The Martensitic Transformation

Manfred Ahlers
Centro Atômico 8400 S.C.de Bariloche. Argentina

Canhale: ahlers@cab.cnea.gov.ar

RESUMEN

La transformación martensítica es una transformación de fase de primer orden sin difusión en el estado sólido. La nueva fase, denominada martensita, se forma por procesos de nucleación y crecimiento. Se demuestra en este trabajo como se puede entender el movimiento coordinado de los átomos durante la transformación, usando como un ejemplo la transformación de la fase ordenada cúbica de cuerpo centrado (bcc) a cúbica de cara centrada (fcc) o a hexagonal (hex), como ocurre en varias aleaciones con base a los metales nobles cobre, plata y oro. Otros sistemas importantes son las aleaciones de hierro, a las cuales pertenecen los aceros, que tienen una transformación martensítica de la fase austenítica fcc a la martensita con estructura bcc o hex. Una condición necesaria para que la transformación martensítica tenga lugar es que aumente la estabilidad termodinámica. Se analizan en este trabajo los factores que determinan la estabilidad de las aleaciones de los metales nobles, usando como prototipo las aleaciones con orden de largo alcance de Cu-Zn-Al, que se están estudiando cuantitativamente en el Centro Atómico Bariloche. El orden de largo alcance también es necesario para obtener transformaciones con poca histéresis entre la transformación y la retransformación, lo que es la base de las aleaciones superelásticas y con memoria de forma. Se presentará la razón de este comportamiento.

Palabras clave: Transformación martensítica, aleaciones de metales nobles, aceros, aleaciones con memoria de forma, estabilidad de fases.

ABSTRACT

The martensitic transformation is a diffusionless first order phase transformation in the solid state, which proceeds by the nucleation and growth of the new phase. It is of importance to understand in which way the coordinated movement of atoms occurs during the transformation. This topic will be discussed for the transformation from the body centered long-range ordered structure to the face centered or hexagonal one, as is observed in many noble metal alloys. Also the transformation in the iron alloys from the face centered austenite to the body centered or hexagonal martensite will be dealt with. The transformation takes place only when the martensite becomes more stable thermodynamically. The factors that control the stability will be evaluated for the noble metal alloys, especially the Cu-Zn-Al alloys which have been studied extensively at the Centro Atómico Bariloche. Since the alloys possess long-range order, the influence of order on the phase stability is analyzed. As will be shown, long-range order is also decisive to obtain alloys with a small hysteresis between transformation and retransformation, which is the basis for the superelastic and shape memory behavior.

Keywords: Martensitic transformation, noble metal alloys, steels, shape memory effect, phase stability.

1 INTRODUCTION

The martensitic transformation is a diffusionless phase transition in the solid state with a large deviatoric component. What this means shall be illustrated by a simple two-dimensional sketch.
Consider in figure 1a a quadratic array of circles, representing the atoms. For some reason this array becomes unstable and distorts to the lattice shown in figure 1b. The distortion shown is large, but the area of the array can remain the same. It is a homogeneous distortion of the original lattice in which an atom does not change its position with respect to its neighbors, it only alters their distances. This is characteristic of diffusionless transformations with large shape changes. For these reasons the transformation can be called deviatoric.

Because the original lattice of figure 1 has quadratic symmetry, an equivalent distortion leads to c in the figure, the only difference being the orientation in space. Suppose now that the stability of the lattices depends on a thermodynamic variable, for example the temperature: At the higher temperature the square lattice may be stable but on cooling it flips over to the new structure at a critical temperature. But since, in our case, two different variants of the same structure are possible, the resulting configuration consists of a mixture of both.

This transformation can be followed by measuring the transformed fraction as a function of temperature on cooling, as done in figure 2.

In a large piece of lattice it is not possible to have all material transformed simultaneously, instead the new structure nucleates first locally in some region in the interior of the array and then grows. It starts at a temperature $M_0$, called the martensitic start temperature. Since the two variants b and c of figure 1 are equivalent energetically, they form with the same probability and start to present obstacles mutually for further growth. In order to continue growing the driving force has to be increased, which means a further cooling. For this reason the fraction that has transformed increases only with decreasing temperature, and is completed at the finish temperature $M_F$. On reheating the retransformation occurs, starting at $A_S$ and being completed at $A_F$. Generally there is a displacement with respect to the cooling curve. This hysteresis can be
quite large. Since no atom redistributions occur, diffusional processes that are time dependent are absent. Therefore the martensitic transformation is temperature, but not time dependent.

It seems reasonable to expect that the change from a to b in figure 1 can be aided by applying a force which helps to stretch the initial lattice in diagonal direction from the lower left to the upper right. This force can be applied at a temperature above $A_F$, at which the lattice does not yet transform spontaneously. But in this way only the variant b is favored, but not c. The behavior can be plotted as a force $F$ versus length change $\Delta l$ at a given temperature, as shown in figure 3.

![Figure 3](image)

**Figure 3**: Relation between length change $\Delta l$ and applied force $F$ at a temperature $T$ above $A_F$. When the transformation is complete the force increases rapidly again.

A minimum force is required before the transformation starts. But since only one variant is induced, the interference between the variants on cooling is absent and the transformation can go to completion at a practically constant force. On unloading, the original lattice is restored since it is the most stable one without load. For this reason it is called superelasticity. It is clear that the force necessary to transform the lattice increases with temperature, if the stability of the square lattice with respect to the transformed one increases with the deviation from the $M_S$ temperature.

This simple picture of figure 1 can easily be extended to real three-dimensional crystals. There is a large group of iron-based alloys which have a close-packed face centered (fcc) structure at elevated temperatures. This structure consists of close-packed planes in which each atom is surrounded by six neighbors, as shown in figure 4 (large open circles). The planes are stacked in such a way that the atoms on the adjacent plane lie in the holes formed by the first (larger filled circles), and those of the next plane lie in the holes on the first two (smaller filled circles). The atoms on the fourth plane lie then above those of the first one, and so on. This corresponds to a stacking $ABCABC...$ This atom array can also be described by the repetition of a cubic unit cell with atoms on the corners and on each face of the cube.

![Figure 4](image)

**Figure 4**: The stacking of close packed planes in an fcc lattice, marked by three different symbols, the large open circles, the larger and the smaller filled circles, corresponding to a stacking ABC. Also shown is the smallest vector $a$ between neighboring holes, and the shortest translation vector $b$. 
On cooling this structure becomes unstable and transforms to a body-centered cubic (bcc) lattice. The elementary bcc cell is a cube with the atoms on the corners and with an additional atom in the center. The transformation from fcc to bcc can be considered as a homogeneous distortion, like that in figure 1. This can be made clear by figure 5. In this figure are drawn two fcc cubic cells, and within them a smaller cell containing an atom in the center. This is a body centered tetragonal (bct) cell. By a homogeneous compression in the vertical direction, indicated by the arrows, and an expansion in the plane normal to it, a cubic bcc lattice is created without the need to change the volume of the bct cell.

![Figure 5: Two unit fcc cells with a smaller body centered tetragonal cell marked in the center (left). By a homogeneous compression in the direction of the arrows and an expansion in the plane normal to it the bcc structure (right) is obtained.](image)

Iron is the base of the steels that have found widespread applications during several thousand years due to their strength and hardness. This is due to the presence of the small carbon atoms. They find sufficient space in the interstitial sites between the iron atoms in the fcc structure and therefore dissolve easily in large quantities at high temperatures. When this fcc ‘austenitic’ lattice transforms martensitically at sufficiently high cooling speeds to the bcc structure, the space for the carbon atom is reduced, and the carbon, having no time to diffuse to more convenient sites produces a high degree of distortion around it which makes the alloy very hard. Homer in his Odyssey knew already very well that his sword had to be quenched rapidly in water to be hard. Often the experience can be made that a steel drill becomes black if one wants to drill a hole too fast and the drill heats up, leading to the diffusion and precipitation of the black carbon and the softening of the drill.

![Figure 6: Surface markings due to the formation of martensite plates in Fe-Ni single crystals. The width of the figure corresponds to 4mm of the sample.](image)

The martensitic transformation leaves also characteristic marks on the surface. In figure 6 is shown the surface contrast due to a partial martensitic transformation in an iron-nickel single crystal. The dark bands are the traces of martensite plates that have grown through the sample volume and have intersected with the surface leading to a surface upheaval. The long ones have formed first, and between them shorter ones have
appeared whose growth has been impeded by the long ones. A crystallographic analysis has shown that the martensite plates have very definite crystal orientations with respect to the original structure. These orientation relationships can nowadays well be accounted for by phenomenological theories, described first by Wechsler, Lieberman and Read, and Bowles and Mackenzie, discussed in the book by Nishiyama [6] and in the book edited by Otsuka and Wayman [7].

It has been observed that in some iron based alloys the hysteresis between the transformation and retransformation is small, 10 K or less, as in Fe-Pt. In other alloys, for example Fe-C, or Fe-Ni steels, it can amount to 400 K. The reason for this puzzle has not been adequately understood till now, and will be a subject of this paper.

There are also martensitic transformations in which the high temperature phase has the bcc structure, which transforms on cooling to fcc, or to a stacking variant thereof, like the 9R structure with the stacking sequence ABCBCACAB, or 2H with ABAB. To this group belong many noble metal alloys, based on copper, silver or gold [14-7], and among others also the nickel-titanium alloys [7]. Due to their compatibility with the human body the Ni-Ti alloys have found widespread medical applications. In this group of alloys the hysteresis between the transformation and retransformation is small, often below 5K in temperature. The most important properties are the superelasticity and the shape memory behavior.

The superelastic effect has already been illustrated by figure 3. By stressing in tension Cu-Zn based single crystals of adequate orientations maximum length changes of around 7% can be obtained, which disappear again on unloading. This can be compared with normal elasticity for which the elongation increases linearly with stress, until at much lower length changes irreversible plastic deformation occurs. The superelastic effect is used, for example, to straighten out irregularly grown teeth by fastening them with bent nickel-titanium wires. They exert a force that remains nearly constant during the movement of the teeth, corresponding to the plateau in figure 2. The stainless steel wires that are used as an alternative have to be changed frequently since the backdriving elastic force decreases with the teeth movement. There is now also a great deal of activity to develop antiseismic devices, which permit to absorb large displacements due to earthquakes, returning to their original shape when the seismic wave has passed.

The shape memory effect can be understood with reference to figure 1. On cooling without a stress all possible variants are formed, as also seen in figure 6, leading to a zero net shape change. If in this state a force is applied, the most favorable variant will grow at the expense of the other less favored ones, provided the interfaces between the different variants are mobile. This finally can lead to the same configuration as that which is obtained when the force is applied above $M_S$. The variant growth is associated with a shape change. On heating, this deformed material transforms back to the original shape of the high temperature phase. Thus the crystal has remembered its shape even after deformation in the martensitic state. The technological and medical applications are numerous, and have been described in the literature, see the articles in [7]. For example, two tubes can be connected by a ring of this material that previously has been expanded at lower temperatures in the martensitic state, and then is warmed up after having been placed over the two tube ends.

Although there are by now many technological and medical applications, and more and more possibilities are sought after, there is still a considerable lack of understanding of the basic processes that take place during and after the transformation. Our research in the Centro Atómico Bariloche is devoted to a better understanding of these basics. In the following, several investigations shall be described which have helped to clarify some phenomena related to the martensitic transformation. Three questions will be mainly discussed. How is it possible that the transformation involving this big homogeneous distortion, as described in figure 5, can take place in the rather rigid surroundings of the matrix lattice? What factors determine the large differences in the degree of hysteresis and lead to the shape memory behavior? What is the thermodynamic driving force for the transformation? It is clear that the reason for these features has to be understood, if materials are to be developed with the desired properties.

2 CRYSTAL PLASTICITY AND THE MARTENSITIC TRANSFORMATION IN Cu AND Fe-Mn ALLOYS

The martensitic transformation leads to a shape change. Shape changes occur also when we deform a common metal, which can even be done in the proximity of the absolute zero temperature. Therefore it is convenient to start the discussion of the transformation mechanism by analyzing the quantities that are responsible for plasticity. In the simplest case the deformation occurs by the displacement of atom planes with respect to their neighbors, as shown schematically in figure 7.
The planes that are favored are those with the largest interplanar distances, since they produce the lowest distortion during the movement. This displacement does not occur simultaneously on the whole plane. Instead it proceeds by the propagation of a local disturbance, which is called a dislocation. For illustration an “edge dislocation” in a cubic primitive lattice is shown in figure 8.

An extra plane is inserted in the lower half, by its end a line can be defined, the dislocation. This extra plane can move to the right or left simply by displacing the upper half of the neighboring plane unto a position above the extra plane. The continuation of this process can thus be described as the movement of a dislocation on the horizontal glide plane, producing a displacement that restores the lattice. This displacement is called the Burgers vector. It is in principle possible to introduce two neighboring extra planes instead of one. Their movement leads also to the restitution of the crystal lattice. It is clear however, that such a configuration produces a large distortion of the surrounding lattice, and therefore is not favored energetically compared to the single extra plane. It can be said that the deformation by the movement of dislocations is characterized by three quantities, the direction of the dislocation line, the glide plane and the Burgers vector with the smallest possible translation vector.

In the fcc structure the generally observed glide plane is a close packed \{111\}_\text{fcc} plane, on which each atom has six nearest neighbors, as shown in figure 4. It is also schematically indicated by the hexagon in figure 7. The shortest Burgers vectors are those between nearest neighbors, given by \((a_{\text{fcc}}/2)<110>_{\text{fcc}}\) in the usual nomenclature, with \(a_{\text{fcc}}\) being the cube distance of the fcc unit cell. In figure 4 has been marked a possible Burgers vector by the double arrow between two neighboring filled circles, denoted by b. In this figure is also indicated a shorter arrow a, between the two different filled circles. It corresponds to a displacement of an atom into the neighboring hole formed by the three atoms on the plane below. Since the arrow a is shorter than b it could be argued that a dislocation with Burgers vector a is favored over that with b.
because of the smaller distortion it produces around its core. However, after moving such a “partial” dislocation on a close packed plane through the lattice, the stacking order has been changed. Instead of ABCABC the stacking is now, after the movement across the second plane, ABABABC. But the fcc structure is the most stable equilibrium phase, then displacements with the smaller Burgers vector lead to an increase in energy and therefore are not favored compared to displacements with the larger Burgers vector which conserves the original stacking.

There are, however, metals and alloys whose relative stability between fcc and hex changes with temperature. Cobalt is the simplest metal that is fcc at high temperatures and hex below \( T_0 = 700 \) K. This means that the fcc structure can transform to hexagonal on cooling through \( T_0 \) provided the appropriate dislocations are available on each second plane. On reheating, the Co retransforms back to the fcc structure. This is a very simple example of a martensitic transformation. The homogeneous distortion is due to the shear in the same direction on each second basal plane. The dislocations are highly mobile since the friction energy during their movement is small, so once \( T_0 \) is reached the dislocations can move, independent of the cooling velocity at \( T_0 \). Therefore the hysteresis between transformation and retransformation can be expected to be small, and superelasticity and the shape memory effect should be present. Unfortunately the \( T_0 \) is so high in Co, that these effects are of little utility.

A great deal of work is now in progress in which it is attempted to develop alloys with this fcc/hex martensitic transformation. One of the main problems with the martensitic transformation is that diffusion controlled processes take over at elevated temperatures. The thermal energy is then sufficiently high that individual atoms can jump via vacant sites, the vacancies. These processes depend on time in contrast to the martensitic transformation that proceeds to completion once the driving force is favorable. In order to eliminate time dependent aging effects, it is therefore necessary to have the martensitic transformation at sufficiently low temperatures at which diffusion cannot take place yet. Alloys with a high melting point like the Fe-alloys are a good choice. There are indeed the Fe-Mn based binary and ternary alloys, which show this transformation from fcc to hexagonal, and which are now being studied worldwide, among other places also at the Centro Atómico Bariloche.

3 THE MARTENSITIC TRANSFORMATION IN THE NOBLE METAL ALLOYS

The transformation from fcc to hexagonal by the movement of partial dislocations is the simplest type of martensitic transformation with a large shape change. The transformation from bcc to fcc cannot be understood in this way, although the former can serve as a basis for the required generalization. Since most of the studies at the Centro Atómico Bariloche have been performed with the ternary Cu-Zn-Al alloys, this system will serve as a prototype in what follows.

The alloys based on the noble metals Cu, Ag and Au belong to the group of alloys that are often called the Hume-Rothery alloys [4]. Depending on composition they form different equilibrium phases, which are mainly controlled by a very simple parameter, the electron concentration \( e/a \). It is the average number of conduction electrons per atom, taking 1 for Cu, 2 for Zn and 3 for Al. The bcc structure, denoted by \( \beta \), is stable at \( e/a \) around 1.5. At the highest temperatures the different atom species are distributed at random on the sites of the bcc lattice. On cooling a long-range ordering to a B2 structure starts at a critical temperature \( T_B2 \). This ordering means that the occupation probability of the atom species is different on the corner sites of the elementary cube and its periodic repetition from that in the centers. When B2 order is perfect in Cu-Zn-Al alloys with a Cu concentration above 50%, the sites at the corners are occupied by Cu atoms, whereas the center sites are randomly occupied by the rest. (An interchange of all corner with center sites leads to the same configuration). A second type of ordering often is observed in the Cu-Zn-Al alloys, in which ordering between the center sites takes place. Center sites occupied by Al atoms alternate with neighboring center sites occupied by the excess Cu atoms that do not enter into the corner sites. The Zn atoms fill up the remaining center sites. This L21 type of ordering is not important for the present discussion, and will be neglected here, considering only the B2 ordered configuration.

This structure starts to transform martensitically during cooling at a temperature \( M_s \), which depends strongly on composition. Its progression has been followed by observations in the transmission electron microscope [8].
Figure 9: Observation by transmission electron microscopy of the growth of thin martensite plates (upper part) and their ensuing thickening (lower part) in a Cu-Zn-Al single crystal.

In figure 9, upper left, appears a thin martensite plate that extends through the field of observation, upper right, together with the growth of two other plates. Subsequently the plates start to thicken by the lateral movement of the planar interface in the lower part of the figure. The growth of the thin plates looks very similar to that from fcc to hexagonal, except that the interface is no longer a simple close packed glide plane, and the martensite has the 9R ABCBCACAB structure. In the lower part of the figure in one of the two plates the projection of the basal plane is in contrast. The complete transformation from the β matrix to the martensite occurs already at the tip, since the martensite structure is fully developed right behind it. What are then the atom movements that occur at the tip? This question will now be treated.

A common shear plane in the bcc structure is the $\{110\}_\beta$ plane. Its orientation is marked within a unit cell in figure 10a. In the B2 ordered lattice the occupation probabilities of the different atom species are different in the center and the corners of the unit cell, marked in the figure by the open and filled circles. The same $\{110\}_\beta$ plane is shown in 10b by the round symbols. The square symbols lie on the plane above. In the bcc structure the atoms, like A and D, lie in the middle of two atoms below.

Figure 10: The transformation from a B2 ordered bcc to an fcc martensite. The original B2 lattice 10a is sheared on the inclined plane leading to the displacement of atoms (square symbols on the plane above that of the round symbols in 10b). For example, A and D move to A' and D'. The shuffle consists of a displacement on an inclined plane restoring the correct stacking of a close packed structure, by moving D' to D'' or D''' and all other atoms on this plane correspondingly (10c). This sequence is illustrated in figure 11.
Figure 11: The transformation from B2 to fcc is illustrated by a hard sphere model. In 11a three layers of the plane corresponding to that in figure 10a and b are shown. By the triangle is marked an inclined plane of the same family as the horizontal one. The shear in the same direction on consecutive planes (as in figure 10b) leads to the distortion shown in 11b. The shuffles on the triangular plane (figure 10c, in the same direction on subsequent planes) leads to the fcc structure (11c).

If we want to make a model of hard spheres, we will find it difficult to keep the atoms there, because they would prefer to move into a position in the center of three atoms below, for example A into A’ and D into D’ in figure 10b. Try to stack three layers of the bcc arrangement as in figure 11a, and you will note that the consecutive planes try to get displaced, all in the same direction, or neighboring ones in opposite directions. In figure 11b is shown the result when all planes are sheared in the same direction (as if looking from top down in figure 10b). If the shear occurs in opposite directions on neighboring planes, then the average shape of the sample is not changed.

If we were permitted to approximate the bcc solid by an arrangement of hard spheres, we would conclude that the bcc structure cannot be stable. That this is not observed is mainly due to another factor, which generally can be as important as the energy balance, namely the entropy. At finite temperatures the atoms vibrate around their equilibrium positions. The more space they have to vibrate, the better. It is clear that the atoms can vibrate more easily at sites corresponding to the bcc structure, like A, than after having fallen into the hole between the three atoms at A’. In other words, the probability to find an atom at a given point in space at a given moment in time is the smaller the larger the vibration amplitude is, i.e. the larger the spacial uncertainty. This is just the vibrational entropy. Thus the stability of the bcc structure is controlled by two opposite forces. Those due to the entropy which would like to have the atoms at positions where they can vibrate easily, and those due to the energy by which the atoms would prefer to have a larger number of neighbors with favorable bonds. At high temperatures the atoms have a high thermal energy, favoring the bcc structure which permits the higher vibration amplitude. But on decreasing the temperature this contribution decreases and the energetically favorable shear takes over, leading to a new structure.

A collapse to a structure in which the shear takes place in opposite directions on neighboring planes has indeed been observed in the form of small precipitates during the dezincification of Cu-Zn at elevated temperatures \[12\]. But such a structure has never been found to form martensitically when on cooling to low temperatures the β phase becomes unstable. The reason is that any pure shear structure is not a low energy equilibrium phase, and second that it is very easy to get to the fcc structure (or the modifications thereof) by a simple additional ‘shuffle’. This can be seen as follows \[1\].

In figure 11a a few spheres have been taken out at the left corner to expose an inclined plane, marked by the triangle. This is a (011)β plane that is transformed to a close packed plane after the horizontal shear. This close packed plane is what is needed in the fcc structure. However, the stacking of subsequent planes is not yet correct. But this can be remedied quite easily. In figure 10c are projected the atoms on the inclined triangular plane as round symbols. The neighboring plane above is presented only by one atom, denoted by D’. It can be seen that this position is not the correct one for an fcc lattice, for which the correct sites are at D” or D’’. But a small displacement of the whole plane, corresponding to the displacement of D’ to D” or double the distance to D’’ produces the correct sequence for a close packed lattice. Since the displacements are small they will be called ‘shuffles’. If the shuffles occur in the same direction on each consecutive inclined plane, an fcc lattice is created, as shown in figure 11c. If two short shuffles on consecutive planes are compensated by one large one in the opposite direction, and this sequence is repeated, then the average shuffle becomes zero, and the resulting structure is 9R with the ABCBCACAB sequence.

One additional condition has to be complied with, namely that the interface between the martensite plate and the matrix should not be distorted on the average, since this would introduce additional distortion energy which would make the transformation less favorable. This produces a restriction on the distribution of
the shuffles in the two directions, which on the average has to be close to zero. The most homogeneous
distribution is that leading to 9R. The fcc lattice can be obtained, if on a certain number of planes the shuffle
occurs in one direction, followed by a sequence in the opposite, the twin direction. In order to obtain the
hexagonal stacking the shuffle directions have to alternate, but this does not lead to a zero average. Therefore
if the hexagonal 2H structure is favored energetically, it becomes necessary to activate another shear system
which permits to comply with an undistorted interface.

It should have become clear why the β phase becomes unstable and how it can transform to an fcc
lattice by a combination of a shear and an adjustment shuffle at the tip of a growing martensite plate. It
replaces the simple movement of a stack of partial dislocations, as in the case of the transformation from fcc
to hexagonal.

An important point to be discussed now is why the hysteresis is found to be small in so many noble
metal alloys leading to the shape memory behavior and to the superelastic effects. The reason can be stated
very simply: In order that after a transformation-retransformation cycle the original undistorted lattice is
restored, the atoms have to move on the same single path during the transformation and back during the
retransformation. It can be illustrated by figure 10c. During the transformation the shortest distances are from
D’ to D” or D’”, whatever the atom distribution on the lattice sites. But if no order were present, on
retransformation atom D’’” has no need to move back to D’. It could as well move in the direction towards
the other two neighboring triangles, and this means that the original lattice is not restored. But in the presence
of order the three different paths are no longer equivalent energetically, and the only path that does not
change the long-range order is the path back to D’.

It is clear that the order plays an important role, and it is suggested that the shape memory effect be
closely related to the reversibility of the transformation-retransformation path. This argument holds also
when the large hysteresis effects in the disordered Fe alloys are compared with the small one in the long-
range ordered Fe-Pt alloys. Here also the order reduces the path multiplicity to one [1]. Therefore a means to
obtain shape memory alloys is to have a long-range ordered high temperature phase. Long-range order is
important not only for pseudoelasticity, it influences also the relative stability of the martensite with respect
to the β phase, and thus controls the Ms temperatures. Furthermore it affects the ease with which the
transformation shear and shuffle can proceed. These questions will therefore be addressed briefly in the
following two chapters.

4 THE INFLUENCE OF LONG-RANGE ORDER ON PHASE STABILITY

This question has been discussed in detail elsewhere [2], and therefore only a few aspects will be
mentioned here, that are also relevant for the description of long-range order in general.

A solid is stable if the bonding between the atoms is favorable. Often the total bonding energy is
described as the sum of interaction energies between all atom pairs. For a binary alloy of atom species A and
B, pair interaction energies \( V_{\text{AA}}(r_i) \), \( V_{\text{BB}}(r_i) \) and \( V_{\text{AB}}(r_i) \) between A-A, B-B and A-B atom pairs, respectively,
can be defined, which generally depend on the crystal structure and on the distance \( r_i \) between the pairs. They
may also depend on their orientation within the lattice, but this complication will be neglected, since it is of
no importance for the noble metal alloys to be discussed here. In the transition metal alloys this may be
different, since bonding is mainly due to the highly directional bonds from the d-electrons.

If the interaction energies between the different pairs A-A, B-B and A-B are the same for each \( r_i \)
then all atom distributions on the sites of a given lattice, for example bcc, have the same energy. This means
that a large bcc crystal with one half consisting only of A atoms, and the other half only of B would have the
same energy as one in which both species are mixed completely. It is however very unlikely that the former
configuration is found. This is also consistent with our daily experience, when we put in a bag and shake a
large number of spheres that are identical except for their color. We will find a complete mixture instead of
the bag, the positional uncertainty is highest. This uncertainty can be expressed quantitatively by the
configurational entropy. Analogously, in the case of the bcc lattice the most likely configuration is the
random distribution of A and B atoms on the sites of the bcc lattice if the interaction energies are the same.
This configuration will be called disordered in the following.

Very rarely the interaction energies are the same for all pairs. Instead it can be less or more
favorable to have A-B pairs instead of A-A and B-B pairs at distance \( r_i \). This tendency can be expressed by the
“pair interchange energy”:

\[
W_{\text{AB}}^{(i)} = V_{\text{AA}}(r_i) + V_{\text{BB}}(r_i) - 2V_{\text{AB}}(r_i)
\]  

(1)

It is the energy difference corresponding to four atoms, two A and two B atoms, between forming an
A-A and B-B or 2 A-B pairs at positions separated by \( r_i \). Thus \( W_{\text{AB}}^{(i)} \) is zero, if all pairs have the same


energy, and is positive, if A-B bonds are favored. The total interaction energy is obtained by summing over all pairs of a given atom distribution in a given lattice. If all pair interchange energies are known then it is a simple matter to do the summation for any configuration.

The question has now to be solved which configuration is the most favorable one, but this is not simple, because two opposing tendencies are present when $W_{AB}^{(ij)}$ is positive. On the one hand, ordering is favorable energetically, and the largest possible number of A-B pairs would be preferred. This means, on the other hand, that the configurational entropy is decreased, since the ways are reduced in which the atoms can be arranged. The quantity that characterizes the configurations is therefore a combination of the energy $E$ and of the entropy $S$, called the Gibbs free energy $G$ which is given by:

$$G = E - T S$$

Where $E$ is the sum of the interchange energies, $T$ the temperature and $S$ the configurational entropy. $E$ and $S$ depend on the atom distribution. The most favorable configuration is that with the lowest Gibbs free energy. Thus at the lowest temperatures order is preferred, but with increasing temperature the $TS$ term plays an ever increasing role. More and more disorder is favored, until at a critical temperature the long-range order breaks down, which means that on the average the probability to find an A atom is the same on each lattice site. There may however remain some short-range order in which locally an A atom has a different number of B atoms at distance $r_i$ than expected from a completely random distribution.

The configurational entropy can be calculated for any configuration by counting the multiplicity of the different atom distributions, as singles, as pairs, and up to atom clusters of the required sizes. Attempts have been made to calculate also the pair interchange energies, but till now the precision is not sufficient to evaluate quantitatively the more distant pair contributions. It is therefore necessary to take recourse to experimentally measured quantities. The most direct experimental method is their determination from short-range ordering by X-ray or neutron diffraction. Many alloy systems have been studied in this way, and a competent summary can be found in [11]. As an example the results for the fcc primary solid solution of fcc Cu-Zn alloys shall be presented, since it gives some important clues to the interaction between atoms in a solid.

The fcc $\alpha$ solid solution is stable from pure Cu to approximately Cu-38at%Zn. It does not show long-range ordering since the critical ordering temperatures are below those at which atom reordering by diffusion can take place. The main reason for the small ordering tendency is the so-called “frustration effect”. In the fcc structure some of the first nearest neighbors of a given atom are simultaneously also nearest neighbors among themselves. It is therefore not possible to have only favorable first neighbor pairs at maximum order, in contrast to the B2 lattice where the first neighbors of an atom are second neighbors among themselves with lower pair interchange energies.

**Figure 12:** The pair interchange energy as a function of pair distance (in units of the bcc lattice constant), as determined experimentally from short-range ordered fcc Cu-31.1at%Zn [10].

Pair interchange energies have been obtained from short-range ordered Cu – 31.1at% Zn by Reinhard et al. [10]. They are replotted in figure 12 as a function of pair distance. Two features are immediately seen. The first and second neighbor pair interchange energies are larger than the more distant
ones, but the contributions up to 20th neighbors are not negligible. It has often been concluded that the
contributions beyond second neighbors can be neglected, since they are small. This is not justified, however,
since the number of atoms in a shell of constant thickness at distance \( r \) increases on the average quadratically
with distance, and therefore the total contribution from the pairs may not be negligible at more distant shells.

It seems surprising that the influence of a given atom species is still felt at such large pair distances.
This has been attributed to the incomplete screening of the ions by the conduction electrons, resulting in the
so-called “Friedel oscillations”. The parameters that control these oscillations are the electron concentration
\( e/a \), the atomic volume and the atom species involved. There is a rather large uncertainty in the
experimentally determined values of the small pair interchange energies at large distances. They are not
sufficiently precise to calculate reliable order energies that are required for a quantitative evaluation of the
martensitic transformation. There is an alternative, based on more general arguments, which is considered to
present a good approximation [2]. At a sufficiently large distance from a given atom the discrete atom
distribution can be replaced by a continuum. Ordering in the disordered lattice involves only atom
redistributions on an atomic scale. It can be expected that the number of A and B atoms on a thin shell at
large distance \( r \) is not changed by ordering. Therefore ordering should not affect the energy contribution from
the more distant pairs, provided that changes in \( e/a \), in atomic volume and in the density of species A and B
on the thin shell at large \( r \) are absent. To a good approximation it should thus be possible to calculate the
order energy including only the first few neighbor pairs. It has been shown indeed that the order energy can
already be well accounted for by the large first and second neighbor pair interchange energies, which are
available from experiment and theory [2].

The martensitic transformation is a low temperature transformation, because at more elevated
temperatures diffusional processes start to take place. Long-range B2 order is stable below the critical order
temperature \( T_{B2} \) around 740 to 800 K in the Cu-Zn and Cu-Zn-Al alloys [2]. Far below \( T_{B2} \) the long-range
order is perfect. By an adequate heat treatment of the \( \beta \) phase at the low temperatures prior to the
transformation it is therefore possible to obtain the nearly perfect long-range order. In this case the
contribution from the configurational entropy can be neglected (the last term in equation 2). This simplifies
considerably the evaluation of the martensitic transformation in the Cu-Zn based alloys. The atom
distribution in the perfectly ordered B2 structure is thus well known. It implies also that the atom positions
are known for the fcc martensite after the diffusionless transformation. It is a simple matter to calculate then
the order contribution to the martensitic transformation by summing over the pair interchange energies from
the first few pairs of the two phases. By adding this to the energy difference between the disordered fcc and
bcc phases which are stable at elevated temperatures, it is possible to determine the total energy difference
involved in the martensitic transformation. The equilibrium temperature between the two phases, only
slightly different from \( T_{B2} \) follows then from measured vibrational entropy changes.

Of course, it is necessary to have available the pair interchange energies from the two phases
involved in the martensitic transformation. For binary Cu-Zn the values from the measurements for Cu-
31.1at%Zn can be used for the fcc martensite at about 40at%Zn, since the calculations have shown that the
pair interchange energies vary only little with composition [13]. For the B2 phase they are not measured but
have to be deduced. Here a very general approximation for nearly free electrons can be made, namely that
the pair interchange energies depend on pair distance but are independent of the lattice structure. This means
that a single continuous curve is sufficient to describe the distance dependence of the pair interchange
energy. The values for the different structures are simply those at the corresponding pair distances. A
theoretical and an experimental argument can justify this.

According to pseudopotential theory the nearly free electron solid consists of atom cores embedded
in the sea of conduction electrons. The core consists of the nucleus and the strongly bound inner core
electrons and is much smaller than the distances between the atoms. The electron distribution is modified by
the interaction with the core. It leads to the effective interaction between the atoms and depends solely on the
pair distance.

The relative stability of the equilibrium \( \alpha \) and \( \beta \) phases depends only on the electron concentration
\( e/a \) in many noble metal alloys. This seems to be surprising since it implies that the large mixing energies that
are measured when the pure elements are mixed together, are the same in both phases, at least the part that
depends on the atom arrangement and is described by the pair interchange energies. Since both phases have
the same \( e/a \), the same atomic volume and the same composition it can be expected, according to the
arguments presented above, that the possibly large contribution from the more distant neighbor pairs is the
same in both phases. This means that also the contribution from the first few neighbors is the same. It has
indeed been shown that the contribution from the 12 first and 6 second neighbor pairs in fcc is equal to that
from the 8 first and 6 second neighbors in bcc. This result is consistent with the structure independence of the
pair interchange energies and with the measured and theoretically deduced quantities.

The same concepts have been applied to the ternary Cu-Zn-Al alloys. In these alloys three different
types of pairs have to be evaluated, in addition to Cu-Zn also Cu-Al and Al-Zn. Since Al-Zn shows little

tendency for order its contribution has been neglected. From short-range order in binary $\alpha$ phase Cu-Al the first and second neighbor pair interchange energies were determined experimentally for the fcc structure. A continuous curve was drawn through the data, and the pair interchange energies for the bcc phase was then interpolated similar to the procedure for the Cu-Zn alloys. It has been shown that in this way the martensitic transformation can be described quantitatively in a wide composition range between Cu-Zn, Cu-Zn-Al and Cu-Al alloys [2].

Thus, the order energy of a given phase, bcc or fcc, can well be approximated by the contributions from the first few neighbor pairs only, without including the more distant pairs that are not known with sufficient precision from calculations or from experiment. This is different if it is attempted to calculate the configurational mixing energies of one phase alone. In this case the alloy is created from the pure elements by large atom redistributions which no longer can be described by the small-scale atom interchanges found during ordering. It has been found, indeed, that the mixing energy is by a factor two too small compared to the experimental values if the contributions from the pairs beyond first and second neighbors are neglected.

Although the martensitic transformation is diffusionless, diffusion processes may take place after the transformation is complete although they are not permitted during the transformation. This new degree of freedom can lead to a reduction in Gibbs free energy. It has been found to occur in the Cu-Zn-Al alloys already at room temperature, and has been called stabilization since it leads to an increase in the retransformation temperature to the $\beta$ phase. This effect is often not desired since it modifies the martensite in an uncontrolled, time dependent manner, which limits the use in technological applications. It depends on two factors, namely the driving force for the redistribution of the atoms, i.e. the change in Gibbs free energy, and the vacancy concentration that makes diffusional atom changes possible. The description of this time dependent behavior involves the determination of changes also in configurational entropy, and therefore is more elaborate, since it implies the determination of the Gibbs free energy according to equation 2. This problem has been discussed elsewhere, and will not be treated here [3].

5 THE CRYSTALLOGRAPHY OF THE MARTENSITIC TRANSFORMATION AND THE SELECTION OF THE SHEAR AND SHUFFLE SYSTEMS

As exposed above, the atomistic description of the martensitic transformation involves in general the combination of a shear and a shuffle. For the transformation from bcc to fcc a combination of displacements on two different \( \{110\}_b \) planes has been used. The question that remains to be answered is why these shear systems have been used, since appropriate combinations of other shear systems can lead to the same final martensitic structure. It is reasonable to expect that those shear and shuffle systems should be activated which are easiest and lead to the smallest unfavorable distortion of the lattice. From the study of the plastic deformation the most favorable shear planes for dislocation slip in the different crystal structures are known. This information is also useful for the evaluation of the shear and shuffle systems during the martensitic transformation.

In the fcc lattice, the common glide plane is the close packed \( \{111\}_f \) plane. The same close packed plane is also observed for the hexagonal lattice, denoted \( (001)_{h} \), when the interplanar distance is sufficiently high, expressed by \( c/a \), where \( c \) is the unit distance between the \( (001)_{h} \) planes and \( a \) is the interatomic distance on this plane. Examples are Zn and Cd with \( c/a > 1.85 \). When \( c/a \) is too small, slip on \( (001)_{h} \) becomes less favored compared to that on the prismatic \( \{10\overline{1}0\}_h \) plane normal to \( (001)_{h} \). An example is Ti with \( c/a = 1.587 \). For the bcc lattice the commonly observed slip planes are \( \{110\}_b \) and \( \{112\}_b \).

A measure of the ease of shear is also the energy change for small shear displacements, which are expressed by the combination of elastic constants. In the cubic crystals three elastic constants are sufficient to describe elasticity, namely \( C_{11}, C_{12} \) and \( C_{44} \). \( C_{44} \) accounts for the shear displacement on the cubic \( \{100\} \) plane in the orthogonal \( \{001\} \) direction. In figure 10a it means the small displacement of the upper horizontal plane with respect to the lower one in horizontal direction parallel to one cubic axis. Generalizing this, any displacement which contains \( <100> \) as the shear direction or as the shear plane normal has the same \( C_{44} \) in cubic crystals. Similarly, the combination \( C' = (C_{11} - C_{12})/2 \) describes the shear displacement on any of the family of \( \{110\} \) planes in the \( <1\overline{1}0> \) direction contained in it. For bcc the \( (110) \) plane lies in the direction of the face diagonal, as marked in figure 10a. Another important combination is \( (2C' + C_{44})/3 \). It describes the shear displacement on a cubic \( \{111\} \) plane, which is the close packed plane in the fcc structure. It also accounts for any shear that contains \( <111> \) as shear direction or shear plane normal.

We are now in a position to start rationalizing the observed martensite crystallographies for transformations from bcc to fcc, bcc to hex and fcc to bcc. As mentioned, in a hard sphere model the bcc structure is not stable, this would mean a negative \( C' \). In the noble metal alloys and many other bcc lattices...
the C’ is positive, though small. Whereas normally C’ increases with decreasing temperature, it decreases when approaching a martensitic transformation from higher temperatures. The ratio C_{44}/C’ is of the order of 10 for brasses, showing that the other possible shear systems have much higher combinations of elastic constants. When the transformation to the fcc lattice is complete these {110} planes have transformed into close packed planes which are the common shear planes in the fcc structure. It can be concluded therefore that the complete transformation path combining the {110} type shear with the shuffle indeed is the favored transformation path. This justifies the transformation model presented above. It is also thought to hold when the transformation proceeds to the 2H martensite with the stacking ABAB. In this case, however, the shuffle cannot satisfy the condition of an undistorted interface, as mentioned before. An additional displacement shear is necessary which indeed is observed. The martensite plate therefore propagates with the creation of stress fields around the tip until the stress is sufficiently high to activate an auxiliary “twin” shear in the 2H region that has already transformed. In this way the stress is relaxed and the growth of 2H can continue, until again a high stress is created. This procedure is repeated continually. It requires additional energy that is dissipated and leads to an increase in hysteresis, when compared with the transformation to the 9R or fcc martensite.

Not all bcc structures that transform martensitically have such a high C_{44}/C’ ratio. In fact there are bcc alloys containing transition elements that have a C’ which is not much smaller than C_{44}. For example, in Ni-Ti alloys the ratio amounts to only two. In Ti a factor of 3 to 4 has been estimated, although due to the high transformation temperature no measurements have been possible. Therefore the different shear systems produce similar elastic distortion energies. Additional arguments are necessary in this case to rationalize the selection of the shear and shuffle systems for the transformation from bcc to 2H or hex. They can be based on the observation that the prismatic {101} BC plane is the common slip plane in hexagonal Ti. This plane corresponds to a {112}BC plane in the fcc lattice prior to the transformation, which is also a common slip plane in bcc. Therefore the transformation from bcc to hex can be considered to proceed by a shear on a (112)BC plane normal to a (110)BC plane. The former is transformed to the prismatic plane, whereas the latter becomes the basal plane of the hexagonal structure after shuffles in opposite directions on neighboring planes has taken place. When the interplanar distance between the (110)BC planes remains unaltered after the transformation to the basal planes, the transformation is completed. Examples are some disordered Ti-Ta alloys and long-range ordered Ni-Ti-Cu. If the interplanar distances change an additional twin shear has to be activated similar to that described for the noble metal alloys. This occurs in the Ni-Ti alloys and leads to an increase in the dissipated energy and consequently to a higher hysteresis compared to Ni-Ti-Cu. It should be mentioned that this mechanism has been proposed already a long time ago by Burgers [5].

In the iron alloys two different martensites have been found. They are characterized by the crystallographic orientation of the interface of the martensite plate with the surrounding matrix, called austenite. One is the (259) martensite whose crystallography can well be accounted for by the phenomenological theories, and the (225) martensite that has presented problems. Why this is the case and why there are different martensites has been an unanswered question till now. It can be well accounted for by the present shear and shuffle model. Since the basal {111}FC planes are common slip planes in the fcc structure, and are also activated for the transformation from fcc to hex, it is straightforward to consider them also as the preferred planes for the transformation shear. It seems reasonable to expect that the shuffle plane is also a plane of the {111}FC type, inclined to the former. It has in fact been shown that it is possible to obtain a martensitic transformation with an undistorted interface. But unfortunately the predicted crystallography is in disagreement with the observations. The interface deviates slightly but clearly from (225). The reason is the following. The shear associated with this transformation is large. In these alloys the elastic constants are high and this means a high distortion energy in the austenite ahead of the growing martensite plate. It would therefore be convenient to reduce the distortion energy by decreasing the transformation shear. This can be done by deforming the martensite by an additional twin shear. Such a twinning has indeed been observed. That twin system will be activated which reduces most favorably the transformation shear.

The austenites that transform to the (259) martensite are ferromagnetic. The ferromagnetism has a very important consequence, namely that it decreases the C’ elastic constant to such an extent that the most favorable shuffle plane becomes the {110}FC instead of the {111}FC plane. With this shuffle plane it is indeed possible to rationalize the formation of the (259) martensite. The low C’ has an additional effect. It reduces the elastic energy in the austenite ahead of the growing martensite plate. Therefore it is not necessary to activate an additional twin shear within the martensite. The combination of the shear on the basal plane and the shuffle on the {110}FC plane is sufficient and leads to the correct crystallography consistent also with the results of the phenomenological theories.
6 SUMMARY AND CONCLUSIONS

Three aspects of the martensitic transformation, which hitherto have not been well understood, are discussed in this paper.

a.) The phenomenological models have been quite successful in describing the crystallography of the martensitic transformation. They do not pretend to account for the actual atom movements during the martensitic transformation. It has been shown in this paper that the atom displacements can proceed by a combination of a long-wave shear and a shuffle, leading to the same predictions as the phenomenological theories. Those shear and shuffle systems that require the least distortion are the most favorable ones.

b.) The stability of the martensite with respect to the high temperature $\beta$ phase in the noble metal alloys is controlled by two contributions. An electronic one due to the conduction electrons described by the electron concentration $e/a$, and a configurational one given in terms of pair interchange energies between the atom pairs. Due to the long-range ordering, the martensitic transformation temperature is lower than that between the corresponding equilibrium phases at elevated temperatures.

c.) Many alloys that transform martensitically show the shape memory and superelastic behavior. It occurs if the atoms are allowed to move back during retransformation solely on the same path they had taken on transformation. In the absence of long-range order generally a multiplicity of transformation paths exist, which lead to the retention of defects after a cycle, and thus to an increase in hysteresis. In the presence of long-range order most of these paths would lead to an unfavorable increase in energy and therefore are prohibited, thus retaining only a single transformation-retransformation path, as occurs for the transformation in the Cu-Zn-Al alloys, and in the long-range ordered Fe-Pt alloys.

7 REFERENCES