Chapter 2
Basic concepts

2.1 Dislocations and grain boundaries

2.1.1 Dislocations

This section concentrates on several relevant aspects of dislocations and the differences in dislocation behavior between face-centered cubic (fcc) and body-centered (bcc) crystals, the two main crystallographic systems studied in this thesis. For a more extensive introduction to the theory of dislocations, reference is made to Hirth and Lothe [1].

The geometry of a dislocation is characterized by the Burgers vector, which defines the displacement of the lattice around the line defect [2]. The Burgers vector is not a vector in the usual mathematical sense, because its sign depends on the line sense of the dislocation and thereby on the orientation of the coordinate system. More precisely, any function $F$ operating on the Burgers vector $\mathbf{b}$ and the line direction $\xi$ must satisfy the condition $F(\mathbf{b},\xi) = F(-\mathbf{b},-\xi)$ and $F(\mathbf{b},\xi) \neq F(-\mathbf{b},\xi)$; $F(\mathbf{b},\xi) \neq F(\mathbf{b},-\xi)$. Two types of dislocations can be distinguished: edge dislocations, which have their Burgers vector perpendicular to the line direction, and screw dislocations, whose Burgers vector is parallel to the line direction. In general, a dislocation may have a mixed edge and screw character so that its Burgers vector is at an arbitrary angle with respect to the dislocation line (see Figure 2.1).

The elastic strain energy $E$ stored per unit length $L$ of a dislocation line is given by

$$\frac{E}{L} = C \mu b^2$$

(2.1)

where $C$ is a constant, which depends on the core structure of the dislocation, $\mu$ is the shear modulus and $b$ is the magnitude of the Burgers vector. Accordingly, dislocations with a small Burgers vector are energetically favored to those with a larger Burgers vector. For a perfect dislocation, the Burgers vector is equal to a translation vector of the lattice; in practice, it is usually equal to the smallest translation vector due to its low energy. Dislocations are consequently of the
½ <110> type in face-centered cubic (fcc) crystals and ½ <111> in body-centered cubic (bcc) crystals. Perfect dislocations may dissociate into partial dislocations, but the extent to which splitting occurs depends strongly on the stacking fault energy of the material. All metal systems studied in the present thesis have a very high stacking fault energy and consequently do not exhibit partial dislocations.

Dislocations may move through the lattice either by glide or by climb. Glide is defined as the movement of a dislocation in a plane that contains both the dislocation line and its Burgers vector. This type of motion is relatively easy because it is conservative, i.e. no volume is created or taken away. Climb occurs when a dislocation moves out of its glide plane in a direction normal to the Burgers vector. This process requires the creation of interstitials or vacancies and is therefore diffusion-controlled. For this reason, climb is appreciable only at temperatures higher than approximately half of the absolute melting temperature (as in Chapter 5 of this thesis). Successive glide of many dislocations in the same direction leads to shearing of the material and is termed slip. Slip is mostly restricted to close-packed planes and directions. In fcc metals, slip occurs almost exclusively on {111} planes in <110> directions, which results in 12 independent slip systems. In bcc metals, the slip direction is <111>, but the slip planes are not well defined on a macroscopic scale. Microscopic evidence suggests that slip occurs mostly on {110} and {112} planes, and that {110} planes are preferred at low temperature. Since for screw dislocations, the Burgers vector is parallel to the dislocation line, they do not have one unique slip plane and can easily cross-slip from one plane to another.
In face-centered cubic crystals, glide on a given slip system is controlled solely by the resolved shear stress on the slip plane in the slip direction. According to Schmid’s law, slip is initiated when the resolved shear stress attains a critical value $\tau_c$, which for a uniaxial tensile stress is given by

$$\tau_c = \sigma_y \cos \phi \cos \lambda$$

(2.2)

where $\sigma_y$ is the yield stress and $\phi$ and $\lambda$ are, respectively, the angles of the slip plane normal and the slip direction with respect to the tensile axis. The critical resolved shear stress is unaffected by other components of the applied stress tensor and independent of the sense of shearing. The reason why fcc metals obey Schmid’s law is that dislocations are confined to the close-packed $\{111\}$ planes and possess planar cores. In bcc metals however, dislocation cores can spread onto several non-parallel planes, thereby giving rise to deviations from Schmid behavior [4,5]. Commonly found manifestations of non-Schmid behavior include tension-compression asymmetry and effects of non-glide stress components on the critical resolved shear stress.

Plastic deformation of bcc metals in general exhibits characteristic features depending on the deformation temperature [6], e.g. a strong temperature dependence of the yield stress and anomalous slip. The latter has been studied most extensively in bcc Fe and Nb, but also in other bcc metals, such as Ta [7-10], V [11-13] and Mo [14,15]. The occurrence of the slip is anomalous as the slip plane observed belongs to one of the $\{110\}$ planes with a small Schmid factor [16-21]. In situ transmission electron microscopy experiments have been performed [22,23] at low and intermediate temperatures on Nb, Mo and bcc Fe. In all these bcc materials, a transition in the dislocation dynamics has been observed between a low temperature behavior, characterized by an almost continuous movement of long screw dislocations and an intermediate temperature behavior, where mixed dislocations predominate.

### 2.1.2 Grain boundaries

Polycrystalline materials generally consist of many grains (single crystals) with different orientations separated by grain boundaries. The geometry of a grain boundary is macroscopically characterized by five degrees of freedom: three angles define the crystallographic orientation of the two crystals with respect to one another, while two parameters describe the inclination of the grain boundary plane. To fully characterize the boundary geometry on a microscopic level, three additional parameters are needed to define the atomic-scale relative translation of
the two grains. Depending on the misorientation, grain boundaries may be of the
tilt type, when the rotation axis lies in the boundary plane, or of the twist type,
when the rotation axis is normal to the boundary plane (Figure 2.2). A general
grain boundary may have both tilt and twist components.

When the misorientation angle between two adjacent grains is small, the
boundary structure is well described by a periodic network of dislocations. The
boundary is accordingly called a low-angle grain boundary and its dislocation
configuration can be determined using Frank’s rule [24]. In the particular case of
a low-angle tilt boundary, the grain boundary simply consists of parallel edge
dislocations with spacing \( d = \frac{b}{\theta} \), where \( b \) is the magnitude of the Burgers
vector and \( \theta \) is the misorientation angle. The grain boundary energy \( E \) can be
approximated by the linear-elastic energy of the dislocation network following
Read and Shockley [25]:

\[
E = E_0 \theta (A - \ln \theta)
\]

where \( E_0 \) is a function of the elastic properties of the material and \( A \) is a constant,
which depends on the core energy of an individual dislocation. When the
misorientation angle increases beyond 10-15°, the dislocation cores start to
overlap so that the derivation of Eq. (2.3) based on linear elasticity ceases to be
valid and the strain field can no longer be approximated by that of isolated
dislocations. Therefore, the description of the boundary structure by Frank’s rule
has little physical significance for high-angle grain boundaries.

High-angle grain boundaries in cubic crystals are often observed to have
special orientation relationships that are described by the concept of the
coincident site lattice (CSL) [26]. In this concept, the lattices of two neighboring
grains are hypothetically allowed to interpenetrate. For certain misorientations, a
proportion of the two sets of lattice points coincides; these lattice points constitute the coincident site lattice. The CSL is characterized by an odd integer $\Sigma$, where $1/\Sigma$ is the proportion of lattice sites that are common to both lattices. Figure 2.3 shows a schematic projection of a $\Sigma = 5$ boundary in an fcc crystal and its corresponding CSL.

Although the relationship between the $\Sigma$ value and the grain boundary energy is not straightforward, grain boundaries of small $\Sigma$ generally have a relatively low energy, as shown by the cusps in the energy plot in Figure 2.4. Furthermore, the width of the cusps in Figure 2.4 suggests that for small deviations from a perfect CSL misorientation, the structure of the coincident boundary is not immediately lost. Such deviations may be accommodated by a network of grain boundary dislocations, analogously to the structure of low-angle boundaries (for which $\Sigma = 1$) as described by Frank’s rule. The maximum angular deviation $\Delta\theta$ from the ideal CSL misorientation for which the boundary still maintains its CSL structure can be estimated from the acceptance criterion introduced by Brandon [28], which requires that

$$\Delta\theta < \frac{15^\circ}{\sqrt{\Sigma}}$$

The dislocations that accommodate the deviation from CSL misorientation are termed secondary grain boundary dislocations. They are defined by reference to the displacement shift complete (DSC) lattice [29,30]. Each CSL has a specific DSC lattice associated with it, which consists of those

Figure 2.3: (a) Projection along the <001> axis of a $\Sigma 5$ boundary on a [013] plane in fcc. (b) CSL constructed by overlaying both lattices. Half-filled circles denote common lattice points. The broken square on the left represents the projection of the CSL unit cell; the DSC lattice is shown on the right by the solid lines (adapted from [3]).
displacement vectors of one grain with respect to the other that leave the CSL unchanged except for a translation. The unit cell of the DSC lattice is smaller than that of the crystal lattice, as shown in Figure 2.3b. Therefore, secondary (or DSC) boundary dislocations generally have a Burgers vector that is not a lattice translation vector and their movement is consequently restricted to the boundary plane.

2.1.3 Interaction between dislocations and grain boundaries
Grain boundaries present effective obstacles to dislocation motion, since dislocations coming upon a boundary generally do not have the correct Burgers vector and slip plane to glide into the next crystal. Most commonly, the elastic interaction between dislocations and grain boundaries is repulsive and consequently the dislocations pile up at the boundary. A schematic representation of a pile-up is shown in Figure 2.5. Under an applied shear stress $\tau_a$, the source emits dislocations of the same sign on a single slip plane. The dislocations interact elastically and their equilibrium spacing, which decreases towards the front of the pile-up, can be calculated accordingly [31]. For a single-ended pile-up of length $l$, the number of dislocations within the pile-up is

$$n = \frac{\pi \tau_a}{\mu b}$$

(2.5)

ignoring the difference between edge and screw parts. The shear stress at the spearhead of the pile-up is equal to $n\tau_a$, i.e. the applied shear stress is multiplied by the number of dislocations. If this shear stress exceeds some critical value, the interaction between the piled-up dislocations and the boundary becomes plastic.

Figure 2.4: Measured relative energy of symmetric $\langle 110 \rangle$ tilt boundaries in aluminum as a function of misorientation angle (taken from [27]). Three cusps can be distinguished at the angles corresponding to $\Sigma 9$, $\Sigma 3$ and $\Sigma 11$ boundaries, respectively.
Basic concepts

In describing the plastic interaction between an incoming lattice dislocation and a grain boundary, several mechanisms may be distinguished. An overview is given in Figure 2.6. All mechanisms require that the sum of the Burgers vectors is conserved, i.e.

$$\sum_i b_{i \text{ IN}} = \sum_j b_{j \text{ OUT}} + \sum_k b_{k \text{ GB}}$$  \hspace{1cm} (2.6)

Dislocations may be transmitted directly across the boundary if the slip planes on both sides intersect along a line that lies in the boundary plane. For pure screw dislocations, the Burgers vector remains unchanged and no residual dislocation is left in the grain boundary (Figure 2.6a). The transmission of dislocations with an edge component requires the creation of a residual grain boundary dislocation with a Burgers vector equal to the difference of the Burgers vectors of the incoming and outgoing lattice dislocations (Figure 2.6b). The residual dislocation is always a DSC dislocation. Alternatively, a dislocation may be absorbed by the boundary without emission of a dislocation in the adjacent grain (Figure 2.6c). In this case, the lattice dislocation fully dissociates into DSC dislocations whose Burgers vectors add up to that of the lattice dislocation. Repeated absorption may lead to a series of glissile DSC dislocations piling up at an obstacle in the grain boundary, e.g. a ledge. The stress generated by this pile-up can nucleate outgoing dislocations in the adjacent grain, thereby giving the appearance of absorption and re-emission (Figure 2.6d). Since the incoming and outgoing dislocations have different Burgers vectors, a residual boundary dislocation is left behind. In the following, all mechanisms by which dislocations are emitted in the adjacent grain (i.e. Figures 2.6a, 2.6b and 2.6d) are collectively referred to as slip transfer.

Experimental observations have led to predictive criteria for the selection of the slip system on which dislocations are emitted during slip transfer. The first of these was put forward in 1957 by Livingston and Chalmers [32] based on observations of slip lines on deformed aluminum bicrystals. They proposed that...
slip transfer takes place on the slip system for which the factor $N$ defined below is a maximum:

$$N = (e_1 \cdot e_2)(g_1 \cdot g_2) + (e_1 \cdot g_2)(g_1 \cdot e_2) \quad (2.7)$$

where $e_1$ and $e_2$ are the slip plane normals of the piled-up and emitted dislocations and $g_1$ and $g_2$ are the slip directions in the pile-up and emission grains, respectively.

More detailed studies of slip transfer were performed later with the availability of transmission electron microscopy (TEM). Based on post mortem TEM observations on deformed specimens as well as in situ tensile deformation...
experiments, Shen et al. [33,34] proposed the following criteria. For the favored slip plane in the emission grain, the angle of the intersection lines $L_1$ and $L_2$ of the pile-up and emission slip planes with the boundary plane should be a minimum, as well as the angle between the slip directions in both grains. This geometric criterion may be expressed as

$$M = (L_1 \cdot L_2)(g_1 \cdot g_2)$$  \hspace{1cm} (2.8)$$

where $M$ is to be maximized. In addition, the Burgers vector of the dislocations on the emission slip plane is calculated to be the one with the maximum Peach-Köhler force [1] from the incoming dislocations acting on it.

Lee et al. [35] proposed a similar set of criteria, which are: (i) the angle between the intersection lines $L_1$ and $L_2$ should be minimized so that

$$M' = (L_1 \cdot L_2)$$  \hspace{1cm} (2.9)$$

is a maximum for the favored slip plane; (ii) the resolved shear stress acting on the outgoing slip system from the piled-up dislocations should be a maximum and (iii) the magnitude of the Burgers vector of the residual dislocations left in the grain boundary plane should be a minimum. A competition exists between criteria (ii) and (iii) regarding the prediction of the Burgers vector of the outgoing dislocations; either of both may be dominant in particular situations. The total set of criteria are independent of the crystal structure and have shown agreement with TEM observations of slip transfer in both fcc [35] and bcc [36] metals.

2.1.4 Grain-size dependence of mechanical properties

The yield stress of polycrystalline materials has long been known to depend on their grain size. This effect is empirically described by the well-known Hall-Petch relation [37,38]:

$$\sigma_y = \sigma_0 + \frac{k_y}{\sqrt{d}}$$  \hspace{1cm} (2.10)$$

where $\sigma_y$ is the yield stress of a polycrystal with average grain size $d$, $\sigma_0$ is the yield stress of a single crystal and $k_y$ is a material-dependent proportionality constant, which is often referred to as the Hall-Petch slope. The Hall-Petch effect has been experimentally validated for many polycrystalline materials, both metallic and non-metallic, and can be physically explained in terms of dislocation mobility and pile-up at grain boundaries [39,40]. Macroscopic yield of a polycrystal requires the activation of simultaneous slip in adjacent grains. Prior to
macroscopic yield, microscopic yield occurs in favorably oriented grains, i.e. those with a high Schmid factor as defined by Eq. (2.2). This causes dislocations to pile up near grain boundaries. The stress at the spearhead of the pile-up can subsequently activate slip in less favorably oriented grains as described in the previous section. Evidently, larger grain sizes allow for longer pile-ups, imposing more stress at the grain boundaries (cf. Eq. (2.5)). This accounts for the grain size dependence expressed through Eq. (2.10). The value of $\sigma_0$ can be rationalized as the lattice friction stress resisting dislocation motion in a single crystal.

While the Hall-Petch relation holds well for conventional grain sizes, the assumptions made about dislocation mobility cease to be valid when the grain size drops into the nanometer regime. For very small grains (< 20 nm), mobile dislocations are not likely to occur, nor can they be brought together close enough to form a pile-up unless very high stresses are applied. Furthermore, grain boundaries and triple junctions start dominating the mechanical behavior at submicrometer grain sizes, e.g. by grain boundary sliding or triple junction strengthening [41]. Measurements of hardness and yield stress at these grain sizes show different results, including both “normal” and “inverse” Hall-Petch behavior [42].

2.2 Imaging techniques

2.2.1 Transmission electron microscopy

Conventional transmission electron microscopy (TEM) is the most widely used technique to image defects in crystals [43]. The image contrast caused by dislocations is based on electron diffraction following Bragg’s law. A useful description of electron diffraction in crystals is given by the concept of the Ewald sphere [44]. The Ewald sphere is constructed in the reciprocal lattice by letting the incident wave vector $k_i$ end on one of the reciprocal lattice points and drawing a sphere around its origin with a radius of $|k_i| = 1/\lambda$, where $\lambda$ is the (relativistic) de Broglie wavelength of the electrons. This is illustrated in Figure 2.7. The diffracted wave vector $k_d$ has the same length as $k_i$ and the magnitude of their difference $K = k_d - k_i$ is related to the diffraction angle $\theta$ by

$$|K| = \frac{2\sin \theta}{\lambda} \tag{2.11}$$
For those reciprocal lattice points that are intersected by the Ewald sphere, the diffraction vector has a length equal to the inverse interplanar spacing $1/d_{hkl}$ and therefore the Bragg condition is satisfied, i.e. Eq. (2.11) becomes:

$$\lambda = 2d_{hkl} \sin \theta$$

(2.12)

In electron diffraction of volumes that are constrained to a small length scale in one or more dimensions (a TEM foil is constrained in at least one dimension: the thickness), diffraction also occurs for wave vectors that do not exactly fulfill the Bragg condition. The excitation error can accordingly be defined as $s_g = k_k - k_g$.

To interpret electron micrographs it is essential to understand the factors that determine the intensities of Bragg-diffracted beams. Various approaches can be followed, i.e. ranging from the kinematical theory to the dynamical theory of electron diffraction. The former is based on the assumptions that only elastic scattering takes place (hence no absorption) and that an electron can be scattered only once, whereas the latter also allows for interaction between the diffracted beams [45]. This interaction is particularly well defined when the crystal is tilted in such a way that, besides the direct beam, only one beam is strongly diffracted (i.e. $|s_g| >> 0$ for the other reflections). As the transmitted wave with amplitude $\phi(z)$ propagates through the crystal, its amplitude is depleted by diffraction and the amplitude $\phi_g(z)$ of the diffracted beam increases, i.e. a dynamical interaction between $\phi(z)$ and $\phi_g(z)$ exists. If we assume that $s_g$ is parallel to the electron
beam, the interaction can be described by the following pair of coupled differential equations, known as the Howie-Whelan equations [45]:

\[
\frac{d\phi_g}{dz} = \frac{\pi i}{\xi_0} \phi_g + \frac{\pi i}{\xi_g} e^{2\pi i s_g z} \tag{2.13}
\]

and

\[
\frac{d\phi_0}{dz} = \frac{\pi i}{\xi_0} \phi_0 + \frac{\pi i}{\xi_g} e^{2\pi i s_g z} \tag{2.14}
\]

where \( s_g \) and \( \xi_0 \) and \( \xi_g \) are the extinction distances for the direct and diffracted waves, respectively. The extinction distance is a characteristic length scale determined by the atomic number of the material, the lattice parameters and the wavelength of the electrons and typically lies between 10 and 100 nm. Eq. (2.14) shows that the change in \( \phi_0(z) \) as a function of depth \( z \) is the sum of forward scattering and scattering from the diffracted beam, taking into account a phase change of \( \pi/2 \) caused by the scattering. Solving Eqs. (2.13) and (2.14) for \( \phi_g \) gives

\[
\phi_g = \frac{i}{\sqrt{1 + s_g^2 \xi_g^2}} \sin \left( \frac{\pi t}{\xi_g} \sqrt{1 + s_g^2 \xi_g^2} \right) \tag{2.15}
\]

where \( t \) is the thickness of the crystal. Accordingly, the intensity of the diffracted beam becomes

\[
I_g = |\phi_g|^2 = \frac{\sin^2 \pi t s_{eff}}{s_g^2 s_{eff}^2} \tag{2.16}
\]

where \( s_{eff} \) is an effective value of \( s_g \) defined by \( s_{eff} = (s_g^2 + \xi_g^{-2})^{1/2} \). In fact, by replacing \( s_{eff} \) by \( s_g \) in Eq. (2.16), the intensity according to the kinematical approximation is found. Absorption can be included in the dynamical theory by adding appropriate terms to both \( \xi_0 \) and \( \xi_g \). Mathematically speaking, absorption is included simply by allowing the arguments of the sines and cosines to become complex.

Dislocations give rise to contrast because they locally distort the lattice and thereby change the diffraction conditions. If the distortion is given by a displacement field \( \mathbf{R} \), Eq. (2.13) becomes

\[
\frac{d\phi_g}{dz} = \frac{\pi i}{\xi_0} \phi_g + \frac{\pi i}{\xi_g} e^{2\pi i (s_g z + g \cdot \mathbf{R})} \exp \left[ -2\pi i \left( s_g z + g \cdot \mathbf{R} \right) \right] \tag{2.17}
\]
In order for a dislocation to contribute to contrast formation, the dot product $\mathbf{g} \cdot \mathbf{R}$ must be nonzero. A screw dislocation in an isotropic elastic medium has a displacement field parallel to its Burgers vector, and therefore produces no contrast when $\mathbf{g} \cdot \mathbf{b} = 0$. General dislocations have a displacement field with more components; their image contrast also depends on $\mathbf{g} \cdot \mathbf{b}_e$ and $\mathbf{g} \cdot (\mathbf{b} \times \mathbf{u})$, where $\mathbf{b}_e$ is the edge component of the Burgers vector and $\mathbf{u}$ is a unit vector along the dislocation line [46]. In practice however, only very faint contrast occurs when $\mathbf{g} \cdot \mathbf{b} = 0$ but $\mathbf{g} \cdot \mathbf{b}_e \neq 0$ and $\mathbf{g} \cdot (\mathbf{b} \times \mathbf{u}) \neq 0$. Therefore, the “invisibility criterion” $\mathbf{g} \cdot \mathbf{b} = 0$ is used commonly to determine Burgers vectors of dislocations in elastically isotropic solids. The determination of a Burgers vector involves finding two reflections $\mathbf{g}_1$ and $\mathbf{g}_2$ for which the dislocation is invisible, so that $\mathbf{b}$ is parallel to $\mathbf{g}_1 \times \mathbf{g}_2$.

The situation where only one beam $\phi_0$ is strongly diffracted is referred to as a two-beam condition. This type of diffraction is widely used in conventional TEM of crystalline materials because the contrast is well defined and the Burgers vectors of the dislocations can be determined as described above. By using the objective aperture in the microscope, either of the two beams can be selected to form an image and accordingly, two imaging modes may be distinguished: bright field (BF) when the direct beam $\phi_0$ is used, and dark field (DF) when the diffracted beam $\phi_g$ is used. An example is shown in Figure 2.8. Since $I_0 = 1 - I_g$, the intensity in the two imaging modes is complementary. Generally, the diffracted beams do not coincide with the optical axis of the microscope and consequently the DF image will not be of maximum quality due to spherical aberration. To overcome this problem, the incident beam is normally tilted in such a way that the desired diffracted beam passes along the optical axis.

While two-beam conditions produce high contrast of dislocations, the resolution at which these defects are resolved is not optimal since the lattice planes around the dislocations are distorted over a relatively large area. In order to obtain the maximum resolution, the crystal should be tilted slightly further by such an amount that the exact Bragg condition is only fulfilled within a small region near the dislocation. In this way, a high-resolution image is obtained in which the dislocation shows up as a bright line. This technique is referred to as weak-beam imaging. The deviation from the exact Bragg condition for perfect crystal is given by $s_g$ as mentioned above and can be determined accurately by the relative position of the so-called Kikuchi pattern with respect to the diffraction pattern. The origin of the Kikuchi pattern lies in the elastic re-scattering of inelastic scattered electrons as further explained in the following section. The
Figure 2.8: Dislocation pile-up configurations near a high-angle grain boundary in an Al-Mg alloy imaged in two-beam condition with $g = <200>$; (a) bright-field image; (b) dark-field image; (c) diffraction pattern. The variation of the contrast along the dislocation lines results from the inclination of the dislocations with respect to the surface.
Basic concepts

width of a dislocation image is approximately \(0.3\xi_g\), i.e. several tens of nanometers in conventional bright- and dark-field imaging. This width can be detrimental to the observations of dislocations that are very closely spaced. Since the effective extinction distance decreases for increasing deviation away from the Bragg condition, the width of a dislocation image can be reduced to values in the order of 1 to 5 nm in weak-beam imaging.

At lower magnifications, the average crystal orientation may vary considerably within the observed area, so that only a small area of the specimen can be set up in two-beam condition. This is especially relevant in specimens that have been deformed prior to preparation, e.g. as in Chapter 5. In these cases, it is often convenient to orient the specimen close to a zone axis, so that many reflections are weakly excited, independently of small changes in orientation. By using the direct beam for imaging, the dislocation structure can be imaged with good contrast over a relatively large area. However, the contrast from individual dislocations is generally smaller than in two-beam condition and not well defined since it results from many different reflections.

2.2.2 Electron backscatter diffraction

The previous section described the nature of electron diffraction patterns due to coherent forward scattering in thin crystals. In contrast, electron backscatter diffraction (EBSD) can give information about the crystal structure and orientation without the requirement of preparation to electron transparency. Consider an electron beam entering a crystal as illustrated in Figure 2.9. The

![Figure 2.9: Schematic of electron backscatter diffraction.](image)
primary electrons are diffusely scattered in all directions, and consequently part of them impinges on a given set of lattice planes \( \{hkl\} \) at the Bragg angle \( \theta_B \). Although the Bragg angle depends on the wavelength of the electrons according to Eq. (2.12), the relative change in wavelength due to inelastic scattering is usually very small and therefore both elastically and inelastically scattered electrons may contribute to Bragg diffraction at approximately the same angle. The trajectories of these electrons constitute two cones on either side of the diffracting set of planes with an apex half-angle of \( 90^\circ - \theta_B \) around the plane normal \( <hkl> \). These are known as Kossel cones. When the backscattered electrons are recorded on a phosphor screen, the interception of a pair of Kossel cones with the screen is seen as a pair of hyperbolas, which approximate to straight lines (Kikuchi lines) because \( \theta_B \) is relatively small.

A typical EBSD pattern is shown in Figure 2.10. The distance between a pair of Kikuchi lines is related to the interplanar distance of the corresponding lattice planes through Eq. (2.12), and the angles between different pairs of Kikuchi lines represent the angular relationships in the crystal. The pattern therefore contains many important characteristics of the crystal. Furthermore, the orientation of the crystal can be determined from the position of the lines; a small change in orientation corresponds to a translation of the pattern.

The principle of EBSD has been implemented in scanning electron microscopy (SEM) for phase determination and orientation imaging [47]. The specimen surface is usually oriented at a small angle (typically \( \sim 20^\circ \)) with respect to the electron beam in order to promote backscatter diffraction. The electron beam subsequently scans across the area of interest on the surface, and at each position the backscatter pattern is recorded and analyzed. This allows for

*Figure 2.10: EBSD pattern of a Mo crystal with the surface normal close to a \( <110> \) zone axis.*
rapid mapping of the texture of polycrystalline materials, together with phase identification from a limited number of pre-selected phases. The orientation resolution that can be obtained from the patterns is of the order of 1°.

2.3 Nanoindentation

2.3.1 Measurement of hardness and elastic modulus

Microhardness testing has been used for many decades to locally measure the hardness of materials. The method is based on the optical measurement of the impression left by pushing an indenter tip of known geometry into the material at a given load. The hardness is consequently defined as the applied load divided by the projected area of the impression. The development of nanoscale materials and the continuing miniaturization of devices have raised the demand to measure mechanical properties on a submicrometer scale. Conventional microhardness testing methods are not suitable for measurement on such small scales because the area of the impression can no longer be accurately measured by light microscopy. Load- and depth-sensing indentation, commonly referred to as nanoindentation, overcomes the requirement to image the impression by extracting the hardness from the indentation load and depth, both of which are continuously measured during the loading and unloading of the indenter. In addition, other properties such as the elastic modulus can be deduced from these data, thus offering a more complete description of the mechanical behavior of the material. Figure 2.11 schematically shows a load versus displacement plot of a complete cycle of loading and unloading.

The most commonly used methods for analyzing indentation data are based on the problem of a rigid punch indenting an elastic half-space. Sneddon [48] established an analytical procedure to show that for any axisymmetric indenter geometry, the relationship between the applied load \( P \) and the elastic displacement \( h \) may be expressed as

\[
P = ah^m
\]  

where \( a \) is a constant and \( m \) is an exponent that was found to be \( m = 1 \) for flat cylinders, \( m = 1.5 \) for paraboloids and \( m = 2 \) for cones. The first experiments in which the indentation load and depth were continuously measured were conducted by Tabor and Stillwell [49,50]. They found that the impression formed in a metal surface by a conical indenter is still conical, but with a larger included tip angle. The diameter of the impression remained unchanged during unloading,
i.e., only the depth recovered elastically. The important implication of these observations is that the unloading behavior may be described by the elastic contact solution while the plasticity is accounted for by incorporating the final shape of the impression into the elastic analysis [51]. This provides a relatively simple approach to plastic indentation.

According to elastic contact theory [52], the elastic modulus of a material indented by a conical indenter may be calculated from the unloading stiffness $S$ and the projected contact area $A$ through the equation

$$S = \frac{dP}{dh} = \frac{2}{\sqrt{\pi}} \beta E_r \sqrt{A}$$

with $\beta = 1$. The finite stiffness of the indenter is taken into account by defining the reduced modulus $E_r$ according to

$$\frac{1}{E_r} = \frac{(1-v^2)}{E} + \frac{(1-v_i^2)}{E_i}$$

where $E$ and $v$ are the elastic modulus and the Poisson’s ratio of the material, respectively, and $E_i$ and $v_i$ are the same parameters for the indenter. While Eq. (2.19) was originally derived for a conical indenter, it has been established that it holds equally with $\beta = 1$ for any body of revolution of a smooth function [51]. Moreover, finite element calculations by King [53] have shown that Eq. (2.19)
also applies to indentation by flat-ended punches with square and triangular cross sections, provided $\beta$ is taken to be 1.012 and 1.034, respectively.

The most commonly used indenter shape for nanoindentation is the Berkovich indenter, which is a three-sided pyramid with a total included angle of 142.3°. Doerner and Nix [54] developed a comprehensive method for extracting the hardness and elastic modulus from nanoindentation load-displacement data of such indenters. Based on observations in selected metals, their method assumes that the first part of the unloading curve is approximately linear. This implies a constant contact area during unloading, as would be the case for a flat-ended punch. The unloading stiffness $S$ is then easily calculated by fitting a straight line to the upper portion of the unloading curve.

More extensive measurements by Oliver and Pharr [55] on a broader range of materials showed that the initial unloading behavior for Berkovich indenters is hardly ever linear and is more adequately described by a power-law relationship analogously to Eq. (2.18):

$$P = \alpha(h - h_f)^m$$  (2.21)

where $h_f$ is the final depth of the remaining impression and the parameters $\alpha$, $h_f$ and $m$ are empirically determined by fitting to the entire unloading curve. The exponent $m$ was found to vary from material to material in the range 1.25 to 1.51. The unloading stiffness $S$ at peak load can be calculated by evaluating the derivative of Eq. (2.21) at maximum depth $h = h_{\text{max}}$. Eq. (2.19) subsequently relates the elastic modulus $E_r$ to this stiffness and the projected contact area $A$ at peak load.

In order to calculate the projected contact area, the indenter geometry needs to be known through the area function $A(h)$, which relates the cross-sectional area of the indenter to the distance from its end $h$. An ideal Berkovich indenter has a square area function of $A(h) = 24.5 \, h^2$. In practice however, blunting at the tip leads to small deviations from the ideal area function, which can be accounted for by using the following approximation:

$$A(h) = a_0 h^2 + a_1 h + a_{1/2} h^{1/2} + a_{1/4} h^{1/4} + a_{1/8} h^{1/8} + ...$$  (2.22)

The coefficients $a_n$ may be determined by calibration on an isotropic material with known mechanical properties. From the calibrated area function, the effective end radius of the tip may be estimated by defining the equivalent radius $r = \sqrt{(A/\pi)}$, writing $h(r)$ as a function of $r$ and (numerically) evaluating $d^2 h / dr^2$ at $r = 0$. Typical values for the end radius of curvature lie between 50 and 200 nm.
Figure 2.12 shows a cross section of an indented surface under load and after load removal, illustrating that elastic recovery only takes place in the depth direction of the applied load and the radius of the impression $a$ remains constant as mentioned before. The total displacement $h$ at any time during the indentation cycle can be written as $h = h_c + h_s$, where $h_c$ is the depth along which contact is made (the contact depth) and $h_s$ is the displacement of the surface at the perimeter of the contact. Using Sneddon’s expression [48] for the surface profile outside the area of contact, Oliver and Pharr [55] obtained the following expression for the contact depth $h_c$ at peak load $P_{\text{max}}$:

$$h_c = h_{\text{max}} - \varepsilon \frac{P_{\text{max}}}{S}$$

(2.23)

where $\varepsilon$ is a parameter that depends on the indenter geometry, being 0.75 for a Berkovich indenter. Once the contact depth is known, the contact area $A(h_c)$ can be evaluated according to Eq. (2.22), and together with the unloading stiffness determined using Eq. (2.21), the elastic modulus is readily obtained following Eq. (2.19). Furthermore, the hardness $H$, defined as the mean pressure of the material under load, can be determined using

$$H = \frac{P_{\text{max}}}{A(h_c)}$$

(2.24)

Another figure that can be obtained from the load-displacement curve is the work done during an indentation cycle. Three quantities can be identified: (i) the total work done by the indenter to the sample $W$, defined by

$$W = \int_{0}^{h_{\text{max}}} Pdh$$

(2.25)

Figure 2.12: Cross section of the surface profiles during and after indentation, showing a graphical interpretation of the depths used in the analysis [55].
(ii) the elastic work $W_e$ recovered during unloading, which is defined analogously as the area under the unloading curve, and (iii) the plastic work $W_p = W_t - W_e$. According to Cheng and Cheng [56], the ratio of plastic to total work is a property of the indented material and the indenter geometry only, and is therefore independent of the load and depth of indentation. This was shown by describing the loading and unloading curves in terms of dimensionless functions and subsequently integrating them following Eq. (2.25).

Oliver and Pharr’s methodology as described above has been validated for various elastically isotropic materials to give results for hardness and elastic modulus within 4% of literature values [55]. There are however materials for which the assumption of elastic contact on which the method is based, leads to a systematic underestimate of the contact area, resulting in an overestimate of the hardness and the modulus. This occurs when a relatively large amount of material piles up around the perimeter of the contact, as is the case for many soft materials [57]. The pile-up results in a significantly larger contact area than predicted by Eq. (2.23). Finite-element analyses [58] have established that the amount of pile-up is well characterized by two parameters, namely the ratio of the final to maximum indentation depth $h_f / h_{\text{max}}$ and the work hardening rate $\eta$. For indentations with $h_f / h_{\text{max}} < 0.7$, very little pile-up occurs; however, as $h_f / h_{\text{max}}$ approaches 1, the amount of pile-up becomes significant if the material shows little work hardening, i.e. $\eta \ll \sigma_y$.

2.3.2 Continuous stiffness measurement

The analysis described in the previous section uses the contact stiffness as calculated from the unloading curve, and consequently the hardness and elastic modulus are only obtained at the maximum indentation load and depth. By applying a small oscillation to the force signal however, the contact stiffness may be measured continuously during loading and unloading. This technique is referred to as continuous stiffness measurement (CSM) [55,59].

The oscillation superimposed on the load signal has a frequency of typically 45 Hz and an amplitude that is sufficiently small not to influence the deformation process significantly. The amplitude of the resulting oscillation of the displacement is approximately 2 nm for the CSM results presented in this thesis. The phase difference between the load signal and the displacement signal is measured using a lock-in amplifier. The contact stiffness of the indentation may be calculated from this phase angle, provided that the mechanical response of the entire system is known. Figure 2.13 shows a simple dynamic model of the
indentation system. The mass of the indenter assembly $m$ (typically 5 g) is supported by three components which are represented by springs in the model: the material with contact stiffness $S$, the load frame with stiffness $K_f$ and the column support springs with stiffness $K_s$. The total damping in the system (mainly due to the capacitive displacement sensor) is included in the model through the damping coefficient $D$. The phase angle $\phi$ can now be expressed as

$$\tan \phi = \frac{\omega D}{\left( S^{-1} + K_f^{-1} \right)^{-1} + K_s - m\omega^2}$$

Figure 2.13: Representation of the indentation system as a damped harmonic oscillator, used to calculate the contact stiffness $S$ during indentation [55].

where $\omega$ is the angular frequency of the applied oscillation. The constants $K_s$ and $D$ can be determined from the response of the system when the indenter is not in contact with the surface so that $S = 0$. The load frame stiffness $K_f$ is usually calibrated together with the indenter area function from a set of indentations with varying depths through an iterative procedure as outlined in reference [55].

Oliver and Pharr used continuous stiffness measurement to show that the contact area decreases continuously from the onset of unloading, supporting their claim that the linear approximation used by Doerner and Nix is not entirely adequate. In addition, the CSM technique has proved useful to measure variations in hardness and modulus as a function of indentation depth in numerous cases. Three types of such variations may be distinguished: (i) those inherent to indentation-induced deformation (termed the indentation size effect, which will be discussed in the next section); (ii) variations due to discrete yielding events, such as nucleation of defects and slip transfer across grain boundaries (see Chapter 3); and (iii) through-thickness variations of mechanical properties in the case of functionally graded materials [60] and coated or multilayered structures [61].
2.3.3 Indentation size effect

Indentations made in crystalline materials show a significant increase in hardness with decreasing indentation size, particularly in the submicrometer depth regime. This effect is commonly known as the indentation size effect and is attributed to enhanced hardening at low depths caused by so-called geometrically necessary dislocations [62,63]. Nix and Gao [64] devised a simple model to quantify the hardening effect using this concept.

During indentation of a single crystal, geometrically necessary dislocations are required to account for the permanent shape change at the surface. In the case of a conical indenter, the arrangement of these dislocations can be idealized as a set of coaxial circular prismatic dislocation loops, equidistantly spaced along the surface of the indentation. Assuming that the injected loops stay within a hemisphere under the contact perimeter, it is easily shown that their density is given by

$$\rho_c = \frac{3}{2bh} \tan^2 \theta$$  \hspace{1cm} (2.27)

where $b$ is the magnitude of the Burgers vector of the dislocations, which is assumed to be normal to the surface, $h$ is the indentation depth and $\theta$ is the apex half-angle of the indenter. The lattice dislocations that are created in addition to the geometrically necessary dislocations by other nucleation processes are called statistically stored dislocations [65]; this term also covers the dislocations already present in the material prior to indentation. The shear stress can be expressed by Taylor’s relation as

$$\tau = \alpha \mu b \sqrt{\rho_c + \rho_s}$$  \hspace{1cm} (2.28)

where $\alpha$ is a constant, $\mu$ is the shear modulus and $\rho_s$ is the density of statistically stored dislocations. We assume von Mises flow to write the equivalent yield stress as $\sigma_y = \sqrt{3} \tau$ and use Tabor’s factor of 3 for the hardness so that $H = 3\sigma_y$. Combining Eqs. (2.27) and (2.28) subsequently gives the following expression for the hardness:

$$\frac{H}{H_0} = \sqrt{1 + \frac{h^2}{h}}$$  \hspace{1cm} (2.29)

where

$$H_0 = 3\sqrt{3} \epsilon \mu b \sqrt{\rho_s}$$  \hspace{1cm} (2.30)
is the hardness that would arise from the statistically stored dislocations alone, and

\[
\hat{h}^* = \frac{81}{2} b \alpha^2 \tan^2 \theta \left( \mu \frac{\mu}{H_0} \right)^2
\]

(2.31)

is a length scale that characterizes the depth dependence of the hardness. From Eq. (2.31) it follows that for intrinsically hard materials, i.e. those with a high value of \(H_0\), the characteristic length scale \(\hat{h}^*\) is very small. Consequently, the indentation size effect is much less pronounced for hard materials than for soft materials when compared at the same indentation depth.

When the square hardness is plotted versus the inverse indentation depth, a straight line results with intercept \(H_0^2\) and slope \(\hat{h}^*\) as given by Eq. (2.29). This is illustrated by the data for an annealed steel specimen plotted in Figure 2.14. The hardness value of \(H_0\) determined from the intercept compares well to the macroscopic hardness of the material. Mechanical polishing leads to work hardening at the surface, giving the appearance of an additional size effect as shown by the dashed line. Whereas uniform work hardening would lead to both an increase of \(H_0\) and a decrease of \(\hat{h}^*\), the surface hardening introduced by polishing is only measured at small indentation depths, and therefore the intercept \(H_0^2\) (corresponding to the limit of large depths) does not change appreciably. However, the derivation of Eq. (2.29) based on a uniform density of statistically

![Figure 2.14: Indentation size effect plot for two specimens of the same duplex steel with a macroscopic hardness of 2.9 GPa. Both specimens were annealed and polished with a colloidal silica suspension; the one represented by the solid data points was subsequently annealed again to relieve residual stresses.](image-url)
stored dislocations does not apply to this situation, and therefore care should be
taken in analyzing size effects in indentation of polished surfaces.

Since classical plasticity theories do not include material length scales,
they cannot account for size effects as described above. In such theories, the yield
stress depends only on the strain and is unrelated to strain gradients that may be
present. However, the geometrically necessary dislocations generated during
shallow indentations involve appreciable strain gradients [66], and their effect on
the hardness can be adequately described by strain gradient plasticity theory. By
recognizing that \( \chi = \tan^2 \theta / h \) is a measure for the strain gradient under conical
indentation, Nix and Gao [64] derived the following relationship for the yield
stress \( \sigma_y \):

\[
\left( \frac{\sigma_y}{\sigma_0} \right)^2 = 1 + \ell \chi
\]

(2.32)

where \( \sigma_0 \) is the flow stress in the absence of a strain gradient and \( \ell = b(\mu/\sigma_0)^2 \) is a
microstructural length scale. Since at the small indentation depths at hand, the
influence of tip blunting becomes significant, Tymiak et al. [67] considered strain
gratings resulting from shallow spherical indentation. Although no strong depth
dependence of the strain gradient was found, experiments and semi-empirical
calculations suggested an indentation size effect with the hardness being
proportional to \( 1/(hR)^2 \), where \( R \) is the radius of the indenter.

2.3.4 Elastic contact

Because of tip rounding, the initial contact of constant-angle indenters (such as
conical or Berkovich indenters) is well described by spherical contact up to a
transition depth \( h_t \), which may be written from geometrical considerations as [68]

\[
h_t = 2R(1 - \sin \theta)
\]

(2.33)

with \( R \) and \( \theta \) as previously defined. Typical values of \( h_t \) are of the order of 10 nm.
Hertzian contact theory [52] gives the elastic pressure distribution \( p(r) \) under a
spherical indenter as

\[
p(r) = p_0 \frac{\sqrt{a^2 - r^2}}{a}
\]

(2.34)

where \( a \) is the radius of the contact circle and \( p_0 \) is the maximum pressure (at
\( r = 0 \)). Integrating Eq. (2.34) over the contact area gives the total indentation load
\( P \).
With \( a = \pi p_0 R / 2E_r \), Eq. (2.35) may be rewritten to give the maximum pressure for a given indentation load \( P \) as

\[
p_h = \frac{3P}{2\pi a^2} = \left( \frac{6PE_r^2}{\pi R^2} \right)^{1/3}
\]  

(2.36)

By calculating the radial and tangential stress components \( \sigma_r \) and \( \sigma_\theta \) along the axis of symmetry, it can be shown that the principal shear stress \( \tau_1 = \frac{1}{2} |\sigma_1 - \sigma_2| \) has a maximum value of approximately \( 0.31p_0 \) at a depth \( z = 0.48a \) below the surface. Consequently, this is the depth at which homogeneous defect nucleation is anticipated during indentation. This has been confirmed through atomistic simulations [69,70].

Finding the components of the elastic stress field under spherical indentation at a given position off the axis of symmetry is not straightforward [52]. However, a good estimate may be obtained by considering the two-dimensional equivalent, i.e. cylindrical contact. For this situation, McEwen [71] derived the stresses in a point \((x, z)\) below the surface to be

\[
\sigma_x = -\frac{p_h}{a} \left( m \left( 1 + \frac{z^2 + n^2}{m^2 + n^2} \right) - 2z \right)
\]  

(2.37)

\[
\sigma_z = -\frac{p_h}{a} \left( 1 - \frac{z^2 + n^2}{m^2 + n^2} \right)
\]  

(2.38)

\[
\tau_x = \frac{p_h}{a} n \left( \frac{m^2 - z^2}{m^2 + n^2} \right)
\]  

(2.39)

where \( m \) and \( n \) are defined by

\[
m^2 = \frac{1}{4} \left[ \left( a^2 - x^2 + z^2 \right)^2 + 4x^2z^2 \right]^{1/2} + a^2 - x^2 + z^2
\]  

(2.40)

\[
n^2 = \frac{1}{4} \left[ \left( a^2 - x^2 + z^2 \right)^2 + 4x^2z^2 \right]^{1/2} - \left( a^2 - x^2 + z^2 \right)
\]  

(2.41)

On the axis of symmetry, the principal stresses \( \sigma_1 \) and \( \sigma_2 \) are given by Eqs. (2.37) and (2.38) with \( x = 0 \), and therefore the principal shear stress \( \tau_1 = \frac{1}{2} (\sigma_1 - \sigma_2) \) reduces to
Basic concepts

\[
\tau_i = p_0 a \left\{ z - z^2 \left( a^2 - z^3 \right)^{1/2} \right\}
\]  \hspace{1cm} (2.42)

which has a maximum value of \(0.30p_0\) for \(z = 0.78a\). Defect nucleation at this depth was indeed observed by Gouldstone et al. [72], who used a bubble raft model to give a two-dimensional representation of indentation of fcc crystals.

The relation between load \(P\) and displacement \(h\) for spherical elastic contact is given by

\[
P = \frac{1}{2} R^{1/2} E_t h^{3/2}
\]  \hspace{1cm} (2.43)

which follows from Eq. (2.36) with \(h = a / R^2\). This expression holds for rounded Berkovich indenters up to the transition depth \(h_t\). For \(h \gg h_t\), the elastic response can be described by that of an ideally sharp Berkovich indenter. Since the elastic contact pressure for constant-angle indenters has a constant value, the load-displacement relation is parabolic:

\[
P = Ch^2
\]  \hspace{1cm} (2.44)

Larsson et al. [73] found by finite element analysis that the curvature \(C\) is given by

\[
C = 2.1891E_t \left( 1 - 0.21\nu - 0.01\nu^2 - 0.41\nu^3 \right)
\]  \hspace{1cm} (2.45)

where \(\nu\) is the Poisson’s ratio of the material under indentation. It should be noted that with an ideally sharp indenter, elastic indentation of metals would be very

\[\text{Figure 2.15: Calculated elastic response of a material with } E_t = 200 \text{ GPa and } \nu = 0.3 \text{ for one sharp and three rounded Berkovich indenters.}\]
hard, if not impossible, to achieve in practice due to the very high stresses generated from the onset of contact. Due to tip rounding however, the elastic response is first described by Eq. (2.43) and only at larger depths, where tip rounding can be neglected, by Eq. (2.44). In fact, experimental results, including those presented in this thesis, often show elastic behavior up to a depth beyond \( h_c \). Figure 2.15 shows the elastic indentation response from three rounded indenters with different radii of curvature compared to that of a sharp Berkovich tip. The curves for the rounded indenters are truncated at their respective value of \( h_t \), assuming \( \theta = 70.3^\circ \) as for a cone with the same area function as a Berkovich indenter.

### 2.3.5 Incipient plasticity

The onset of plastic deformation during indentation is evidenced in the load-displacement curve by the first deviation from elastic response. For load-controlled indentation of crystalline materials, this deviation commonly has the form of a displacement burst, i.e. a plateau in the load-displacement curve where the indentation depth suddenly increases at constant indentation load as shown in Figure 2.16a. The onset of this yield excursion is defined as the yield point. The length of the excursion at a given load can be calculated, provided that the loading condition is fully plastic after indentation. The elastoplastic loading curve for a Berkovich indenter has the functional form of a parabola

\[
P = C' h^2
\]

(2.46)

where, assuming no strain hardening, the curvature \( C' \) can be related to the yield stress \( \sigma_y \) by [73]

\[
C' = 12.0348 \sigma_y \left( 1 + \ln \frac{E \tan 24.7^\circ}{3 \sigma_y} \right)
\]

(2.47)

The elastic shear stress sustained prior to the yield point is often much higher than predicted by conventional yield criteria and can even attain values close to the theoretical shear strength, as was initially observed by Gane and Bowden [74]. The physical origin of the enhanced elastic loading and the subsequent displacement burst has been the subject of extensive discussions in literature. The proposed mechanisms may roughly be divided into two main categories: (i) mechanisms that attribute the high elastic loads sustained to surface strengthening by adsorbed films (e.g. native oxides) and (ii) mechanisms based solely on the difficulty of nucleating dislocations within the small volume of
material that is elastically deformed under point contact. A great deal of experimental evidence has been produced in support of both types of mechanisms, a review of which can be found in the introduction of reference [75].

An important experimental result supporting the view of surface film controlled plasticity is the observation that the extent of elastic loading is greatly reduced when the surface is carefully cleaned to remove possible oxide or

---

**Figure 2.16**: Schematic load vs. displacement plots showing (a) a single yield excursion and (b) staircase yielding. The elastic and elastoplastic loading curves have been calculated with the parameters $E_r = 200$ GPa, $\sigma_y = 1$ GPa and $\nu = 0.3$. 
contamination layers [76,77]. Gerberich et al. [78] conducted a series of nanoindentation experiments on Fe-Si crystals, in which they found that placing a drop of HCl solution on the surface in order to dissolve the oxide film led to a contact that was plastic from the outset. Repassivation of the surface restored its original high elastic strength. This led them to the conclusion that the oxide film controls the initial plasticity and the yielding is associated with the escape of previously nucleated and piled-up dislocations to the surface upon fracture of the oxide. This view was shared by Kramer et al. [75] based on indentation of W and Fe-Si single crystals. The notion that limited dislocation activity takes place before the displacement burst implies that the initial deformation is not purely elastic but quasi-elastic, and only the subsequent collective dislocation motion leads to detectable plastic deformation.

In contrast to these findings, nanoindentation experiments on well-annealed gold single crystals [79,80] showed elastic loading up to the theoretical shear strength in the absence of significant surface contamination. Kiely and Houston [80] investigated the orientation dependence of the yield point for atomically flat (111), (001) and (110) surfaces. By resolving the applied stress on (111) slip planes, they found that for each of these orientations, plastic deformation occurs when the resolved shear stress reaches 1.8 GPa on all {111} planes that are active during deformation. These observations provide strong evidence in favor of the mechanisms that attribute the yield excursion to dislocation nucleation. The excursion can be rationalized as the nucleation of a series of dislocations from a dislocation source under the indenter, e.g. a Frank-Read source. The source may be present already prior to indentation or be produced by homogeneous nucleation of the first dislocation. In the latter case, the yield point corresponds to a shear stress close to the theoretical shear strength.

While all proposed mechanisms may be relevant in particular situations, many researchers [68,81-83] agree that the onset of permanent plastic deformation is generally controlled by dislocation nucleation and/or multiplication, and not by oxide fracture, although the presence of an oxide film may significantly affect the value of the yield point. The question whether dislocations are nucleated prior to the apparent yield point is still under debate; if so, the initial loading behavior should be classified as quasi-elastic as mentioned above.

As first observed by Corcoran et al. [79], the initial yield behavior of metals is in some cases characterized by a series of yield excursions rather than a single one. This “staircase yielding” phenomenon is illustrated in Figure 2.16b.
The excursions are separated by loading portions which are predominantly elastic, i.e. they are well described by Eq. (2.43) or (2.44), depending on the value of the indentation depth compared to $h_t$. The plasticity is thus confined to the yield excursions at this stage of deformation. An explanation of this phenomenon may be as follows: the dislocations generated from a Frank-Read source during the first excursion form a pile-up, which exerts a back force on the indenter. When the forces sum to zero, the source stops operating, and loading continues until another source is activated on a parallel slip band. This process repeats until fully plastic loading is established. Bahr et al. [82] suggested that staircase yielding occurs if the shear stress prior to the yield point is only slightly higher than the flow stress, so that upon yielding, the shear stress drops below the nucleation shear stress and further elastic loading is needed to activate the same or another dislocation source. From these viewpoints, the load at which each excursion occurs depends on the availability of dislocation sources under the indenter and on the shear stress required to nucleate dislocations from them. This accounts for the variation observed between indentations in the number of excursions and their size. Gouldstone et al. [68] found that the size of the first excursion in staircase yielding of Al thin films is well predicted by a simple energetic model, which relates the elastic energy accumulated prior to the yield point to the interaction energy of subsequently nucleated geometrically necessary dislocation loops. The fact that heterogeneous nucleation from dislocation sources, rather than homogeneous nucleation, is responsible for the repeated yielding has been confirmed through molecular dynamics simulations [83].

References
2. F.C. Frank, Philos. Mag. 42 (1951), 809
5. V. Vitek, Prog. Mater. Sci. 36 (1992), 1
Chapter 2
51. G.M. Pharr, W.C. Oliver, F.R. Broiten, J. Mater. Res. 7 (1992), 613
55. W.C. Oliver, G.M. Pharr, J. Mater. Res. 7 (1992), 1564
57. T.Y. Tsui, W.C. Oliver, G.M. Pharr, J. Mater. Res. 11 (1996), 752; A. Bolshakov, W.C. Oliver, G.M. Pharr, ibid., 760
61. N.J.M. Carvalho, Ph.D. thesis (University of Groningen, 2001)
63. Q. Ma, D.R. Clarke, J. Mater. Res. 10 (1995), 853
65. M.F. Ashby, Philos. Mag. 21 (1970), 399
71. E. McEwen, Philos. Mag. 40 (1949), 454
74. N. Gane, F.P. Bowden, J. Appl. Phys. 39 (1968), 1432
76. N. Gane, J.M. Cox, Philos. Mag. 22 (1970), 881
77. J.B. Pethica, D. Tabor, Surf. Sci. 89 (1979), 182