small electron orbit and rarely overlap with each other. This leads to the narrow energy band. On the other hand, the electrons in the outer orbit have higher energy, and thus are easily disturbed. This causes more overlap between different atoms, forming wide energy band. Ignoring the interaction between atomic states, there is a simple corresponding relation between the energy band and the energy level. According to the subshell electron energy levels

https://doi.org/10.1515/9783110495379-010

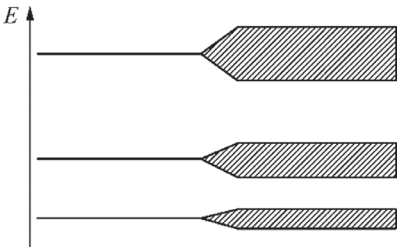


Fig. 10.1: Energy levels and energy bands.

of angular momentum quantum number, *s*, *p*, *d*, . . ., the corresponding energy bands are named as *ns* band, *np* band, *nd* band, and more. Some regions between the energy bands, called forbidden band or band gap, are energy levels that electrons do not possess. Considering a solid consisting of *N* atoms, each energy band contains *N* divisive energy sublevels. Each sublevel can accommodate two electrons with opposite direction of spin. For example, 2*s* energy band contains *N* divisive energy sublevels and can accommodate 2*N* electrons. Three 2*p* energy bands contain 3*N* energy sublevels and can accommodate 6*N* electrons. It is also possible that the energy band is partially filled. For instance, a copper atom contains one electron in 4*s*. *N* copper atoms should contain *N* 4*s* energy sublevels and accommodate 2*N* electrons. But *N* electrons are actually available. Only half of the 4*s* energy band are filled.

10.1.2 Fermi energy

Different from the classical electron theory, the free electrons which are 10^4 times higher than gas molecules in the density obey Fermi-Dirac distribution. In thermal equilibrium, the probability of the free electrons in energy state E can be indicated by the following formula:

$$f = \frac{1}{e^{(E-E_F)/(kT)} + 1} \tag{10.1}$$

where f denotes the Fermi–Dirac distribution function, E_F is the Fermi energy at T temperature, which is equal to the free energy increment after adding one electron. k is the Boltzmann constant.

At 0 K, the relationship of f and E is presented with the solid line as shown in Fig. 10.2. In this case, $f = 1$ at $E \leq E_F$, while $f = 0$ at $E \geq E_F$. All energy states that are less than Fermi energy are occupied by electrons ($f = 1$). According to the lowest energy principle, the electrons gradually fill up the energy levels below E_F starting from the lowest energy level. Fermi energy is the highest energy level of free electrons at 0 K. All the energy states above E_F have no electrons ($f = 0$), which are empty energy state (also called empty state). If $T > 0$ K, $f = 1/2$ at $E = E_F$, $1/2 < f < 1$

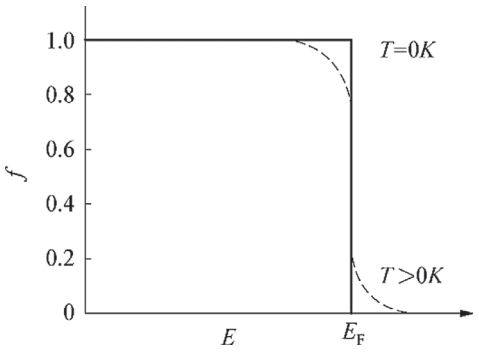


Fig. 10.2: Fermi–Dirac distribution.

at $E < E_F$, and $0 < f < 1/2$ at $E > E_F$. These are shown as dashed line in Fig. 10.2. The steep part of f locates in a small energy region (approximately 0.1 eV) near E_F. The value of f changes sharply from 1 at $E < E_F$ to 0 at $E > E_F$. In other words, when $T > 0$ K, the energy levels below E_F are filled up by electrons, and those above E_F are almost empty. It can be inferred that the unfilled energy level must be close to E_F. The distribution function indicates that a small amount of electrons with energy near E_F can absorb heat to jump into the high energy state above E_F. As a result, the empty energy levels above E_F may partially be occupied by the jumped electrons. Because the electrons with the energy far below E_F cannot be excited via absorbing heat, the electrons jumped to the high energy state by this way are very limited.

10.2 Electrical property

10.2.1 Description of electrical property

The electrical property of materials refers to the response to the external electrical field. This section focuses on the apparent description of the conductive behavior and discussion on the conductance mechanism of materials. The effect of energy band structure on the conductive ability is also highly emphasized. The principles involved are suitable for interpretation of electrical characteristic of various materials including metal, semiconductor, and insulator.

The ability to transfer current is generally an important electrical property of material. The current, I (unit: ampere), depends on the applied voltage, V (unit: volt), according to Ohm's law:

$$V = IR \tag{10.2}$$

where R denotes the electrical resistance of material (unit: ohm). It is influenced by the shape of sample, but independent on the electrical current for most materials.

The electrical resistivity, ρ, has nothing to do with the shape of materials, but has a relation to the electrical resistance:

$$\rho = \frac{RA}{l} \tag{10.3}$$

where l indicates the length of a rod material. A indicates the cross sectional area of this rod. The unit of ρ is ohm-meter ($\Omega \cdot \mathrm{m}$). According to eq. (10.2) it can be expressed as

$$\rho = \frac{VA}{Il} \tag{10.4}$$

The electrical conductivity, σ, is another way of describing the electrical property of materials. It is inversely proportional to the electrical resistivity:

$$\sigma = \frac{1}{\rho} \tag{10.5}$$

The electrical conductivity is a kind of ability of conduction current. The unit of σ is the reciprocal of the ohm-meter $(\Omega \cdot \mathrm{m})^{-1}$. So the electrical property described by the electrical conductivity or the electrical resistivity is the same.

Except for eq. (10.2), the Ohm's law can be also expressed as

$$J = \sigma\, \zeta \tag{10.6}$$

where J indicates the current per unit area (= I/A), which is also called electrical current density. ζ indicates the electrical field intensity, equal to the ratio of the voltage between the two ends of a rod material and its length, l:

$$\zeta = \frac{V}{l} \tag{10.7}$$

The equivalency of the two expressions (eqs. (10.2) and (10.6)) of Ohm's law is easy to be proven.

Solid materials exhibit an amazing wide range of electrical conductivity, up to 27 orders of magnitude. According to their conductivity, solid materials are classified into three groups: conductors, semiconductors, and insulators. Metals are good conductors, and have unique electro-conductivity with the order of magnitudes 10^7 $(\Omega \cdot \mathrm{m})^{-1}$. Insulators have the quite low electro-conductivity with the order of magnitudes 10^{-20}–10^{-10} $(\Omega \cdot \mathrm{m})^{-1}$. The electroconductivity of semiconductors is moderate, in the range of 10^{-6}–10^{4} $(\Omega \cdot \mathrm{m})^{-1}$.

The current is derived from the motion of charges. It is the response to the acting force of external electrical field. The positive charges are accelerated in direction of the electrical field applied, while the negative charges are accelerated in the opposite direction of the electrical field. For most of metallic materials, the current is caused by the motion of electron charges, which is called electron conduction.

Additionally, the motion of ion charges may also generate electrical current in ionic materials, which is called ion conduction. In this section, only electron conduction will be discussed.

10.2.2 The conductivity based on the energy band theory

Only the electrons with energy above the Fermi energy can be affected by the electrical field. They are involved in conducting currents, and are called free electrons in metals. In semiconductors and insulators, the vacancies among the charged entities, called holes, can act as a conductive medium. The holes generally have energy below the Fermi energy. The conductivity is a function of the free electrons or/and the number of vacancies. Obviously, the discrepancy among conductors, semiconductors and insulators is the difference in number of free electrons and vacancies.

In metals, the electrons become free only when they are excited into the empty energy states above the Fermi energy. The typical energy band structures of metals are shown in Fig. 10.3(a) and (b). There is an empty state near the highest filled state, E_F. The electrons can be activated into the lower empty states by a very little energy. The electrical field can usually activate a large number of electrons to jump into the empty state to conduct.

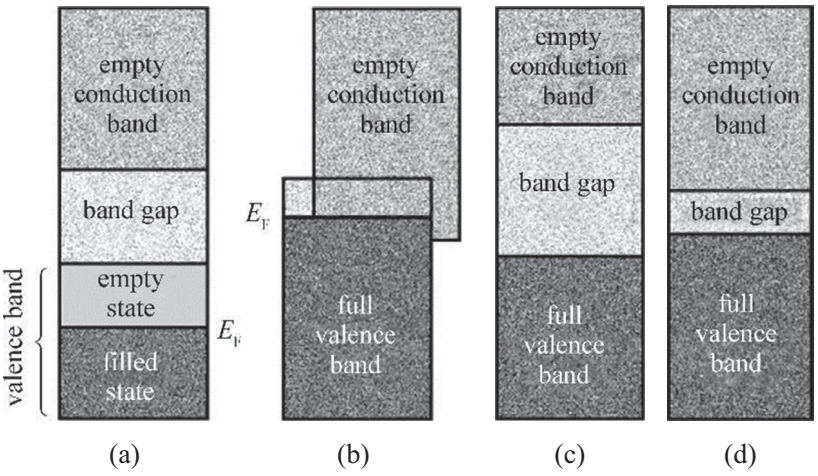


Fig. 10.3: The electron energy band structures of various materials at 0 K: (a) Energy band structures of metals (i.e., Cu), (b) overlap of empty band and full valence for some metals (i.e., Mg), (c) insulator: a wide band gap (>2 eV) exists between full valence band and empty band, and (d) semiconductor: a narrow band gap (<2 eV) exists between full valence band and empty band.

In semiconductors and insulators, there is no empty state on the top of the full valence band. If electrons are to become free, they must be excited and should be able to jump over the band gap to enter into the bottom of the empty conduction band.

It occurs when electrons gain enough energy. This energy should correspond to the energy difference between the two states of the electrons, and be approximately equivalent to the band gap energy E_g. Generally, the band gap is about a few electron volts in wide. This indicates that a large electrical field is needed to excite an electron to jump over the band gap. Ordinarily, the excited energy does not always come from the electrical field, does come from light or heat. The amounts of electrons that are heat excited to enter into the empty conduction band are dependent on temperature and the band gap energy. At a certain temperature, the wider energy gap corresponds to the smaller probability of the electrons excited into the conduction band. This means that the material with a wide band gap generally has low electrical conductivity. Therefore, the difference between semiconductors and insulators is that they have different band gap widths as shown in Fig. 10.3(c) and (d). The band gap of semiconductors is relatively narrow, while that of insulators is relatively wide. Obviously, as increasing temperature, both semiconductor and insulator can gain more thermal energy that excites more electrons jumping into the empty conduction band. This causes the enhancement of the electrical conductivity.

10.2.3 Electron mobility

When an electric field is applied, the free electrons in material are impacted by a force arisen from the electric field. These electrons with negative charges are accelerated along the opposite direction of the electric field. The accelerated electrons generally do not react with atoms in an ideal crystal in accordance with the principle of quantum mechanics. All the electrons should be accelerated when a electric field exists. This leads to the continuous increase of the electric current with time. However, at the moment of applying an electric field, the electric current reaches a constant value. This indicates that some frictions on the electrons exist to hinder the acceleration of the external electric field. The friction comes from the electron scattering caused by defects in crystal lattice. These defects include vacancies, interstitial and impurity atoms, dislocations, and even vibration of the atoms themselves. All these defects cause kinetic energy loss of the electrons and change of motion direction. After balancing these obstacles, the net electron movement in the opposite direction of the electric field reaches a stable value. The stable electron flow is called electric current.

The electron scattering is considered to be a resistance to the current, which can be measured by the electron drift velocity and electron mobility. The drift velocity, v_d, is defined as the average electron velocity in the direction of external electric field force. It is proportional to the electric field intensity, ζ:

$$\upsilon_d = \mu_e \zeta \tag{10.8}$$

where the proportional constant μ_e is called electric mobility. Its unit is $m^2/V{\cdot}s$.

For most materials, the electrical conductivity can be described as

$$\sigma = n|e|\mu_e \tag{10.9}$$

where n denotes the quantities of free electrons per unit volume. $|e|$ is the absolute value of an electron charge (1.6×10^{-19} C). The electrical conductivity is proportional to the electron mobility and the quantities of free electrons.

10.2.4 The electrical resistivity of metal

As mentioned earlier, most metals are good electrical conductors. They exhibit high electrical conductivities at room temperature. The typical values of several common metals are collected in Table 10.1. The high conductivity is attributed to the abundance of free electrons in metals. In other words, the electrons in metals can be easily excited to jump into the empty state above Fermi energy. So the n has a large value in eq. (10.9).

Table 10.1: Electrical conductivities of some common metals and alloys at room temperature[1, 2].

Metals	Electrical conductivity $(\Omega \cdot m)^{-1}$	Metals	Electrical conductivity $(\Omega \cdot m)^{-1}$
Silver	6.9×10^7	Brass($Cu_{70}Zn_{30}$)	1.6×10^7
Copper	6.0×10^7	Iron	1.0×10^7
Gold	4.3×10^7	Mild steel	0.6×10^7
Aluminum	3.8×10^7	Stainless steel	0.2×10^7

Since the electrical resistivity is inversely proportional to the electrical conductivity, it is often used to represent the electrical conductivity of metals. The conduction electrons in metal can be scattered by crystal defects that are main obstacles to electron flow. More defects correspond to large electrical resistivity or small electrical conductivity. The amount of defects in a metal depends on its composition and processing state (e.g., heat treatment and deformation extent). It is also related to temperature. It has been proved experimentally that the total electrical resistivity is attributed from the three contributions: thermal vibration, impurities and plastic deformation. Because each of them affects electron scattering independently, the total electrical resistivity can be expressed as

$$\rho_{\text{total}} = \rho_t + \rho_i + \rho_d \tag{10.10}$$

where ρ_t, ρ_i, and ρ_d represent the contributions of temperature, impurities, and deformation, respectively, to the electrical resistivity. The above equation is also called Matthiessen's rule. As an example, the electrical resistivity of pure copper and its alloys (Cu–Ni) plotted in a resistivity–temperature coordinates clearly shows the respective contributions of the three factors as shown in Fig. 10.4. After adding different amounts of nickel in copper, the curve moves upward for a certain distance accordingly, indicating the impurity contribution superimposed on the basic resistivity of pure copper. The deformed alloy also reveals the contribution of deformation to the resistivity.

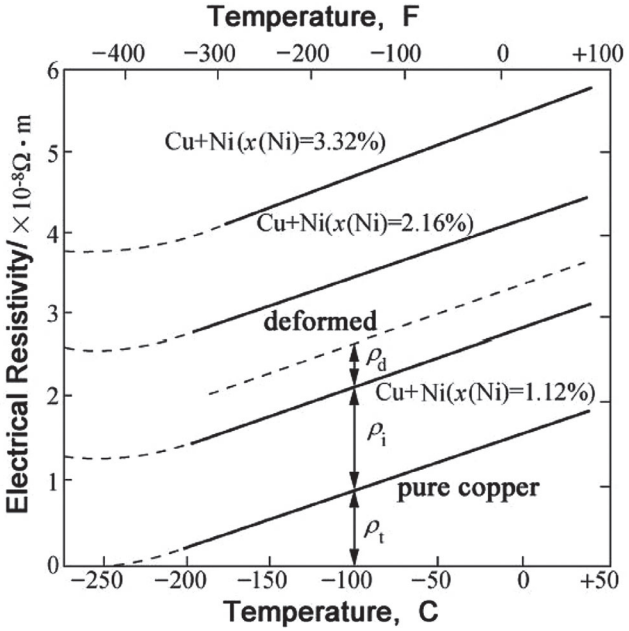


Fig. 10.4: Relationships between electrical resistivity and temperature for copper and Cu–Ni alloys.

10.2.5 The electrical conductivity of intrinsic and extrinsic semiconductor

Unlike metals, semiconductors exhibit lower electrical conductivity. Their unique electrical characteristics make them of some special applications. These materials are particularly sensitive to a small amount of impurity. High-purity materials exhibit semiconducting behavior via their inherent electron structure, which are called intrinsic semiconductors. The materials with impurity that dominates the semiconducting property are called extrinsic semiconductors.

The electron band structure of the intrinsic semiconductor at 0 K conforms to the characteristics shown in Fig. 10.3(d). Such an electron energy band structure is separated into a fully filled valence band and an empty conduction band by a narrow band gap that is typically smaller than 2 eV. Silicon and germanium are two typical intrinsic semiconductors. Their band gap energy is about 1.1 and 0.7 eV,

respectively. They both belong to group ⅣA in the periodic table, and have covalent bond. Some compound semiconductors also exhibit intrinsic semiconducting behavior.

The elements between groups ⅢA and ⅤA can form such compounds, for example, GaAs and InSb. The elements in groups ⅡB and ⅤA form compounds that also exhibit semiconductor behavior, for example, CdS and ZnTe. Since the two elements are far each other in the periodic table, their atom bonding tends to be ionic. Their band gap energy also becomes larger. Thus, these compounds exhibit more insulative. For easy reference, some electrical properties of the intrinsic semiconductors and compounds discussed above are collected in Table 10.2.

Table 10.2: Electrical properties of the intrinsic semiconductors and compounds at room temperature[3].

Materials	Band gap energy (eV)	Electrical conductivity $(\Omega \cdot m)^{-1}$	Electron mobility $(m^2 \cdot (V \cdot s)^{-1})$	Vacancy mobility $(m^2 \cdot (V \cdot s)^{-1})$
Elements				
Si	1.11	4×10^{-4}	0.14	0.05
Ge	0.67	2.2	0.38	0.18
III–V compounds				
GaP	2.25	–	0.05	0.002
GaAs	1.35	10^{-6}	0.85	0.45
InSb	0.17	2×10^{4}	7.7	0.07
II–VI compounds				
CdS	2.40	–	0.03	–
ZnTe	2.26	–	0.03	0.01

In the semiconductor, an electron excited to conduction band leaves a vacancy in the covalent band as shown in Fig. 10.5(b). The position of a runaway electron (vacancy) is called hole and confirmed to move under an electric field, which can be realized by other valence electrons filling the incomplete bond (Fig. 10.5(c)). Hole is considered to be a positive charge with about 1.6×10^{-19}C (same measure as the negative charge in an electron). The holes and the electrons excited in the semiconductor move in opposite directions under the electric field. Both can be scattered by the defects in the semiconductor.

Considering effect of the holes and the excited electrons in the intrinsic semiconductor, the conductivity formula (10. 9) can be modified by adding the contribution of hole current:

$$\sigma = n|e|\mu_e + p|e|\mu_h \tag{10.11}$$

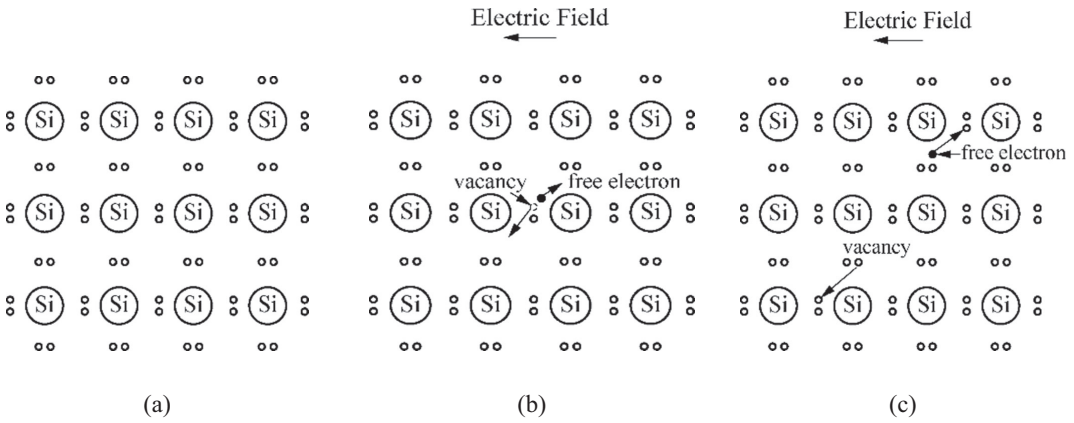


Fig. 10.5: Illustration of conductance behavior in intrinsic semiconductor silicon: (a) covalent bond structure of silicon, (b) and (c) motion of electrons and vacancies under an external electric field after excitation.

where p denotes the amount of holes per unit volume. μ_h indicates the migration rate of the holes. Since each electron excited into conduction band leaves a vacancy in the valence band, one has $n = p$. Thus:

$$\sigma = n|e|(\mu_e + \mu_h) = p|e|(\mu_e + \mu_h) \tag{10.12}$$

It is found that μ_e value is always larger than that of μ_h for those intrinsic semiconductors and compounds discussed above as listed in Table 10.2.

In practice, the commercial products used are always extrinsic semiconductors. Their semiconducting behavior depends on impurities in the material. The small amount of impurities can introduce extra electrons or holes. As an example, even one impurity in 10^{14} atoms can make the silicon to be extrinsic.

The silicon semiconductor is considered again in order to explain how the extrinsic semiconductor behaves. There are four valence electrons in a silicon atom. Every valence electron combines with the adjacent silicon atom to form covalent bond. If a pentavalent impurity, from group ⅤA, such as P, As, and Sb, substitutes silicon, only four of the five electrons form covalent bonds with the four adjacent silicon atoms. The remaining one is attracted by weak static electricity around the impurity atoms as shown in Fig. 10.6(a). Because of the relative small bond energy (on the order of 0.01 eV), the electron easily removes from the impurity atom and becomes a free electron or a conducting electron.

The electron with weak bond energy has its own single energy level located in band gap. The level is near the conduction band as shown in Fig. 10.7(a). The bond energy of an electron corresponds to the energy required to excite an electron from an impurity state to an energy state in the conduction band. An electron is donated to the conduction band for every excitation event. The impurities are called donors.

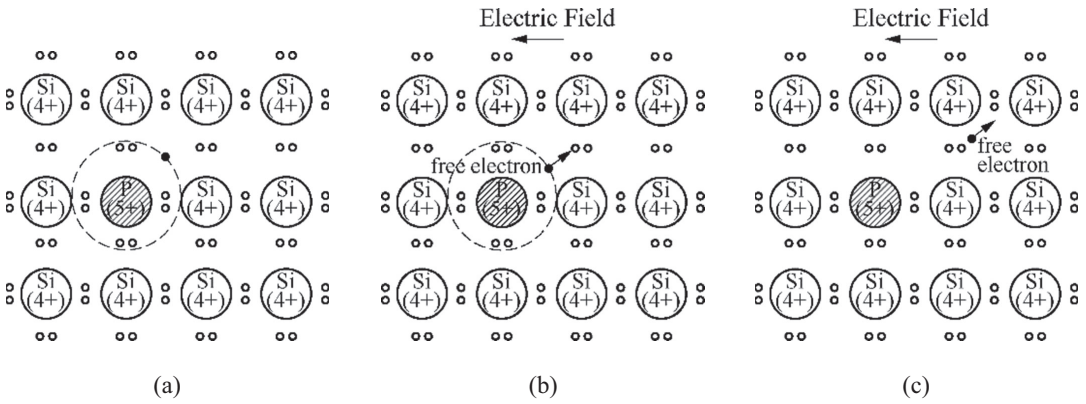


Fig. 10.6: Electronic bonding model for nonintrinsic n-type semiconductors. (a) Five-valent phosphorus atom substituted silicon atom, (b) superfluous electrons become free electrons after excitation, and (c) the motion of free electrons in an electric field.

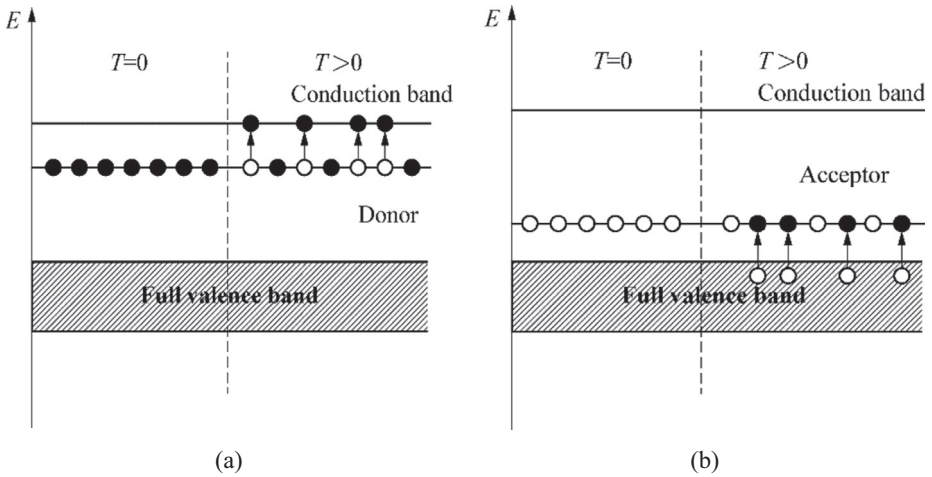


Fig. 10.7: Donor and acceptor: (a) n-type and (b) p-type.

Each donor electron comes from the impurity energy state in the band gap. Unlike the intrinsic semiconductors, there is no hole generated in the valence band.

The heat energy acquired at room temperature can excite a large number of electrons from the donor state. The escaping electrons from the valence band in intrinsic semiconductors are very few (Fig. 10.5(b)). Therefore, the amount of free electrons in conduction band is much larger than that of holes in valence band, namely $n \gg p$, thus eq. (10.12) can be approximately rewritten as

$$\sigma \approx n|e|\mu_e \tag{10.13}$$

This kind of material is described as *n*-type extrinsic semiconductor. Their electrical conductivity is mainly determined by the electron concentration.

If trivalence impurities such as Al, B, Ga in the group IIIA in the periodic table substitute the matrix atoms in the intrinsic semiconductors, for example, silicon or germanium, the opposite effect occurs. Lack of an electron in the covalent bonds around each impurity atom can be taken as a hole (vacancy). It is weakly bonded to the impurity atom, and can move by adjacent electron transferring as shown in Fig. 10.8. A hole moves via exchanging positions with an electron. The essence of hole movement is electron movement. So, a hole moving contributes to the conductive process similar to a donor electron mentioned earlier.

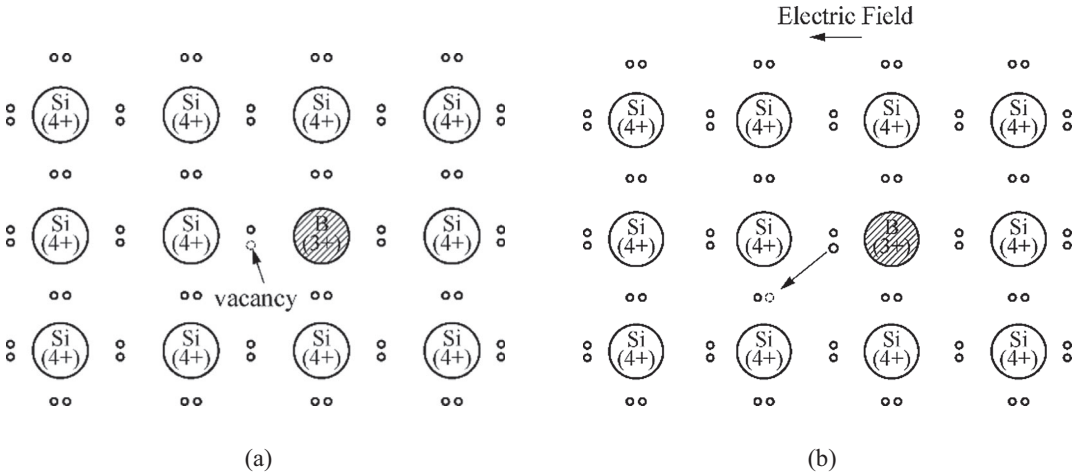


Fig. 10.8: Electronic bonding model for intrinsic *p*-type semiconductors: (a) trivalence boron atom substituted a silicon atom and (b) vacancy movement under an electric field.

The extrinsic activation generating holes can be illustrated by the energy-band model. Every excited impurity atom brings an energy level in band gap. This level is above but near the valence band. Electrons in valence band can be thermally excited into the impurity energy level, and leaves holes in the valence band as shown in Fig. 10.7(b). In this process, only holes are generated in valence band. Free electrons cannot be created in either conduction band or impurity state. This impurity is called acceptor, because it can accept one electron from valence band and leave an electron vacancy. The energy state created by this kind of impurity in the band gap is called acceptor state.

In such extrinsic semiconductor, the number of holes is much larger than that of free electrons (i.e., $p \gg n$), because the holes can be created not only by extrinsic excitation of acceptor, but also by intrinsic excitation of valence band. This material is called p-type semiconductor. Since positive charge carries dominate conductance, holes are main carries contributed to the electrical conductivity. Accordingly, eq. (10.12) can be simplified as

$$\sigma \approx p|e|\mu_h \tag{10.14}$$

Extrinsic semiconductors (including *n*-type and *p*-type) can be firstly made of high purity material (total impurity is under 10^{-7} at.%). Then one can mix the impurity into the high-purity material as donors or acceptors. The semiconducting properties may be tailored by adjusting the impurity concentration. This procedure in synthesizing semiconductors is called doping.

In extrinsic semiconductors, the charge carriers can be created by heat activation at room temperature. They exhibit relatively high electrical conductibility at ambient temperature. Most of these semiconductors are designed as electronic components worked at room temperature.

10.2.6 Electrical conductivity and dielectricity of insulator

Insulators have a wide band gap between the fully filled valence band and the empty conduction band as shown in Fig. 10.3(c). The band gap energy is generally larger than 2 eV, so that only few electrons can be thermally excited to jump over the band gap at room temperature. This means very small electrical conductivity. In fact, all the ceramic materials and covalent bond polymers are insulators. Some of them are listed in Table 10.3 to show their electrical conductivities at room temperature. In many cases they are used because of their insulativity. In this regard, high electrical resistivity is very much expected. Similar with semiconductors, insulators increase in electrical conductibility as temperature increases.

Table 10.3: Electrical conductivities of several nonmetals at room temperature[4].

Materials	Electrical conductivity $(\Omega \cdot m)^{-1}$	Materials	Electrical conductivity $(\Omega \cdot m)^{-1}$
Graphite	10^{5}	**Polymers**	
Ceramics		Phenolic aldehyde	10^{-9}–10^{-10}
Alumina	10^{-10}–10^{-12}	Nylon	10^{-9}–10^{-12}
Porcelain	10^{-10}–10^{-12}	Poly(methyl methacrylate)	$<10^{-12}$
Na,Ca-glass	$<10^{-10}$	Polyethylene	10^{-13}–10^{-17}
Mica	10^{-11}–10^{-15}	Polystyrene	$<10^{-14}$
		Polytetrafluoroethylene	$<10^{-16}$

Dielectric materials are a kind of nonmetallic materials that have a large electrical resistivity more than $10^{8}\ \Omega \cdot m$. Their response to the electrical field is electric induction rather than electrical conductance. Under an electrical field, the dielectric materials exhibit an electric dipole structure. The positive and negative charges tend to be separated at atomic or molecular level. Because of the interaction of dipole and electric field, dielectric materials are usually used as capacitor.

When a voltage is applied on a capacitor, one plate of the capacitor presents positive charge, and the other plate presents negative charge. The positive and negative charges on the two plates are balanced by the external electric field. The capacitance, C, depends on the quantity of electricity, Q, stored in the plate, which can be expressed as

$$C = \frac{Q}{V} \tag{10.15}$$

where V denotes the voltage applied on the capacitor. The capacitance unit is Coulomb per volt (C/V), that is, farad (F).

Considering a parallel plate capacitor with a vacuum cavity between the plates, its capacitance can be calculated with the below formula:

$$C = \varepsilon_0 \frac{A}{l} \tag{10.16}$$

where A indicates the plate area. l indicates the plate spacing. ε_0 is the vacuum permittivity that is a universal constant, that is, 8.85×10^{-12} F/m.

When the dielectric material is placed between the two plates, it becomes

$$C = \varepsilon \frac{A}{l} \tag{10.17}$$

where ε denotes the dielectric permittivity that is far larger than ε_0. The ratio of the two permittivities is called relative dielectric permittivity ε_r:

$$\varepsilon_r = \frac{\varepsilon}{\varepsilon_0} \tag{10.18}$$

The value of ε_r is much greater than unity. This means that the storage capacity of charge increases when the dielectric material is inserted into the space between plates. This effect mentioned above was first studied by Faraday in 1873 (Fig. 10.9). The relative dielectric permittivity represents a property of material. It should be considered as a key parameter when designing capacitors.

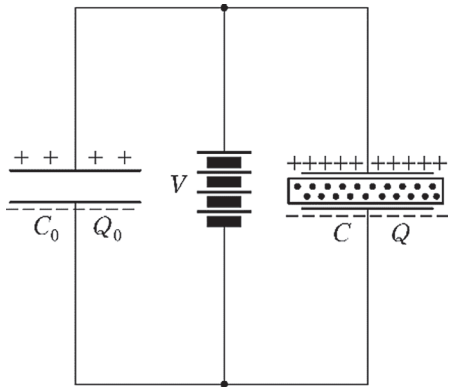


Fig. 10.9: Parallel plate capacitance.

The capacitance phenomenon can be well understood with the help of the field vector. The positive and negative charges in a dipole tend to separate. The electric dipole moment, P, related to the dipole charge, q, and the distance, d, between the two poles can be expressed as

$$P = qd \tag{10.19}$$

There is a potential between the two poles, which starts from the negative pole toward the positive pole. This potential vector is equivalent to the dipole moment. It can rotate toward the direction of the external electric field applied. The directional arrangement of dipoles is called polarization.

For the capacitor, the density of surface charge, D, that is, quantity of electric charge per unit area on capacitor plate (C/m^2), is determined by the intensity of the electric field, ζ. Under vacuum, it is

$$D_0 = \varepsilon_0 \zeta \tag{10.20}$$

where ε_0 is the proportional coefficient. There is the similar expression for the dielectric case:

$$D = \varepsilon \zeta \tag{10.21}$$

D is also named the dielectric displacement.

The polarization model discussed above for the dielectric material can be used to explain the increase of capacitance or dielectric constant. Considering a vacuum capacitor as shown in Fig. 10.9, the dielectric material inserted in the plates is polarized after an electric field is applied. This leads to significant increase in charges on the two plats of the capacitor. In this process, the surface of the dielectric material toward the positive plate of the capacitor accumulates negative charges, while another surface toward the negative plate accumulates positive charges.

In presence of the dielectric material, the surface charge density of the two plates in a capacitor can also be expressed as

$$D = \varepsilon_0 \zeta + P \tag{10.22}$$

where P represents the contribution of polarization. It can be taken as the increase in charge density after insertion of the dielectric material in the capacitor.

$$P = \frac{Q'}{A}$$

where Q' indicates the increase in charge on the plate of the capacitor after a dielectric material is inserted. A indicates the area of the plate. P and D are in the same unit of C/m^2.

The polarization can be considered as a polarization field in the dielectric material, which is caused by the rotation and adjustment of the dipole moments of atoms or molecules under the external electric field. It is simply equivalent to the total dipole moments per unit volume in the material. For many dielectric materials, P is proportional to ζ:

$$P = \varepsilon_0(\varepsilon_r - 1)\ \zeta \tag{10.23}$$

In this case, ε_r is independent on the electric field intensity. Some dielectric parameters of several dielectric materials are listed in Table 10.4 for reference.

Table 10.4: Dielectric constants and dielectric strengths of several dielectric materials.[5, 6].

Materials	Dielectric constants		Dielectric strengths ($V \cdot km^{-1}$)
	60 Hz	1 Hz	
Ceramics			
Titanate ceramics	–	15–10,000	80–483
Mica	–	5.4–8.7	1,609–3,218
Steatite ($MgO–SiO_2$)	–	5.5–7.5	322–563
Na,Ca-glass	6.9	6.9	402
Molten silicon oxide	4.0	3.8	402
Porcelain	6.0	6.0	64–644
Polymers			
Phenolic aldehyde	5.3	4.8	483–644
Nylon 66	4.0	3.6	644
Polystyrene	2.6	2.6	805–1,126
Polyethylene	2.3	2.3	724–805
Polytetrafluoroethylene	2.1	2.1	644–805

Since ceramics have larger dipole moment, their dielectric constants are mostly larger than that of polymers. The dielectric constants, ε_r, of polymers are generally in the range of 2–5. These materials are usually used in the insulation of electric wire, electric cable, engine, electric generator. In addition, they also can be used in the capacitor.

For example, a parallel capacitor with the plate area of 6.45×10^{-4} m^2 and the plate spacing of 2×10^{-3} m is applied with an electric voltage of 10 V. A dielectric material with relative dielectric constant 6.0 is placed in the zone between the two plates. The followings can be calculated:

a) Capacitance
b) The amount of charge stored in each plate
c) Dielectric displacement
d) Intensity of polarization

Solution:

a) $\varepsilon = \varepsilon_r \varepsilon_0 = (6.0)\left(8.85 \times 10^{-12}\right) = 5.31 \times 10^{-11}\text{F/m}$

$$C = \varepsilon \frac{A}{l} = 5.31 \times 10^{-11} \times \frac{6.45 \times 10^{-4}}{2 \times 10^{-3}} = 1.71 \times 10^{-11}\text{F}$$

b) $Q = CV = 1.71 \times 10^{-11} \times 10 = 1.71 \times 10^{-10}\text{C}$

c) $$D = \varepsilon\zeta = \varepsilon \frac{v}{\iota} = \frac{5.31 \times 10^{-11} \times 10}{2 \times 10^{-3}} = 2.66 \times 10^{-7}\text{C/m}^2$$

d) $$P = D - \varepsilon_0 \zeta = D - \varepsilon_0 \frac{v}{\iota} = 2.66 \times 10^{-7} - \frac{8.85 \times 10^{-12} \times 10}{2 \times 10^{-3}} = 2.22 \times 10^{-7}\text{C/m}^2$$

10.3 Thermal behavior

The thermal behavior of a material refers to the response of the material to heat. When a solid material is heated, the temperature rise and volume expansion will happen. If there is a temperature gradient in a material or medium, the heat will transfer from high temperature to low temperature until the temperature gradient disappears. The thermal property of materials can be characterized by their heat conductivity, heat capacity and thermal expansion coefficient.

10.3.1 Heat capacity

The temperature rise of a solid material after heated indicates that some energy is absorbed. The heat capacity indicates the ability of the material to absorb heat energy from the surrounding environment. It is defined as the heat energy needed for the temperature rise of the material by one degree. The heat capacity can be described mathematically as follows:

$$C = \frac{dQ}{dT} \tag{10.24}$$

where dQ indicates the heat energy that is needed to produce a temperature increment, dT. The energy needed for 1 mol material to increase 1 K in temperature is called molar heat capacity with the unit of $\text{J} \cdot (\text{mol} \cdot \text{K})^{-1}$. The heat capacity per mass is defined as the specific heat capacity, c, with the unit of $\text{J} \cdot (\text{kg} \cdot \text{K})^{-1}$.

The heat capacity can be measured at either constant volume or constant pressure. The constant volume heat capacity is expressed by C_v, while the constant pressure heat capacity is expressed by C_p. C_p is always larger than C_v. At or below room temperature the difference between C_p and C_v is very small.

In most of solid materials, heat energy is mainly consumed by increasing the vibrational energy of atoms. The vibration occurs with high frequency but low amplitude,

which can couple and transfer via the atom bonds. It can be considered as elastic waves or simple harmonic waves with short wavelengths and high frequencies. The vibration waves propagate in sound velocity in crystal solids. Generally, the vibrational heat energy is composed of a series of elastic waves with certain frequency range and distribution. The vibrational wave or heat energy can be quantized. The vibration energy of a single quantum is called a phonon that is similar to a quantum in the electromagnetic radiation. Sometimes the vibration wave is also called phonon.

In the process of electron conduction, the thermal scattering of free electrons occurs via the vibration waves that also contribute to the energy transfer during heat conduction.

For the solid with simple crystal structure, the values of C_v depend on temperature as shown in Fig. 10.10. C_v equals zero at 0 K, but increases rapidly as temperature increases. This is because the vibration wave becomes stronger at relatively higher temperature. At low temperature, the relation of C_v to temperature, T, can be written as

$$C_v = AT^3$$

where A is a constant independent of temperature. Above the Debye temperature, θ_D, C_v is independent on temperature, and is approximately equal to $3R$ (R: gas constant). Thus, the heat energy needed for increasing 1 K in temperature is constant, although the total energy in material increases with the increase of temperature. Because the value of θ_D is below room temperature for many materials, their C_v values are about 25 $J \cdot (mol \cdot K)^{-1}$ at room temperature. Table 10.5 collects several thermal parameters including C_p of some common materials for reference.

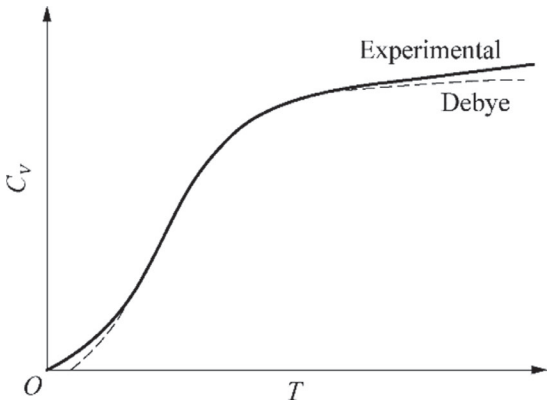


Fig. 10.10: Relations of the constant volume heat capacity to temperature.

Heat capacity increase also involves other thermal absorption mechanisms. For example, the electrons can absorb more energy via providing the higher kinetic energy for free electrons. When ferromagnetic materials are heated above Curie point, randomization of electron spin absorbs energy. In most cases, these contributions are relatively smaller than that of the vibration.

Table 10.5: Thermal properties of common materials[7, 8].

Materials	c_p $J \cdot (kg \cdot K)^{-1}$	a_l $\times 10^{-6}\ (°C)^{-1}$	k $W \cdot (m \cdot K)^{-1}$	L $\times 10^{-8} (\Omega \cdot W) \cdot K^{-2}$
Metals				
Aluminum	900	23.6	247	2.24
Copper	386	16.5	398	2.27
Gold	130	13.8	315	2.25
Iron	448	11.8	80.4	2.66
Nickel	443	13.3	89.9	2.10
Silver	235	19.0	428	2.32
Tungsten	142	4.5	178	3.21
1025 steel	486	12.5	51.9	
316 stainless steel	502	16.0	16.3	
Brass (70Cu–30Zn)	375	20.0	120	
Ceramics				
Aluminum oxide (Al_2O_3)	775	8.8	30.1	
Beryllium oxide (BeO)	1,050	9.0	220	
Magnesium oxide (MgO)	940	13.5	37.7	
Spinel ($MgAlO_4$)	790	7.6	15.0	
Ca,Na-glass	840	9.0	1.7	
Molten silicon oxide (SiO_2)	740	0.5	2.0	
Polymers				
Polyethylene	2,100	60–220	0.38	
Polypropylene	1,880	80–100	0.12	
Polystyrene	1,360	50–85	0.13	
Polytetrafluoroethylene	1,050	100	0.25	
Phenolic aldehyde	1,650	68	0.15	
Nylon 66	1,670	80–90	0.24	
Polyisoprene		220	0.14	

10.3.2 Thermal expansion

Most solid materials are thermal expansion and cold-shrink. Their changes in length are the function of temperature:

$$\frac{l_f - l_0}{l_0} = \alpha_l \left(T_f - T_0\right) \tag{10.25(a)}$$

or

$$\frac{\Delta l}{l_0} = \alpha_l \Delta T \tag{10.25(b)}$$

where l_0 and l_f are initial length and ultimate length when temperature changes from T_0 to T_f. The linear expansion coefficient, α_l, represents the dilation property

when the material is heated. The unit of α_l is K^{-1}. In general, heating or cooling affects the three dimensional size of an object, resulting a change in volume. The relationship between volume and temperature can be expressed as the following formula:

$$\frac{\Delta V}{V_0} = \alpha_v \Delta T \tag{10.26}$$

where V_0 is the initial volume. ΔV denotes the volumetric increment. α_v is the coefficient of volume expansion. Its value depending on the crystallographic orientation is often anisotropy. For the materials with isotropic expansion, the value of α_v is about $3\alpha_l$.

On atomic scale, the thermal expansion implies the increase in average distance between atoms. The potential energy of atoms has relations to their distance. At a certain temperature, there is an equilibrium atomic distance, at which the potential energy has a minimum value. At 0 K, the atomic vibration can be ignored. The equilibrium atomic distance is r_0 in this case. When heated up to a higher temperature, for example, T_1, T_2, T_3, the vibrational energy becomes E_1, E_2, E_3, respectively. The stronger vibration (i.e., higher potential energy) corresponds to the larger vibration amplitude. The average atomic distance shifts to right side as temperature increases according to the asymmetric curve as shown in Fig. 10.11(a). As the average atomic distance increases, the material expands.

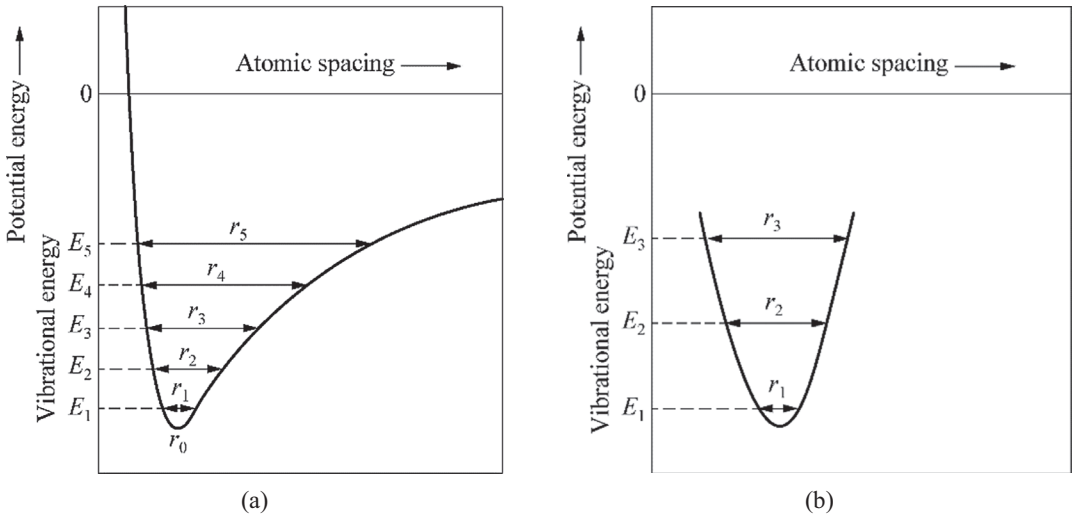


Fig. 10.11: The relationship between potential energy and atomic spacing: (a) asymmetric curve and (b) symmetric curve.

As discussed earlier, the larger vibration amplitude does not directly contribute to the increment of the atomic distance. The asymmetric potential energy curve causes larger average atomic distance as temperature increases, and thus results in the thermal expansion. If the curve is symmetric (Fig. 10.11(b)), the average atomic