Chapter 4
Defects

Abstract No crystal is perfect. Various point defects and their thermodynamics, diffusion and distribution effects are discussed. Also dislocations and extended defects such as cracks, stacking faults, grain boundaries and antiphase domains are covered.

4.1 Introduction

In an ideal lattice each atom is at its designated position. Deviations from the ideal structure are called defects. In the following, we will briefly discuss the most common defects. The electrical activity of defects will be discussed in Sects. 7.5 and 7.7. For the creation (formation) of a defect a certain free enthalpy $G_{\mathrm{D}}^{\mathrm{f}}$ is necessary. At thermodynamical equilibrium a (point) defect density $\propto \exp\left(-G_{\mathrm{D}}^{\mathrm{f}}/kT\right)$ will always be present (cf. Sect. 4.2.2).

Point defects (Sect. 4.2) are deviations from the ideal structure involving essentially only one lattice point. The formation energy for line defects (Sect. 4.3) or area defects (Sect. 4.4) scales with $N^{1/3}$ and $N^{2/3}$, respectively, N being the number of atoms in the crystal. Therefore, these defects are not expected in thermodynamic equilibrium. However, the path into thermodynamical equilibrium might be so slow that these defects are metastable and must be considered quasi-frozen. There may also exist metastable point defects. By annealing the crystal, the thermodynamic equilibrium concentration might be re-established. The unavoidable two-dimensional defect of the bulk structure is the surface, discussed in Chap. 11.

4.2 Point Defects

4.2.1 Point Defect Types

The simplest point defect is a vacancy V, a missing atom at a given atomic position. If an atom is at a position that does not belong to the crystal structure an interstitial I

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7_4

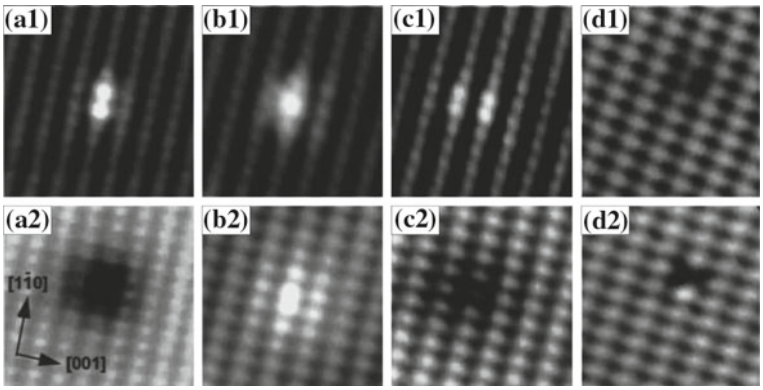

Fig. 4.1 Images of occupied (*upper frames*) and empty (*lower frames*) density of states of typical defects on Si-doped GaAs (110) surfaces. (**a1**, **a2**) show a Ga vacancy, (**b1**, **b2**) a Si_{Ga} donor, (**c1**, **c2**) a Si_{As} acceptor and (**d1**, **d2**) a Si_{Ga}–V_{Ga} complex. Adapted from [227]

(or Frenkel defect) is formed. Depending on the position of the interstitial different types are distinguished. An interstitial atom that has the same chemical species as the crystal is called 'self-interstitial'.

If an atom site is populated with an atom of different order number Z, an impurity is present. An impurity can also sit on interstitial position. If the number of valence electrons is the same as for the original (or correct) atom, then it is an isovalent impurity and quasi fits into the bonding scheme. If the valence is different, the impurity adds extra (negative or positive) charge to the crystal bonds, which is compensated by the extra, locally fixed charge in the nucleus. This mechanism will be discussed in detail in the context of doping (Chap. 7). If in an AB compound an A atom sits on the B site, the defect is called an antisite defect A_B.

A Ga vacancy, a silicon impurity atom on Ga- and As-site and a Si_{Ga}-vacancy complex at the (110) surface of Si doped GaAs are shown in Fig. 4.1 as observed with STM [227, 228]. Also antisite defects in GaAs can be observed with STM [229, 230].

A point defect is typically accompanied by a relaxation of the surrounding host atoms. As an example, we discuss the vacancy in Si (Fig. 4.2a). The missing atom leads to a lattice relaxation with the next neighbors moving some way into the void (Fig. 4.2b). The bond lengths of the next and second-next neighbor Si atoms around the neutral vacancy are shown in Fig. 4.2c. The lattice relaxation depends on the charge state of the point defect (Jahn–Teller effect) which is discussed in more detail in Sect. 7.7. In Fig. 4.2d the situation for the positively charged vacancy with one electron missing is shown. One of the two bonds is weakened since it lacks an electron. The distortion is therefore different from that for V^0. Also the (self-)interstitial is accompanied with a lattice relaxation as shown in Fig. 4.3 for a silicon interstitial at tetrahedral place. Self-interstitials in silicon and germanium are reviewed and compared in [231] for their various charge states.

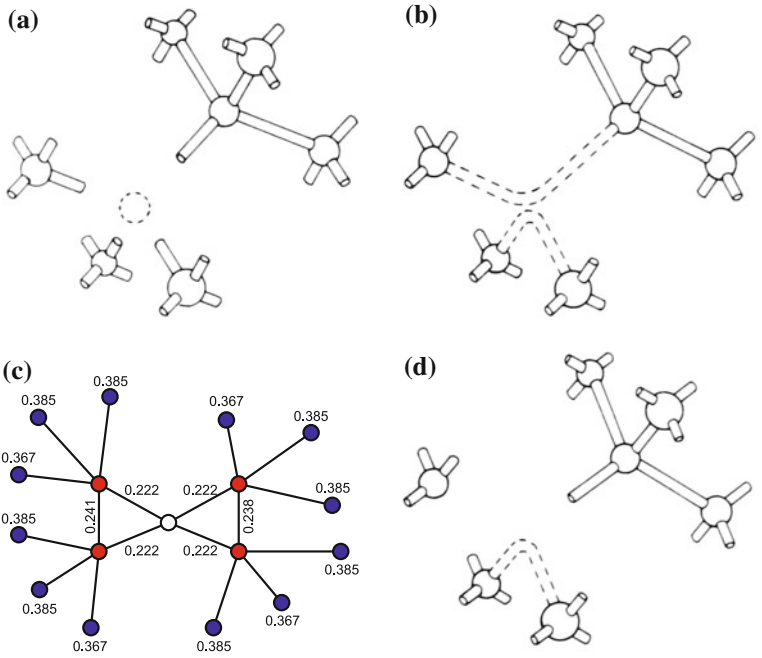

Fig. 4.2 (**a**) Schematic diamond lattice with vacancy, i.e. a missing Si atom without relaxation. (**b**) Si with neutral vacancy (V^0), lattice relaxation and formation of two new bonds. (**c**) Schematic diagram showing the (inward) relaxation of the neighbors around the neutral Si vacancy defect site (*empty circle*) calculated by an ab initio method. The distances of the outer shell of atoms (*red circles*) from the vacant site is labeled (in nm). The bond lengths of the two new bonds and the second-neighbor (*blue circles*) distance are also indicated. The bond length in bulk Si is 0.2352 nm, the second-neighbor distance 0.3840 nm. Adapted from [232]. (**d**) Si unit cell with positively charged vacancy (V^+). Parts (**a**, **b**, **d**) reprinted with permission from [233]

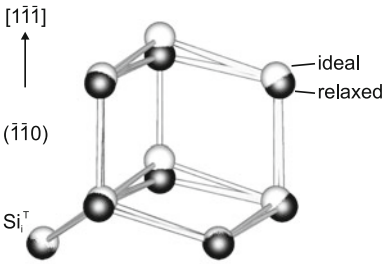

Fig. 4.3 Silicon tetrahedral interstitial Si_i^T and its next atoms in ideal (*white spheres*) and relaxed (*black spheres*) position. Adapted from [155]

4.2.2 Thermodynamics

For a given temperature, the free enthalpy G of a crystal (a closed system with regard to particle exchange)

$$G = H - TS \tag{4.1}$$

is minimum. H is the enthalpy and S the entropy. The enthalpy $H = E + pV$ is the thermodynamic potential for a system whose only external parameter is the volume V. It is used when the independent variables of the system are the entropy S and pressure p. The free enthalpy is used when the independent parameters are T and p. G_0 (H_0) is the free energy (enthalpy) of the perfect crystal. H^{f} is the formation enthalpy of an isolated defect. This could be, e.g., the enthalpy of a vacancy, created by bringing an atom from the (later) vacancy site to the surface, or an interstitial, created by bringing an atom from the surface to the interstitial site. In the limit that the n defects do not interact with each other, i.e. their concentration is sufficiently small, they can be considered independent and the enthalpy is given by

$$H = H_0 + n\,H^{\mathrm{f}}. \tag{4.2}$$

The increase of entropy due to increased disorder is split into the configurational disorder over the possible sites, denoted as S^{d}, and the formation entropy S_{f} due to localized vibrational modes. The total change ΔG of the free energy is

$$\Delta G = G - G_0 = n\,(H^{\mathrm{f}} - T\,S^{\mathrm{f}}) - T\,S^{\mathrm{d}} = n\,G^{\mathrm{f}} - T\,S^{\mathrm{d}}, \tag{4.3}$$

where $G^{\mathrm{f}} = H^{\mathrm{f}} - T\,S^{\mathrm{f}}$ denotes the free enthalpy of formation of a single isolated defect. In Table 4.1 experimental values for the formation entropy and enthalpy are given for several defects. Surprisingly, despite their fundamental importance in semiconductor defect physics, these numbers are not very well known and disputed in the literature.

The defect concentration is obtained by minimizing ΔG, i.e.

$$\frac{\partial \Delta G}{\partial n} = G^{\mathrm{f}} - T\,\frac{\partial S^{\mathrm{d}}}{\partial n} = 0. \tag{4.4}$$

The entropy S^{d} due to disorder is given as

$$S^{\mathrm{d}} = k_{\mathrm{B}} \ln W, \tag{4.5}$$

where W is the complexion number, usually the number of distinguishable ways to distribute n defects on N lattice sites

Table 4.1 Formation enthalpy H^{f} and entropy S^{f} of the interstitial (I) and vacancy (V) in Si and the Ga vacancy in GaAs

Material	Defect	H^{f} (eV)	S^{f} (k_{B})
Si	I	3.2	4.1
Si	V	2.8	~ 1
GaAs	V_{Ga}	3.2	9.6

Data for Si from [234, 235], for GaAs from [236]

$$W = \binom{N}{n} = \frac{N!}{n!\,(N-n)!}. \tag{4.6}$$

With Stirling's formula $\ln x! \approx x(\ln x - 1)$ for large x we obtain

$$\frac{\partial S^{\mathrm{d}}}{\partial n} = k_{\mathrm{B}} \left[\frac{N}{n} \ln\left(\frac{N}{N-n}\right) + \ln\left(\frac{N-n}{n}\right) \right]. \tag{4.7}$$

If $n \ll N$, $\partial N/\partial n = 0$ and the right side of (4.7) reduces to $k_{\mathrm{B}} \ln(N/n)$. The condition (4.4) reads $G^{\mathrm{f}} + k_{\mathrm{B}} T \ln(n/N)$, or

$$\frac{n}{N} = \exp\left(-\frac{G^{\mathrm{f}}}{kT}\right). \tag{4.8}$$

In the case of several different defects i with a degeneracy Z_i, e.g. a spin degree of freedom or several equivalent configurations, (4.8) can be generalized to

$$\frac{n_i}{Z_i\, N} = \exp\left(-\frac{G_i^{\mathrm{f}}}{kT}\right). \tag{4.9}$$

In [237] the equilibrium concentration of interstitials C_I^{eq} in silicon has been given as

$$C_I^{\mathrm{eq}} = \left(1.0 \times 10^{27}\,\mathrm{cm}^{-3}\right) \exp\left(-\frac{3.8\,\mathrm{eV}}{kT}\right), \tag{4.10}$$

about $10^{14}\,\mathrm{cm}^{-3}$ at 1200 °C. The vacancy concentration has been investigated in [238]. Around a temperature of 1200 °C it is in the 10^{14}–$10^{15}\,\mathrm{cm}^{-3}$ range. Due to the reaction

$$0 \rightleftharpoons I + V, \tag{4.11}$$

a mass action law holds for the concentrations of interstitials and vacancies

$$C_I\, C_V = C_I^{\mathrm{eq}}\, C_V^{\mathrm{eq}}. \tag{4.12}$$

4.2.3 Diffusion

The diffusion of point defects is technologically very important, in particular for silicon as host material. Typically a dopant profile should be stable under following technological processing steps and also during device performance. Also defect annihilation is crucial after implantation processes. Diffusion of an interstitial I and a

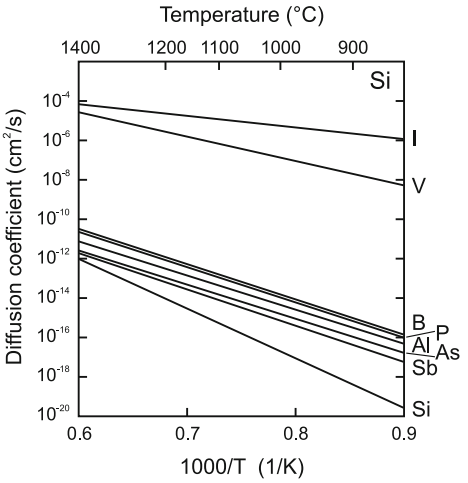

Fig. 4.4 The temperature dependent diffusion coefficient of Si interstitials I, vacancies V and various impurities in silicon. Also the self-diffusion coefficient, labeled with 'Si' is shown. Based on data from [239]

vacancy V to the same site is prerequisite for recombination of defects (so called *bulk process*) according to the scheme $I + V \rightarrow 0$. We note that the process $0 \rightarrow I + V$ is called *Frenkel pair process*.[1] Also the self-diffusion of silicon has been studied, e.g. using radioactively marked isotopes [237]. The diffusion of point defects including dopants in silicon has been reviewed in [239, 240]. Usually Fick's law is applied, stating how the flux J depends on the concentration gradient, for an interstitial it reads:

$$J_I = -D_I \nabla C_I, \tag{4.13}$$

D_I being the interstitial diffusion coefficient. For interstitials in Si it was found [237] that

$$D_I = 0.2 \exp\left(-\frac{1.2\,\text{eV}}{kT}\right) \text{cm}^2/\text{s}. \tag{4.14}$$

The diffusion of neutral vacancies occurs with [241]

$$D_V = 0.0012 \exp\left(-\frac{0.45\,\text{eV}}{kT}\right) \text{cm}^2/\text{s}. \tag{4.15}$$

The temperature dependent diffusion coefficients of point defects and dopants in silicon are shown in Fig. 4.4.

The self-diffusion coefficient of silicon has been determined from the annealing of isotope superlattices (Sect. 12.5) of sequence $^{28}\text{Si}_n/^{30}\text{Si}_n$, $n = 20$ to be [242]

[1] At higher temperatures a silicon atom can occasionally acquire sufficient energy from lattice vibrations to leave its lattice site and thus an interstitial and a vacancy are generated.

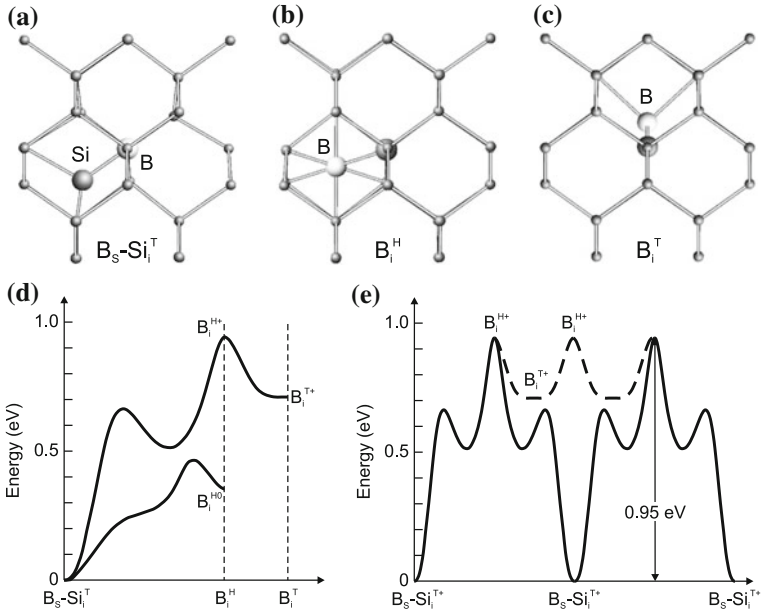

Fig. 4.5 Configurations of boron in Si: (**a**) Substitutional boron and Si self-interstitial at 'T' site (B_S–Si_i^T). Interstitial boron at (**b**) 'H' (B_i^H) and (**c**) 'T' site (B_i^T), each with the Si atoms on the Si lattice sites. The *large bright ball* represents the boron atom, *large* and *small dark balls* represent Si atoms. (**d**) Lowest energy barrier diffusion paths for positively charged and neutral B–Si states, total energy versus configuration. (**e**) Two diffusion pathways for positively charged B–Si, kick-out (*dashed line*) and pair diffusion (*solid line*); the activation energy is labeled. Adapted from [245]

$$D_{\mathrm{Si}}^{\mathrm{SD}} = \left[2175.4\ \exp\left(-\frac{4.95\,\mathrm{eV}}{kT}\right) + 0.0023\ \exp\left(-\frac{3.6\,\mathrm{eV}}{kT}\right)\right]\ \mathrm{cm}^2/\mathrm{s}, \quad (4.16)$$

the first (second) term being due to interstitial (vacancy) mechanism, dominant for temperatures larger (smaller) than 900 °C. The enthalpy in the exponent, e.g. $H_V = 3.6^{+0.3}_{-0.1}$ eV [242], consists of the formation and migration enthalpies,

$$H_V = H_V^{\mathrm{f}} + H_V^{\mathrm{m}}. \quad (4.17)$$

Using the experimental value $H_V^{\mathrm{f}} = 2.8 \pm 0.3$ eV [235] from Table 4.1, for the migration enthalpy a value around $H_V^{\mathrm{m}} \approx 0.8$ eV is obtained.

As an example for a dopant diffusion process that has been understood microscopically, we discuss here boron in silicon. In Fig. 4.5a the lowest-energy configuration of a boron-related defect in silicon is depicted, B_s–Si_i^T, i.e. boron on a substitutional site and a self-interstitial Si on the 'T' place with highest symmetry[2] (see Fig. 3.18).

[2]The positive charge state is stable, the neutral charge state is metastable since the defect is a negative-U center (see Sect. 7.7.5).

Due to its importance as an acceptor in Si, the configuration and diffusion of B in Si has found great interest [243–245]. The diffusion depends on the charge state of boron. The diffusion of positively charged boron has been suggested [245] to occur via the following route: The boron leaves its substitutional site and goes to the hexagonal site ('H') (Fig. 4.5b) with an activation energy of about 1 eV (Fig. 4.5d). It can then relax (∼0.1 eV) without barrier to the tetrahedral 'T' position (Fig. 4.5c). The direct migration B_s–$Si_i^{T+} \rightarrow B_i^{T+}$ has a higher activation energy of 1.12 eV and is thus unlikely. The boron atom can then diffuse through the crystal by going from 'H' to 'T' to 'H' and so on (Fig. 4.5e). However, long-range diffusion seems to be not possible in this way because the kick-in mechanism will bring back the boron to its stable configuration. The pair diffusion mechanism for neutral boron B_s–$Si_i^T \rightarrow B_i^H \rightarrow B_s$–$Si_i^T$ via the hexagonal site has an activation energy of about 0.5 eV (Fig. 4.5d) while the path via B_i^T has a larger 0.9 eV barrier. The concentration dependence of the diffusion mechanism has been discussed in [246].

Similarly, indium diffusion in silicon has been investigated suggesting a minimum energy In_s–$Si_i^T \rightarrow In_i^T \rightarrow In_s$–$Si_i^T$ diffusion pathway via the tetrahedral site with 0.8 eV activation energy [247]. Microscopic modeling has been reported also for diffusion of phosphorus [248].

4.2.4 Dopant Distribution

The introduction of impurities into a semiconductor (or other materials such as glasses) is termed *doping*. The unavoidable incorporation of impurities in the nominally pure (nominally undoped) material is called *unintentional* doping and leads to a residual or background impurity concentration. Several methods are used for doping and the creation of particular doping profiles (in depth or lateral). All doping profiles underly subsequent diffusion of dopants (Sect. 4.2.3).

Various methods of doping are used. A straightforward method of doping is the incorporation during crystal growth or epitaxy. For semiconductor wafers a homogeneous doping concentration is targeted, both laterally and along the rod from which the wafer is cut (Sect. 12.2.2). When a crystal is grown from melt, containing a concentration c_0 of the impurity, the concentration in the solid is given by ('normal freezing' case [249–251])[3]

$$c(x) = c_0\, k\, (1-x)^{k-1}, \tag{4.18}$$

where $c(x)$ is the impurity concentration in the crystal at the freezing interface, x is the frozen melt fraction (ratio of solid mass to total mass, $0 \le x \le 1$). k is the *distribution*

[3] Mass preservation of the impurities can be written at any time $c_m(1-x) + \int_0^x c(x')\,dx' = c_0$, where c_m is the (remaining) concentration in the melt. At the beginning $c_m(0) = c_0$. At the interface $c(x) = k\, c_m(x)$. Putting this into the mass preservation, building $c'(x)$ and solving the resulting differential equation $c' = c(1-k)/(1-x)$ with $c(0) = k\, c_0$ leads to (4.18).

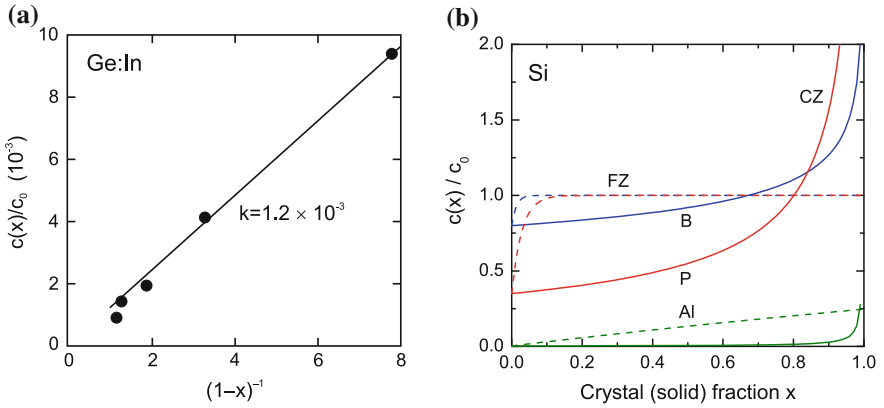

Fig. 4.6 (**a**) Relative concentration of indium along a CZ-grown germanium crystal. Absolute concentration is in the $10^{16}\,\mathrm{cm}^{-3}$ range. *Solid line* follows (4.19) with $k = 1.2 \times 10^{-3}$. *Symbols* are experimental data from [254]. (**b**) Impurity distribution (relative concentration $c(x)/c_0$) for CZ (4.18) (*solid lines*) and FZ (4.20) (*dashed lines*, $z = 0.01$) silicon crystals for B (*blue*), P (*red*), and Al (*green*). Distribution coefficients have been taken from Table 4.2. Note crossing of B and P lines and possibly associated change from p-type to n-type (cmp. Fig. 1.7)

coefficient (or segregation coefficient) which is the fraction of impurities that is built into the crystal at the liquid–solid interface. Since the melt volume reduces during the solidification, the impurity concentration rises over time. For small distribution coefficients (4.18) can be approximated to

$$c(x) \approx c_0 \frac{k}{1-x}\,, \tag{4.19}$$

An experimental example for Ge:In is shown in Fig. 4.6a.

In Table 4.2 the distribution coefficients for various impurities in Si, Ge and GaAs is given. The modification of distribution coefficients in SiGe alloys is discussed in [252]. Equilibrium values (k_{eq}) are obtained for 'slow' crystal growth. For finite growth rates, k becomes a function of the growth rate and is then called the *effective* distribution coefficient. For $k < 1$, $k_{\mathrm{eff}} > k_{\mathrm{eq}}$. k_{eff} approaches 1 for high growth rates, i.e. all impurities at the rapidly moving interface are incorporated.

Equation (4.18) applies to Czrochalski growth where the crystal is pulled out of the melt [257]. In float-zone (FZ) growth [257] a polycrystalline rod is transformed into a crystalline one while a RF-heated and liquid 'float' zone is moved through the crystal. In this case the impurity distribution is given by[4]

[4]When the float zone moves through the crystal, the change of mass of impurities $m_{\mathrm{m}} = c_{\mathrm{m}} z$ in the liquid is $m'_{\mathrm{m}} = c_0 - k c_{\mathrm{m}}$. The first term stems from the melting of the polycrystalline part, the second from the solidification of the crystal. Solving the resulting differential equation $c'_{\mathrm{m}} = (c_0 - k c_{\mathrm{m}})/z$ with $c_{\mathrm{m}}(0) = c_0$ and using $c(x) = k c_{\mathrm{m}}(x)$ yields (4.20).

Table 4.2 Equilibrium distribution coefficients (at melting point) of various impurities in silicon, germanium and GaAs

Impurity	Si	Ge	GaAs
C	0.07	>1.85	0.8
Si		5.5	0.1
Ge	0.33		0.03
N	7×10^{-4}		
O	≈1		0.3
B	0.8	12.2	
Al	2.8×10^{-3}	0.1	3
Ga	8×10^{-3}	0.087	
In	4×10^{-4}	1.2×10^{-3}	0.1
P	0.35	0.12	2
As	0.3	0.04	
Sb	0.023	3.3×10^{-3}	<0.02
S	10^{-5}	$> 5 \times 10^{-5}$	0.3
Fe	6.4×10^{-6}	3×10^{-5}	2×10^{-3}
Ni	$\approx 3 \times 10^{-5}$	2.3×10^{-6}	6×10^{-4}
Cu	8×10^{-4}	1.3×10^{-5}	2×10^{-3}
Ag	$\approx 1 \times 10^{-6}$	10^{-4}	0.1
Au	2.5×10^{-5}	1.5×10^{-5}	
Zn	2.5×10^{-5}	6×10^{-4}	0.1

Data for Si from [251, 253], for Ge from [146, 254–256] and for GaAs from [146]

$$c(x) = c_0 \left[1 - (1 - k) \exp\left(-\frac{k\,x}{z} \right) \right], \tag{4.20}$$

where x is the ratio of the crystal mass to the total mass, i.e. crystal, liquid and feed rod. z is the relative mass of the (liquid) float zone, i.e. the ratio of liquid mass to the total mass. The impurity distribution for CZ and FZ crystals is compared in Fig. 4.6b. Obviously the FZ process can create much more homogeneous profiles.[5]

Using epitaxy arbitrary doping profiles along the growth directions can be created by varying the impurity supply during growth. Impurities can be introduced through the surface of the material by diffusion from a solid or gas phase. In ion implantation [258] the impurity atoms are accelerated towards the semiconductor and deposited with a certain depth profile due to multiple scattering and energy loss events, depending on the acceleration voltage (increasing deposition depth with increasing voltage, Fig. 4.7a and ion mass (decreasing deposition depth with increasing mass, Fig. 4.7b. The depth profile is often investigated using secondary ion mass

[5]We note that during directed solidification of Si:(B,P) a pn-junction forms due to the different distribution coefficients of boron and phosphorus. This has been used in [79].

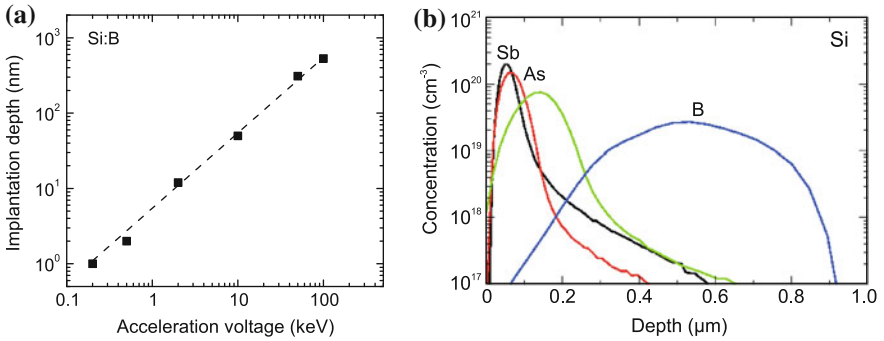

Fig. 4.7 (**a**) Depth of peak concentration of boron implanted in silicon for various acceleration voltages U. Data from various sources, for $U < 1$ keV from [264]. *Dashed line* is linear dependence. (**b**) Simulated depth profiles of impurity concentration for B, P, As, and Sb implanted into crystalline silicon with $U = 100$ keV and a dose of 10^{15} cm^{-2}. Adapted from [265]

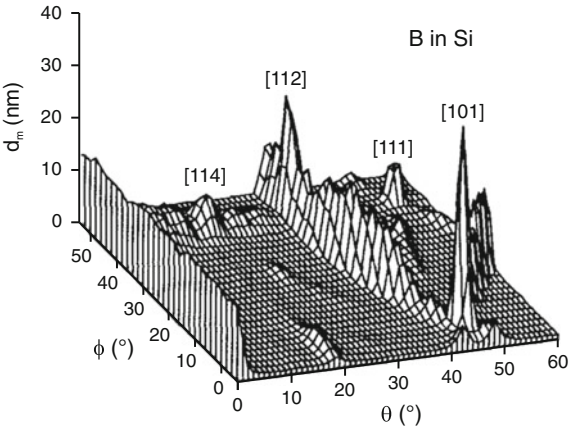

Fig. 4.8 Simulated mean path length as a function of implantation direction (azimuthal angle ϕ and polar angle θ) near [001] for 5 keV boron in silicon. The [001] channeling peak appears as a ridge at the left side of the plot ($\theta = 0$, any value of ϕ). Adapted from [261]

spectrometry (SIMS) [259]. The profile also depends on the matrix material whose stopping power depends on its density and atomic mass. While an implantation depth of about 50 nm is reached for boron in silicon ($A \approx 28$) for 10 keV, 20 keV are necessary in germanium ($A \approx 72.6$) [260]. The mean path length[6] d_m depends also on the crystallographic direction (channeling effects, Fig. 4.8) [261]. A simulation of the interaction of ions and solids can be performed using the SRIM software [262, 263].

[6]The mean path length is the distance integrated along the ion trajectory until its direction deviates by more than 4° from the incident direction.

4.2.5 Large Concentration Effects

Lattice Constant

At high doping concentration, a noticeable effect on the lattice constant a_0 is found. For silicon the atom density[7] is $N_{Si} = 5 \times 10^{22}\,cm^{-3}$. A doping level of $N = 10^{19}\,cm^{-3}$ corresponds thus to a dopant fraction of 0.02 %. Such crystal could also be considered a very dilute alloy. About each $(N_{Si}/N)^{1/3} \approx 17$th atom in a given direction is a dopant.

The effect of high doping on the lattice constant is due to different ionic radius of the dopant and the hydrostatic deformation potential of the band edge occupied by the free carriers [266]. In a linear approach, the effect is summarized in the coefficient β via

$$\beta = \frac{1}{N}\frac{\Delta a_0}{a_0}. \tag{4.21}$$

The effect due to charge carriers on β is negative (positive) for p-doping (n-doping). Experimental data for Si, Ge, GaAs and GaP are compiled in [267, 268] and theoretically discussed. The effect is in the order of $\beta = \pm(1 - 10) \times 10^{-24}\,cm^3$. For example, in the case of Si:B, the shrinkage of the lattice constant is mostly due to the charge carrier effect, for Si:P both effects almost cancel. In [269] it is shown that boron incorporation in silicon changes the lattice constant in various directions quite differently, e.g. d_{333} is shrunk by 0.4 % for a doping level of $10^{19}\,cm^{-3}$ while the {620} lattice constant remains constant.

Clustering

Point defects can cluster, i.e. several point defects aggregate at neighboring sites. An example the configuration of five nearby vacancies in silicon, the so-called V_5 cluster is shown in Fig. 4.9a. In [270] the ring-like hexavacancy in silicon is predicted a very stable defect. A large number of clustered vacancies is equivalent to a void. An example is depicted in Fig. 4.9b for an In_2O_3 crystal which has locally 'decomposed' into an indium particle and a void as revealed by TEM [271]. Also impurities can exhibit clustering (see Sect. 4.2.5).

Typically a random distribution of dopants in the host is assumed (cmp. Sect. 3.7.1 on random alloys). The introduction of several impurities can lead to pairing effects, e.g. described for Se and B, Ga, Al or In in silicon [273]. A high concentration of a single impurity makes the existence of clusters, i.e. two or more neighboring dopant atoms, more probable. This effect has been extensively studied for B in Si [274], showing that several boron atoms with interstitials I form thermodynamically stable clusters, e.g. B_3I_2. This cluster forms from B_2I and BI with only 0.2 eV activation barrier [275] as shown in Fig. 4.10. The formation is limited by diffusion of the smaller clusters to the same site. The number of free carriers (here holes) released

[7] Eight atoms per cubic unit cell of length $a_0 = 0.543\,nm$.

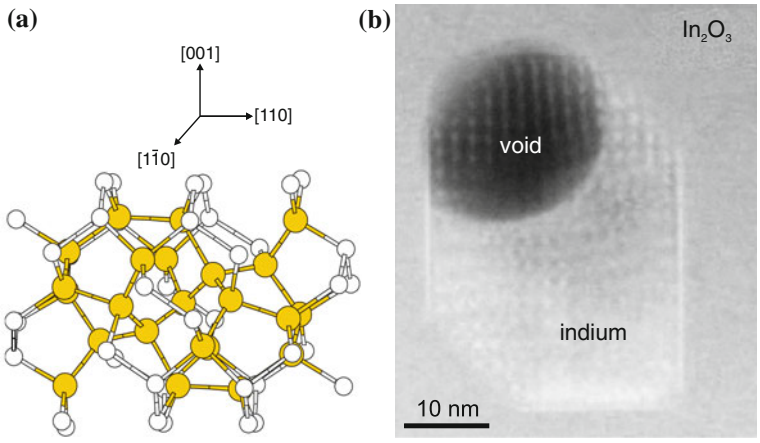

Fig. 4.9 (**a**) Predicted configuration of the V_5-cluster (five vacancies) in silicon. *Yellow spheres* indicate more distorted atoms than the rest of the lattice atoms (*white spheres*). Adapted from [272]. (**b**) Indium particle with adjacent void embedded in In_2O_3 (STEM image revealing Z-contrast in [001]). Adapted from [271]

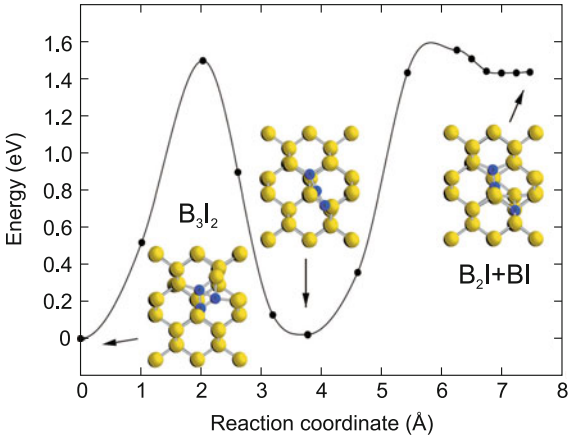

Fig. 4.10 Minimum energy path for the breakup of a B_3I_2 cluster into B_2I and BI. Silicon (boron) atoms are shown as *yellow* (*blue*) *spheres*. Adapted from [275]

from such cluster is smaller than the number of boron atoms since it forms a deep acceptor [274]. This autocompensation mechanism is thus limiting the maximum achievable free carrier concentration due to doping and is technologically unfavorable. Reactions between boron atoms and silicon self-interstitials often lead to boron clustering in the peak region of an implantation profile and require detailed optimization of the annealing process [276].

Solubility Limit

The steady-state impurity solubility can be defined as the maximum concentration of impurity atoms in a crystal allowing thermodynamic balance between the crystal and

Table 4.3 Maximum solubility N_s of some impurities in silicon

Impurity	N_s (10^{20} cm^{-3})
B	4
P	5
As	4
Sb	0.7
Al	0.13
Cu	1.4×10^{-2}
Au	1.2×10^{-3}
Fe	3×10^{-4}

Data for B, P, As, Sb from [278], other data from [281]

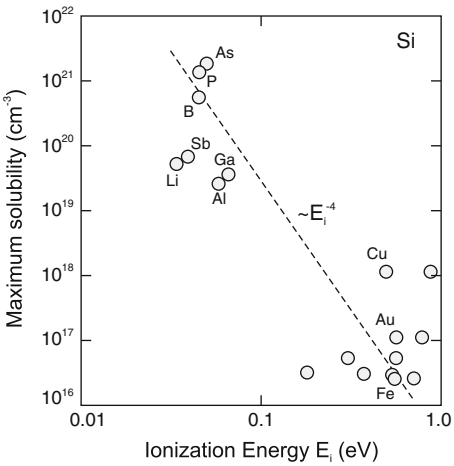

Fig. 4.11 Solubility limit for various impurities in silicon versus their ionization energy. Adapted from [279]

another phase, e.g. a liquid phase, an extended defect or a precipitate. Precipitates are small inclusions of a second phase in a crystal, exhibiting a high concentration of 'gathered' impurities that cannot be solved in the crystal. Solubility limits for impurities in silicon have been first determined in [277] with a bulk of subsequent research [278] due to its practical relevance in device fabrication. The solubility limits for a few impurities in silicon are listed in Table 4.3. It is related to the ionization energy of the defect (cmp. Sect. 7.4) as shown in Fig. 4.11.

The temperature dependence of the solubility for a few dopants is depicted in Fig. 4.12a. The solubility depends also on the present strain [280]. The simple empirical relation $x_s = 0.1\,k$ (Fig. 4.12b) between the maximum molar solubility x_s and the distribution coefficient k in silicon and germanium has been pointed out in [281].

A typical example for the formation of precipitates is Fe in InP, used for compensation of shallow donors in order to produce semi-insulating material (Sect. 7.7.8). The solubility of Fe in InP is fairly low, about 10^{17} cm^{-3} at growth temperature [282]. In Fig. 4.13 a high-resolution TEM image of a precipitate in InP doped with

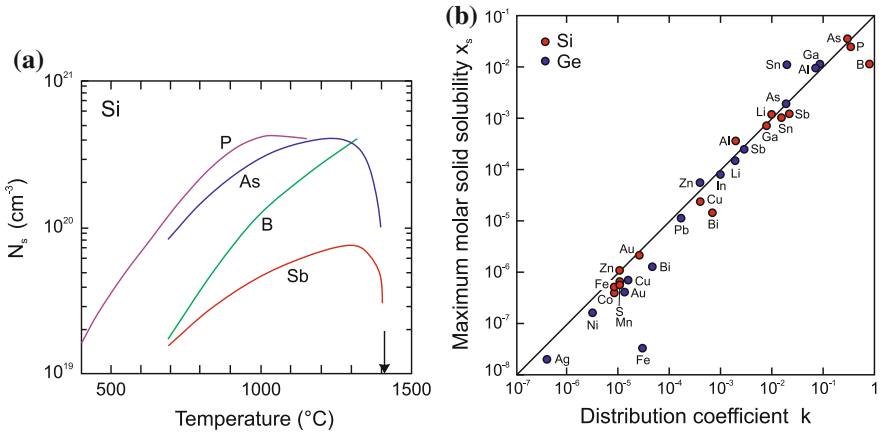

Fig. 4.12 (**a**) Steady-state solubility of impurities (P, As, B and Sb as labeled) in silicon. *Solid lines* are theoretical model matching various experimental data. *Arrow* denotes the melting point of silicon (1410°). Adapted from [278]. (**b**) Maximum molar solid solubility x_s versus the distribution coefficient for various impurities in crystalline silicon and germanium. *Solid line* follows $x_s = 0.1\,k$. Adapted from [281]

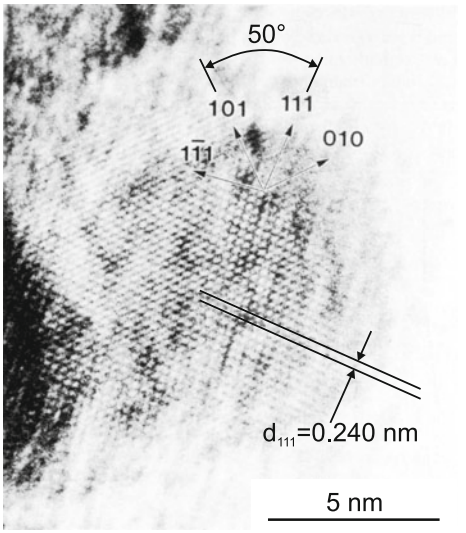

Fig. 4.13 High resolution TEM image of a FeP precipitate in iron-doped InP. Adapted from [283]

$3 \times 10^{18}\,\mathrm{cm}^{-3}$ Fe is shown. The precipitate exhibits a lattice constant of $d_{111} = 0.240\,\mathrm{nm}$ in [111]-direction, much different from that of InP ($d_{111}^{\mathrm{InP}} = 0.339\,\mathrm{nm}$). The angle between the [101] and [111] direction is 50° instead of 35° for InP. This is consistent with orthorhombic FeP [283]. Typically FeP and FeP_2 precipitates are found in highly Fe-doped InP [284].

4.3 Dislocations

Dislocations are line defects along which the crystal lattice is shifted by a certain amount. The vector along the dislocation line is called line vector **L**. A closed path around the dislocation core differs from that in an ideal crystal. The difference vector is called the Burger's vector **b**. Dislocations for which the Burger's vector is a vector of the lattice are called *full* dislocations. In contrast, dislocations with Burger's vectors that are not translation vectors of the lattice are called *partial* dislocations. The history of dislocation theory is described in [285].

Since the energy of a dislocation is proportional to $\mathbf{b}^2$, only dislocations with the shortest Burger's vector are stable. The plane spanned by **L** and **b** is called the *glide* plane. In Fig. 4.14 a high-resolution image of the atoms around a dislocation and the phase and amplitude of the (111) reflection are shown. The phase corresponds to the atomic columns, the amplitude to the displacement of the atoms at the dislocation core (see also Fig. 5.34).

4.3.1 Dislocation Types

Edge Dislocations

For an *edge* dislocation (Fig. 4.15a) **b** and **L** are perpendicular to each other. An extra half-plane spanned by **L** and $\mathbf{b} \times \mathbf{L}$ is inserted (Fig. 4.16).

Screw Dislocations

For a *screw* dislocation (Fig. 4.15b) **b** and **L** are collinear. The solid has been cut along a half-plane up to the dislocation line, shifted along **L** by the amount **b** and reattached.

Around the intersection of a screw dislocation with a surface, the epitaxial growth occurs, typically in the form of a growth spiral that images the lattice planes around the defect.

60° Dislocations

The most important dislocations in the zincblende lattice (Fig. 4.17) have the line vector along $\langle 110 \rangle$. With the Burger's vector $a/2\,\langle 110 \rangle$ three different types of dislocations can be formed: edge, screw and 60° dislocations. The vicinity of the core of the latter is shown in more detail in Fig. 4.17d. We note that the atomistic structure of 60° dislocations is different for **L** along [110] and [−1 10]; depending on whether the cations or anions are in the core, they are labeled α or β dislocations.

Misfit Dislocations

When materials with different lattice constants are grown on top of each other, the strain can plastically relax via the formation of misfit dislocations. A typical network of such dislocations is shown in Fig. 4.18 for SiGe on Si.

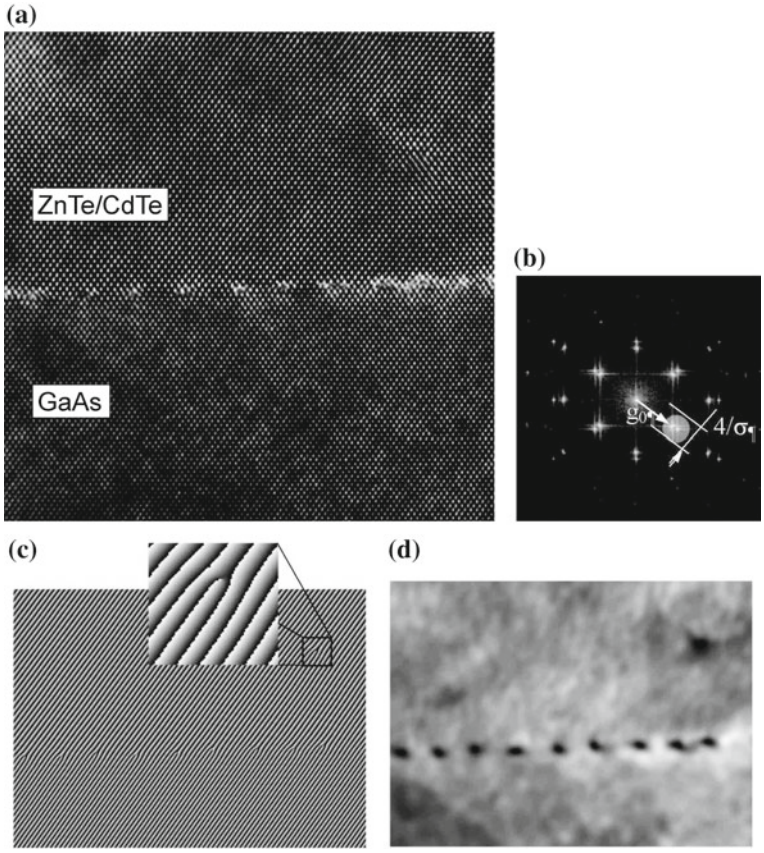

Fig. 4.14 (**a**) High-resolution transmission electron microscopy image (HRTEM) in the ⟨110⟩ projection of a network of misfit dislocations at a GaAs/CdTe/ZnTe interface. Substrate: GaAs (001), 2° off ⟨110⟩, ZnTe buffer layer is 2 monolayers thick. (**b**) Fourier transform with round mask around the (111) Bragg reflection. (**c**) Phase and (**d**) amplitude images for the mask from (**b**). From [286]

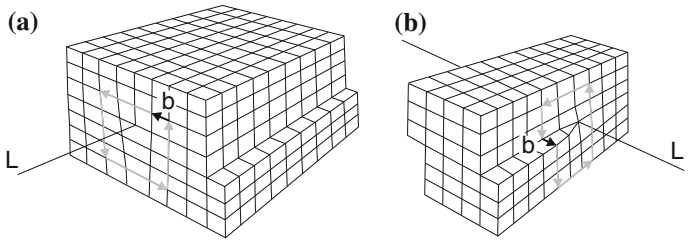

Fig. 4.15 Model of (**a**) an edge and (**b**) a screw dislocation. The line vector **L** and the Burger's vector **b** are indicated

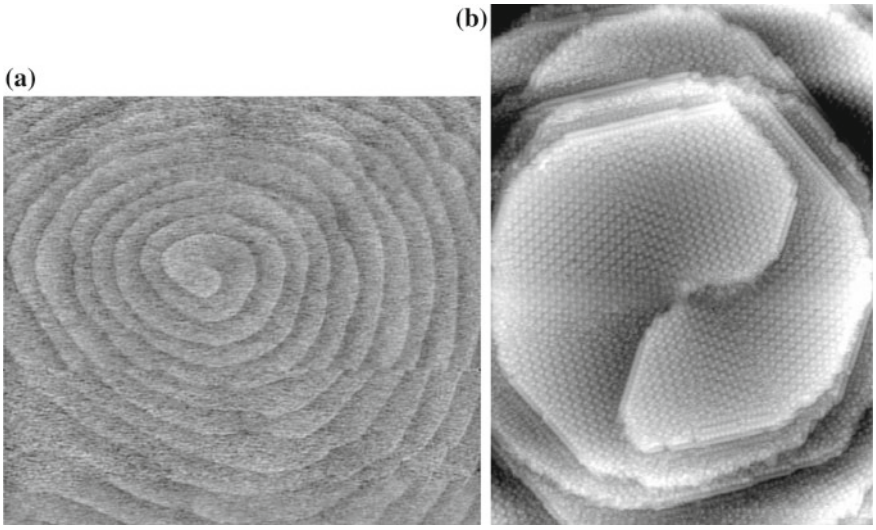

Fig. 4.16 (**a**) Atomic force microscopy image of growth spiral around a screw dislocation on a silicon surface; image width: 4 μm. (**b**) STM image (width: 75 nm) of a screw-type dislocation with a Burgers vector of [000-1] on the N-face of GaN. The reconstruction is c(6 × 12). The c(6 × 12) row directions correspond to $\langle \bar{1}100 \rangle$. Reprinted with permission from [287], © 1998 AVS

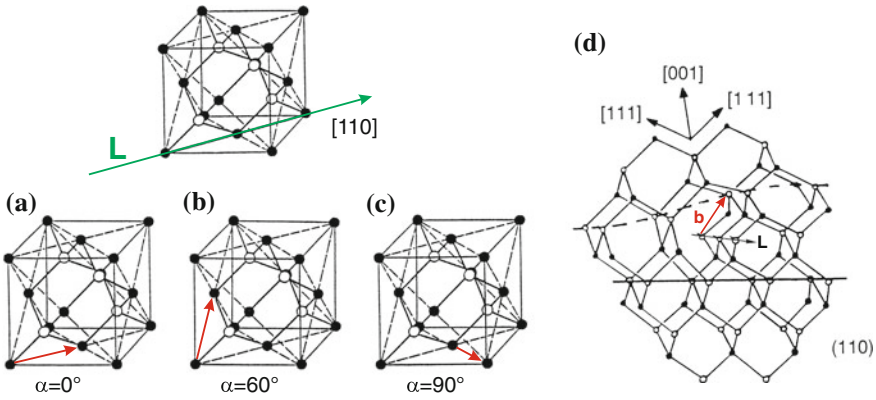

Fig. 4.17 Dislocations in the zincblende structure. The line vector is along [100]. The Burger's vector $a/2\,\langle 110 \rangle$ can create an (**a**) edge dislocation, a (**b**) screw dislocation, and (**c**) a 60° dislocation. (**d**) Atomistic structure of a 60° dislocationeps

Partial Dislocations

Partial dislocations, i.e. the Burger's vector is not a lattice vector, must necessarily border a two-dimensional defect, usually a stacking fault (Sect. 4.4.2). A typical partial dislocation in diamond or zincblende material is the Shockley partial dislocation (or just Shockley partial) with Burger's vector $b = (a_0/6)\,\langle 112 \rangle$. Another important partial is the Frank partial with $b = (a_0/3)\,\langle 111 \rangle$. A perfect dislocation

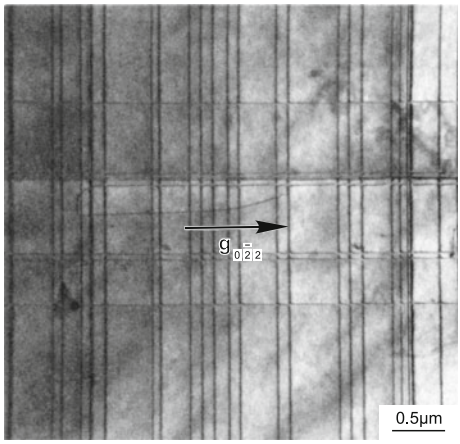

Fig. 4.18 Plan-view transmission electron microscopy image of a network of ⟨110⟩ dislocation lines in InGaAs on InP (001) with a lattice mismatch of about 0.1 %. The TEM diffraction vector is $\mathbf{g} = [2\bar{2}0]$. Adapted from [288]

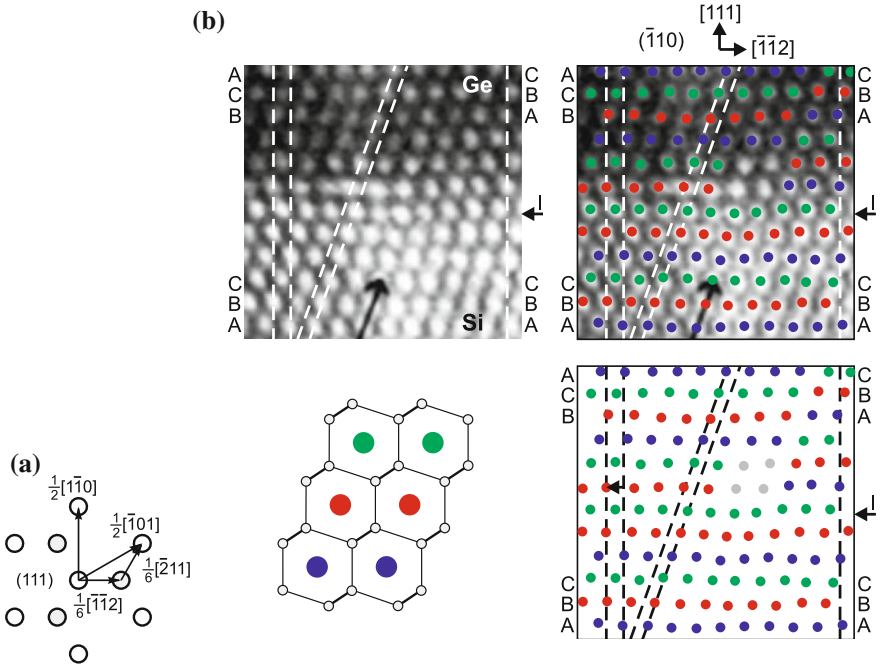

Fig. 4.19 (**a**) Graphical representation of the dislocation reaction of (4.22). (**b**) TEM image of the interface of a Ge/Si heterostructure with a $\left[\bar{2}11\right]/6$ Shockley partial dislocation. The image is overlayed with empty rod positions (as schematically shown in the *lower left* part of the figure) colored according to the stacking position (A: *blue*, B: *red*, C: *green*). The *arrows* labeled 'I' denote the position of the interface. Based on [289]

can be dissociated into two partials. This is energetically favorable. As an example we consider the reaction (Fig. 4.19a)

$$\frac{1}{2}\left[\bar{1}01\right] \rightarrow \frac{1}{6}\left[\bar{1}\bar{1}2\right] + \frac{1}{6}\left[\bar{2}11\right]. \tag{4.22}$$

The length of the full dislocation is $a_0/\sqrt{2}$. The length of the Shockley partial is $a_0/\sqrt{6}$. Thus the energy $E = G\,b^2$ of the full dislocation is $E_1 = Ga_0^2/2$ and the sum of the energies of the partials is smaller, $E_2 = 2Ga_0^2/6 = Ga_0^2/3$. In Fig. 4.19b a TEM image of a Ge/Si interface with a Shockley partial is shown.

4.3.2 Visualization of Dislocations by Etching

Defects can be made visible using etching techniques. This is particularly popular for finding dislocations. Many etches are anisotropic, i.e. the etch velocity varies for different crystal directions. As an example the result of etching a silicon sphere in molten KOH and a germanium sphere in a HNO_3/HF solution are shown in Fig. 4.20. The remaining bodies exhibit those planes with low etching velocity. The etch velocity of various etch solutions has been investigated in detail in particular for silicon (Fig. 4.21).

In a planar geometry, etch pits indicate the presence of dislocations, as shown in Fig. 4.22 for Ge of different orientation. The anisotropic etch prepares {111} planes. The dislocation core is at the intersection of the planes. In Fig. 4.23 hexagonal etch pits stretched along $[1\bar{1}0]$ are developed by molten KOH [292, 293]. The sides of the base are along [110], ⟨130⟩ and ⟨310⟩. The depth and width of the pits increases with increasing etching time. On the $(00\bar{1})$ surface, the orientation of the pits is rotated by 90° because of the polar [111]-axis of the zincblende structure [292]. Such etch

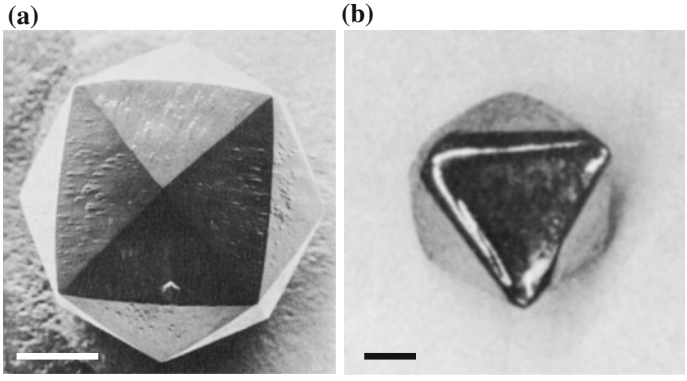

Fig. 4.20 (**a**) Resulting shape of Si sphere ('Lösungskörper') after 3 h at 100 °C in molten KOH. (**b**) Resulting shape of Ge sphere after etching in HNO_3:HF:CH_3COOH, 35:30:35 weight percent. The octaedric form indicates {111} faces. Markers are 1 mm. Adapted from [290]

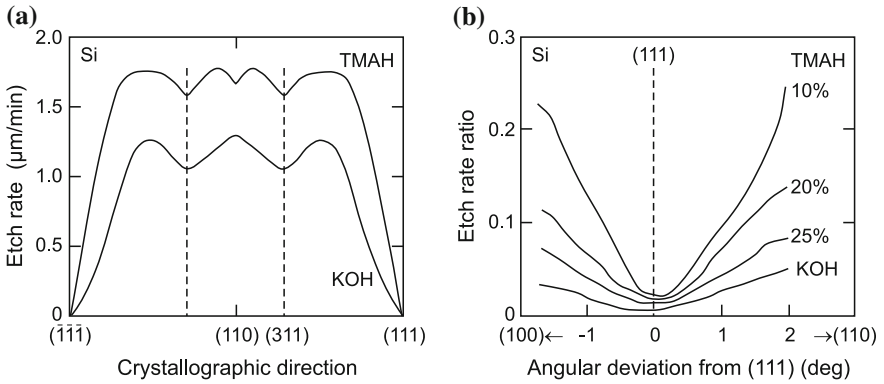

Fig. 4.21 (**a**) Etch rate of silicon for tetramethyl-ammonium-hydroxide (TMAH) water solution (25 %) at 86 °C and 40 % KOH at 70 °C as a function of crystallographic direction. (**b**) Detail of the anisotropy around the (111) direction for TMAH solutions with three different concentrations and 40 % KOH, all at 86 °C. Adapted from [291]

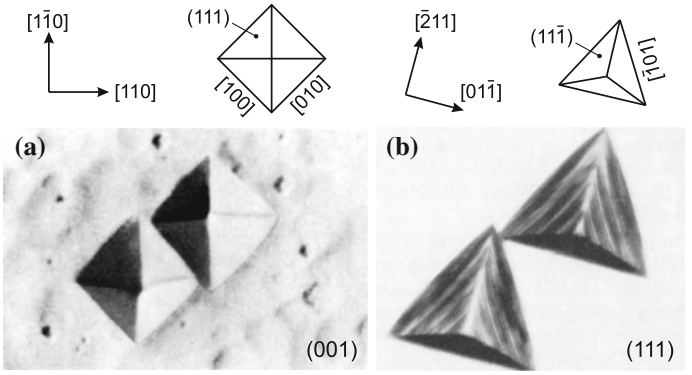

Fig. 4.22 Etch pits on germanium with (**a**) (001) and (**b**) (111) surface orientation. In both cases {111} facets are prepared by the etch. As etch in (**b**) a HNO_3/HF/CH_3COOH solution with $AgNO_3$ additive has been used. Width of the triangular etch pits is about 100 µm. Adapted from [296]

pit develops at a dislocation with Burger's vector $a/2\,[011]$ (inclined to the (001) surface) [294]. Other types of etch pits indicate dislocations with other Burger's vectors [294, 295]. Recipies how to wet chemically etch various semiconductors can be found in [290, 296–299]. Other etching techniques include dry processes such as plasma etching or reactive ion etching (RIE) [300–303].

4.3.3 Impurity Hardening

It has been found that the addition of impurities can lead to a substantial reduction of the dislocation density. This effect is known as impurity hardening and is caused by a hardening of the lattice due to an increase of the so-called critical resolved

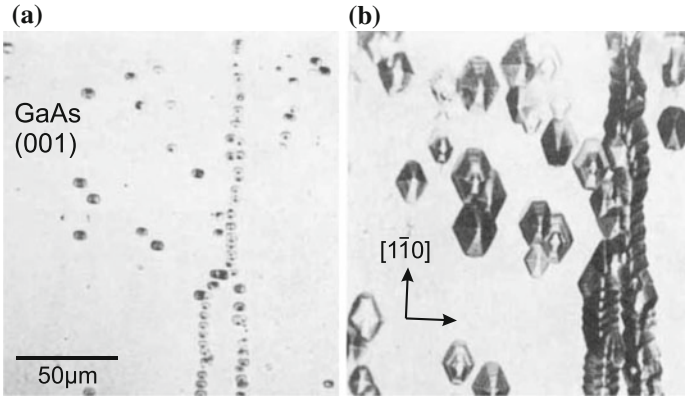

Fig. 4.23 Etch pits on GaAs (001) after (**a**) 3 min and (**b**) 10 min etch time in molten KOH at 300 °C. Adapted from [292]

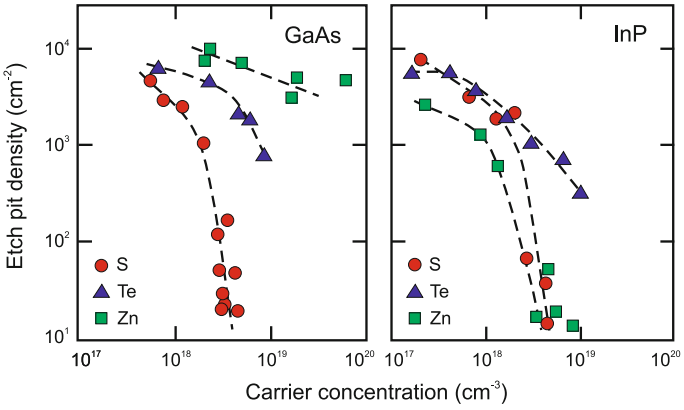

Fig. 4.24 Dislocation density (as revealed by etch pits) for GaAs and InP as a function of the carrier concentration for various concentrations of impurities (S, Te, and Zn). Adapted from [305]

shear stress [304]. In Fig. 4.24 the dependence of the dislocation density in GaAs and InP is shown as a function of the carrier density that is induced by the incorporation of (electrically active) group-II or group-VI atoms (acceptors or donors, cf. Sect. 7.5). The high carrier concentration is unwanted when semi-insulating substrates (cf. Sect. 7.7.8) or low optical absorption (cf. Sect. 9.8.1) are needed. Thus the incorporation of isovalent impurities, such as In, Ga or Sb in GaAs and Sb, Ga or As in InP, has been investigated and found to be remarkably effective. Material containing such impurities in high concentration ($>10^{19}\,\mathrm{cm}^{-3}$) must be considered a low-concentration alloy. The lattice constant is thus slightly changed, which can cause problems in the subsequent (lattice-mismatched) epitaxy of pure layers.

4.4 Extended Defects

4.4.1 Micro-cracks

If the stress in a material becomes too big to be accommodated by dislocations, cracks may form to release strain energy.[8] In Fig. 4.25 an example is shown. In this case, micro-cracks have formed in a bulk mercury indium telluride crystal upon incorporation of residual stress and thermal stress during cooling of the material from growth temperature (about 1000 K) to room temperature. See also Fig. 12.19 for micro-cracks in an epitaxial layer.

4.4.2 Stacking Faults

The ideal stacking of (111) planes in the zincblende structure, ABCABC. . ., can be disturbed in various ways and creates area defects. If one plane is missing, i.e. the stacking is ABC*AC*ABC, an *intrinsic* stacking fault is present. If an additional plane is present, the defect is called an *extrinsic* stacking fault, i.e. ABCAB*A*CABC. An extended stacking fault in which the order of stacking is reversed is called a twin lamella, e.g. ABCABC*BAC B*ABCABC. If two regions have inverted stacking order they are called twins and their joint interface is called a twin boundary, e.g. . . .ABCABCAB*C*BACBACBA. . . (Fig. 4.28). The various types of stacking faults are shown in Fig. 4.26. In Fig. 4.27 a cross-sectional image of stacking faults in GaAs on Si is shown. They block each other and thus partially annihilate with increasing thickness.

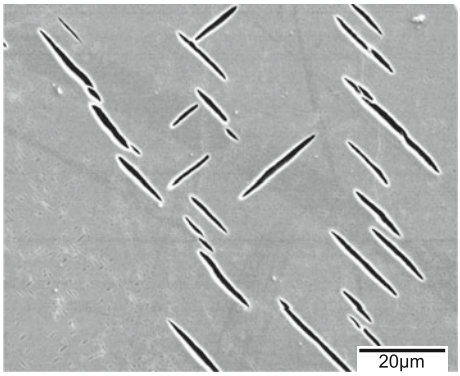

Fig. 4.25 Micro-cracks in a mercury indium telluride crystal. Adapted from [306]

[8] We note that in elasticity theory a *continuous* deformation is assumed. Obviously the separation (fracture) into two unstrained blocks is the lowest strain energy state of a stressed piece of material.

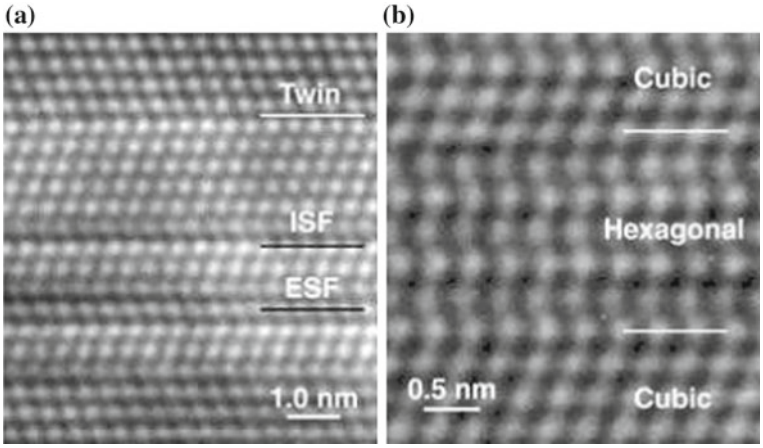

Fig. 4.26 HRTEM images of (**a**) thin-film silicon with intrinsic (labeled 'ISF') and extrinsic ('ESF') stacking faults and twin boundary ('Twin'). (**b**) Six monolayer thick hexagonal (wurtzite) CdTe layer in cubic (zincblende) CdTe. Stacking order (from bottom to top) is: ABCABABABABC... Reprinted with permission from [307]

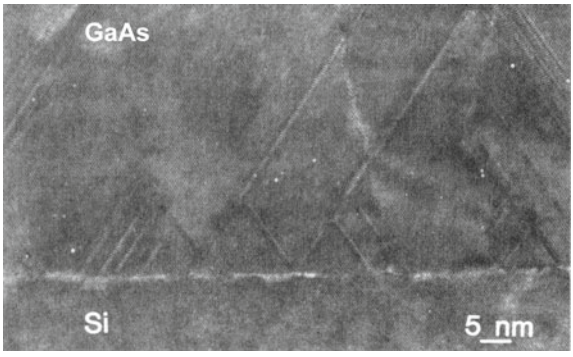

Fig. 4.27 Cross-sectional TEM image showing stacking faults in heteroepitaxial GaAs on Si. Adapted from [308]

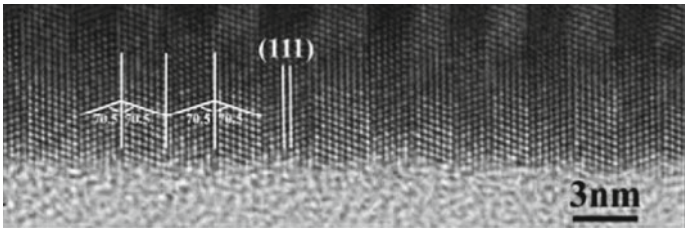

Fig. 4.28 High resolution TEM image of ZnS nanowire exhibiting periodical twin structures. Adapted from [309]

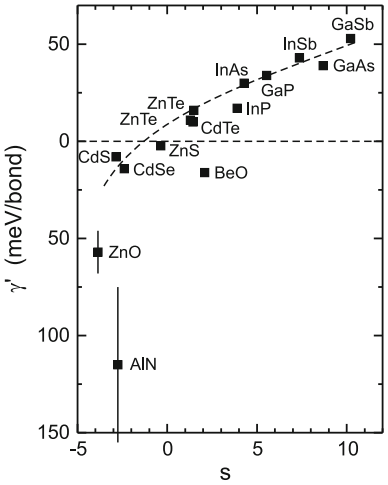

Fig. 4.29 Reduced stacking fault energy (stacking fault energy per bond) γ' for various compound semiconductors plotted as a function of the s-parameter. *Dashed line* is guide to the eye. Data from [164]

A stacking fault is bounded by two partial dislocations (Sect. 4.3.1) formed by the dissociation of a perfect dislocation. A full (or *perfect*) dislocation with Burger's vector $a/2[110]$ in a III–V compound is dissociated into two Shockley partials according to (4.22) [310]. Since the dislocation energy is proportional to $|\mathbf{b}|^2$, the dissociation is energetically favored (see Sect. 4.3.1).

The stacking-fault energy in pure silicon is $\gamma = 47\,\mathrm{mJ\,m^{-2}}$ [311]. A similar value is found for Ge, $\gamma = 60\,\mathrm{mJ\,m^{-2}}$ [312] and undoped GaAs, $\gamma = 45\,\mathrm{mJ\,m^{-2}}$ [313]. In diamond a much larger value is found, $\gamma = 285\,\mathrm{mJ\,m^{-2}}$ [314]. Impurity incorporation typically reduces the stacking fault energy. The systematics of stacking fault energy for various III–V and II–VI compounds has been discussed [164, 315, 316]. It can be correlated with the s-parameter (2.11) as depicted in Fig. 4.29.

4.4.3 Grain Boundaries

The boundaries of crystal grains are called *grain boundaries*. They are defined by five parameters, three rotation angles (e.g. Euler angles) to describe how the orientation of grain II results from grain I and two parameters to define the boundary plane of the two grains in the coordinate system of reference grain I.

Such defects can have a large impact on the electric properties. They can collect point defects and impurities, act as barriers for transport (Sect. 8.3.8) or as carrier sinks due to (nonradiative) recombination. Details of their structure and properties can be found in [317, 318]. The two crystal grains meet each other with a relative tilt and/or twist. The situation is shown schematically in Fig. 4.30a for a small angle between the two crystals. A periodic pattern of dislocations forms at the interface that is called a small-angle grain boundary (SAGB) (Fig. 4.30b). In

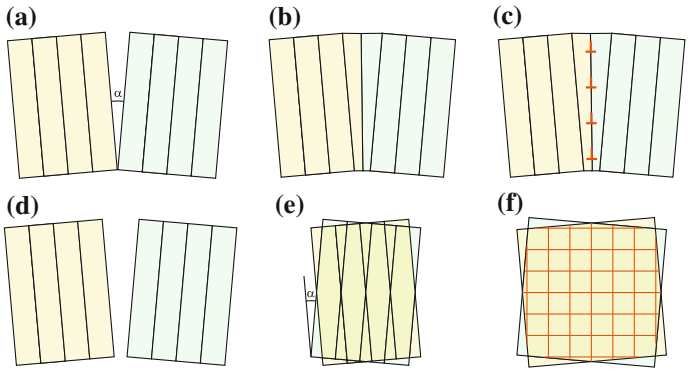

Fig. 4.30 Schemes of (**a**, **b**, **c**) pure tilt and (**d**, **e**, **f**) pure twist boundary, dislocation formation in (**c**) pure tilt and (**f**) twist boundaries

Fig. 4.31, experimental results for pure tilt SAGB are shown. The dislocation spacing is inversely proportional to the tilt angle θ. An image of a twist SAGB is shown in Fig. 4.32.

Special large angle boundaries possess (for a certain angle) a *coincident site lattice* (CSL). Some of these grain boundaries have a low energy and are thus commonly observed. The ratio of lattice points of the CSL and the lattice unit cell is an odd integer number n; the corresponding grain boundary is then labeled Σn. SAGB are also termed $\Sigma 1$. $\Sigma 3$ grain boundaries are always twin boundaries. An example with (111) grain boundary is schematically shown in an example in Fig. 4.33a. A $\Sigma 3$ (twin) boundary in silicon with {112} grain boundary [319, 320] is depicted in Fig. 4.34 together with the atomic arrangement of the grain boundary itself. A $\Sigma 5$ (001) grain boundary is schematically shown in Fig. 4.33b; the special angle is $\theta = \arctan 3/4 \approx 36.87°$.

Real grain boundaries may not be flat, contain impurities or precipitates and even consist of a thin amorphous layer.

4.4.4 Antiphase and Inversion Domains

Antiphase domains occur when one part of the crystal is shifted with respect to another by an antiphase vector **p**. This does not form a twin. If the polar direction changes between two domains they are called inversion domains.

In the zincblende structure the [110] and [$\bar{1}$10] directions are not equivalent. In one case there is a Zn-S lattice and in the other a S-Zn lattice. Both lattices vary by a 90° rotation or an inversion operation (which is not a symmetry operation of the zincblende crystal). If, e.g., a zincblende crystal is grown on a Si surface with monoatomic steps (Fig. 4.35, cmp. Fig. 11.6), adjoint regions have a different phase;

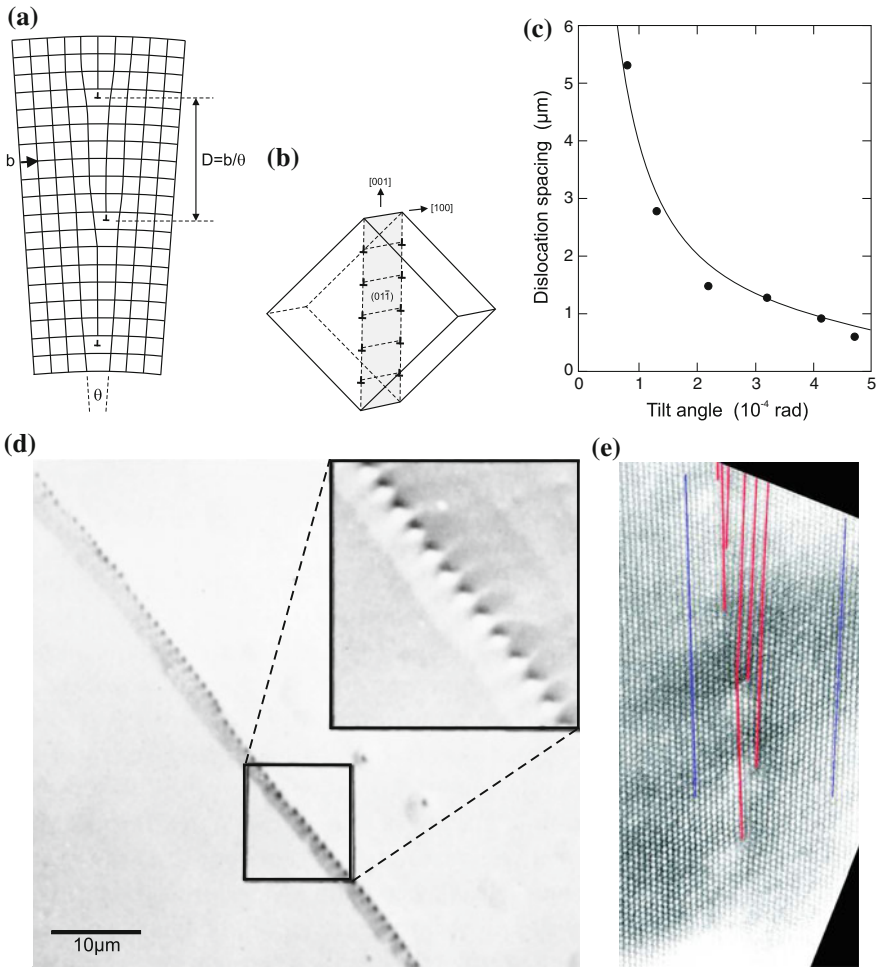

Fig. 4.31 (**a**) Scheme of a small-angle (pure tilt) grain boundary. (**b**) Model of edge dislocations in a {110} plane in Ge. (**c**) Relation of dislocation distance d and tilt angle θ for various small-angle grain boundaries in Ge. *Solid line* is relation $d = 4.0 \times 10^{-8}/\theta$. (**d**) Optical image of an etched (CP–4 etch) Ge sample with a small-angle grain boundary. Adapted from [321]. (**e**) HRTEM image of a small-angle grain boundary in Si with dislocations highlighted. From [322]

they are called antiphase domains (APD). The antiphase vector is $(0, 0, 1)\, a_0/4$. At the boundaries a two-dimensional defect, an antiphase domain boundary, develops. The APD boundary contains bonds between identical atom species. In Fig. 4.36, intertwining APD boundaries are shown on the surface of InP layers on Si. The antiphase domains can be visualized with an anisotropic etch.

In Fig. 4.37a, inversion domains in iron-doped ZnO are shown. Between domains the direction of the c-axis is reversed. The iron is found preferentially in the inversion domain boundary (IDB) (Fig. 4.37b) and plays an important role in its formation [326, 327].

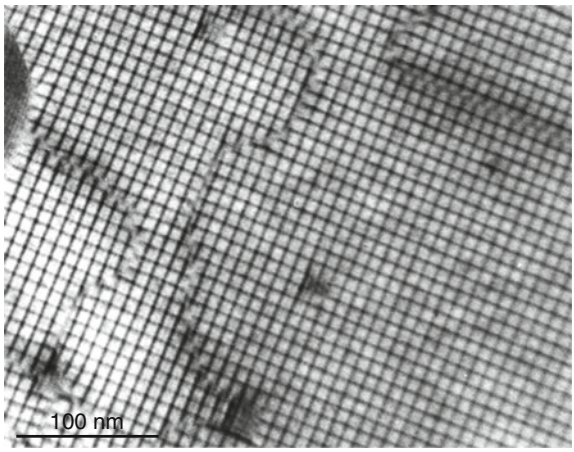

Fig. 4.32 Bright-field TEM image of pure twist boundary with network of pure twist dislocations fabricated by wafer bonding of two Si (001) surfaces with a relative twist. Adapted from [323]

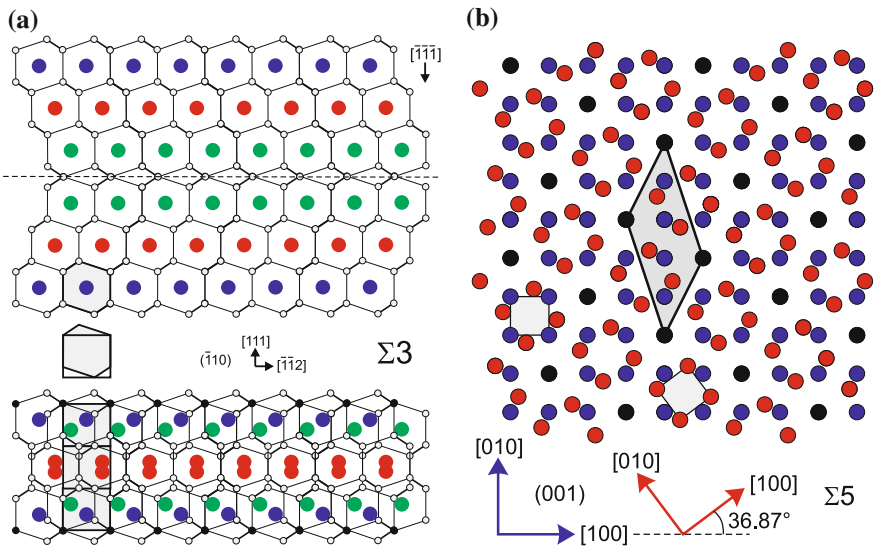

Fig. 4.33 (**a**) Schematic of $\Sigma 3$ (111) twin boundary in a diamond or zincblende structure (cmp. Fig. 4.28). The grain boundary is marked by a *dashed line* shown in side-view. The hexagonal and rectangular *grey boxes* have the same area. The lattice points of the coincident site lattice (CSL) are shown with *black circles* in the lower part of the figure. The unit cell of the CSL has three times the volume of the unit cell of the fcc lattice. (**b**) Schematic of a $\Sigma 5$ (001) grain boundary in a (simple) cubic crystal shown in plane-view. The *blue* and the *red* lattice are rotated by 36.86°, the lattice points of the CSL are shown in *black*. The unit cell of the CSL lattice (*dark grey*) has five times the volume of the cubic unit cell (*light grey*)

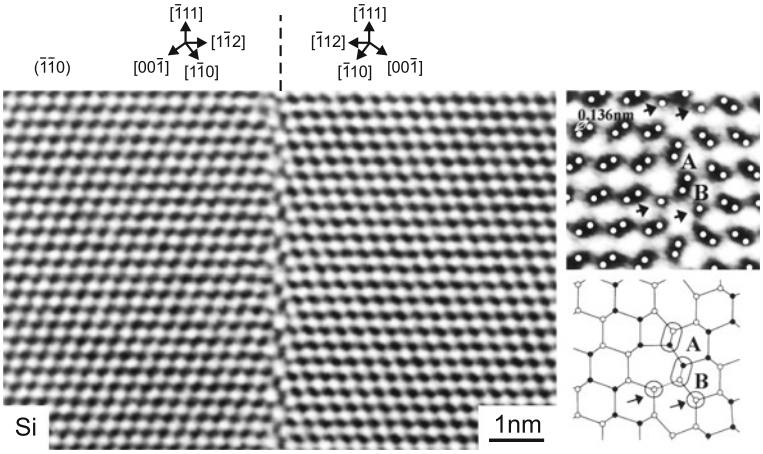

Fig. 4.34 TEM images in two magnifications of a $\Sigma 3\{112\}$ boundary in silicon together with a schematic of the atomic arrangement. Adapted from [320]

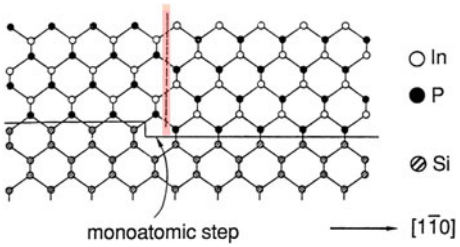

Fig. 4.35 Monoatomic step on the Si (001) surface and subsequent formation of an antiphase boundary in InP (zincblende)

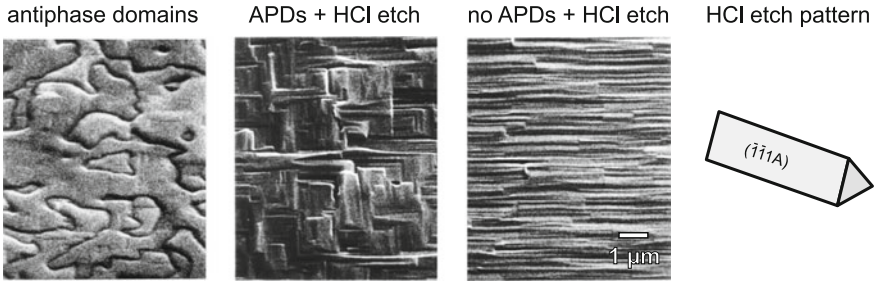

Fig. 4.36 Antiphase domains in InP on Si. HCl etchs InP anisotropically and prepares (111)A planes. The etch patterns of layers with (without) APDs are cross-hatched (linear). Adapted from [324]

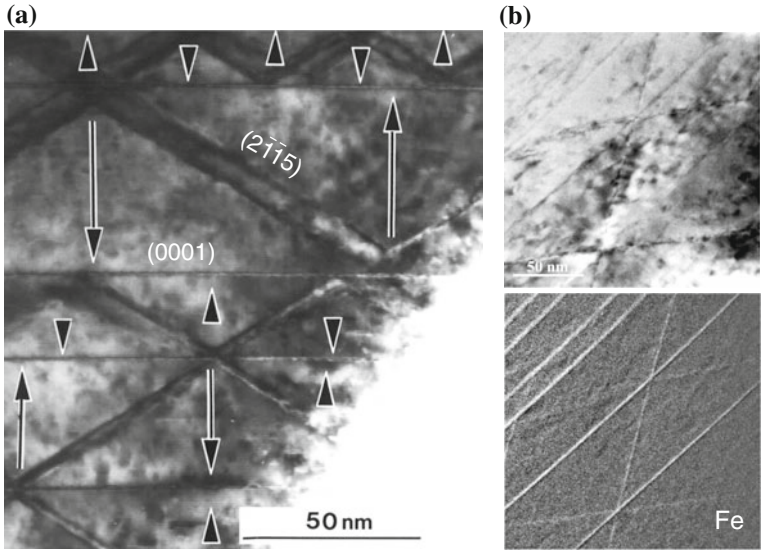

Fig. 4.37 Transmission electron microscopy of inversion domains in ZnO:Fe. (**a**) Inversion domains in iron-doped ZnO ($ZnO:Fe_2O_3 = 100:1$). *Arrows* denote the orientation of the *c*-axis in the respective domains. (**b**) *Top*: bright field TEM, *bottom* Fe distribution from energy-filtered image. Adapted from [325]

4.5 Disorder

Disorder is a general term for deviations from the ideal structure on a microscopic scale. Apart from the various structural defects discussed in the previous chapters, further examples of disorder are

- The presence of various isotopes of an element. This introduces disorder with regard to the mass of the atoms and impacts mostly phonon properties (see Fig. 8.24).
- The occupation of lattice sites in alloys (Sect. 3.7) ranging from a random alloy, clustering to (partially) ordered phases.
- The thermal and zero-point motion of atoms around their equilibrium position.