two-dimensional numerical simulations have shed light on various aspects of the applied magnetic field in LPEE. However, these models, being two dimensional, have naturally neglected the contributions of the circumferential velocity and magnetic body force components. Therefore, the results presented earlier were qualitative and did not include information about three-dimensional effects such as deviations from axisymmetry, variations of the magnetic field components, and mixing in the solution.

In order to shed light on these three-dimensional effects, three-dimensional models were developed in [29.35, 36] for the LPEE growth of a binary system. To the best of our knowledge, these were the first three-dimensional simulation models developed for LPEE. Here, we present the key features of these models.

29.5.1 Simulation Model

The three-dimensional simulation studies carried out in [29.35, 36] were focused on the growth of a binary system (GaAs) for computational simplicity. In addition, the inclusion of the third component (In) in the analysis would not have a significant effect on the flow structures of the solution. The effect of magnetic field nonuniformity was also investigated. Below we present the model and the simulation results of [29.35, 36].

The growth cell selected for simulation is shown in Fig. 29.1. The governing equations are written explicitly in cylindrical coordinates for a binary system as follows

Continuity

$$\frac{1}{r}\frac{\partial}{\partial r}(ru)+\frac{1}{r}\frac{\partial v}{\partial \varphi}+\frac{\partial w}{\partial z}=0\,. \tag{29.20}$$

Momentum

$$\begin{aligned}&\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial r}+\frac{v}{r}\frac{\partial u}{\partial \varphi}+w\frac{\partial u}{\partial z}-\frac{v^2}{r}\\&=\nu\left(\nabla^2 u-\frac{u}{r^2}-\frac{2}{r^2}\frac{\partial v}{\partial \varphi}\right)-\frac{1}{\rho_{\mathrm{L}}}\frac{\partial p}{\partial r}-\frac{\sigma_{\mathrm{E}}B^2u}{\rho_{\mathrm{L}}}\,,\end{aligned} \tag{29.21}$$

$$\begin{aligned}&\frac{\partial v}{\partial t}+u\frac{\partial v}{\partial r}+\frac{v}{r}\frac{\partial v}{\partial \varphi}+w\frac{\partial v}{\partial z}+\frac{uv}{r}\\&=\nu\left(\nabla^2 v-\frac{v}{r^2}+\frac{2}{r^2}\frac{\partial u}{\partial \varphi}\right)-\frac{1}{r\rho_{\mathrm{L}}}\frac{\partial p}{\partial \varphi}-\frac{\sigma_{\mathrm{E}}B^2v}{\rho_{\mathrm{L}}}\,,\end{aligned} \tag{29.22}$$

$$\begin{aligned}&\frac{\partial w}{\partial t}+u\frac{\partial w}{\partial r}+\frac{v}{r}\frac{\partial w}{\partial \varphi}+w\frac{\partial w}{\partial z}\\&=\nu\nabla^2 w-\frac{1}{\rho_{\mathrm{L}}}\frac{\partial p}{\partial z}-g\beta_{\mathrm{t}}(T-T_0)+g\beta_{\mathrm{c}}(C-C_0)\,.\end{aligned} \tag{29.23}$$

Mass transport

$$\begin{aligned}&\frac{\partial C}{\partial t}+(u+\mu_{\mathrm{E}}E_r)\frac{\partial C}{\partial r}+\frac{v}{r}\frac{\partial C}{\partial \varphi}\\&+(w+\mu_{\mathrm{E}}E_z)\frac{\partial C}{\partial z}=D_{\mathrm{C}}\nabla^2 C\,.\end{aligned} \tag{29.24}$$

Energy

$$\frac{\partial T}{\partial t}+u\frac{\partial T}{\partial r}+\frac{v}{r}\frac{\partial T}{\partial \varphi}+w\frac{\partial T}{\partial z}=\alpha\nabla^2 T\,. \tag{29.25}$$

Electric charge balance

$$\frac{\partial^2\phi}{\partial r^2}+\frac{1}{r}\frac{\partial \phi}{\partial r}+\frac{\partial^2\phi}{\partial z}=0\,, \tag{29.26}$$

where the gradient operator is defined as

$$\nabla^2=\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right)+\frac{1}{r^2}\frac{\partial^2}{\partial \varphi^2}+\frac{\partial^2}{\partial z^2}\,. \tag{29.27}$$

The associated boundary and interface conditions are given as follows. Along the vertical wall

$$\begin{aligned}&u=0\,,\quad v=0\,,\quad w=0\,,\quad T=T_{\mathrm{g}}-\frac{z-z_0}{H}\Delta T\,,\\&\frac{\partial\phi}{\partial r}=0\,,\quad \frac{\partial C}{\partial r}=0\,.\end{aligned} \tag{29.28}$$

Along the growth interface

$$\begin{aligned}&u=0\,,\quad v=0\,,\quad w=0\,,\\&k_{\mathrm{S}}\frac{\partial T}{\partial z}-k_{\mathrm{L}}\frac{\partial T}{\partial z}=-\pi J\,,\\&-\sigma_{\mathrm{E}}\frac{\partial\phi}{\partial r}=J\,,\quad C=C_1\,.\end{aligned} \tag{29.29}$$

Along the dissolution interface

$$\begin{aligned}&u=0\,,\quad v=0\,,\quad w=0\,,\\&k_{\mathrm{S}}\frac{\partial T}{\partial z}-k_{\mathrm{L}}\frac{\partial T}{\partial z}=+\pi J\,,\quad \phi=0\,,\quad C=C_2\,.\end{aligned} \tag{29.30}$$

Initial conditions at $t=0$

$$C=C_0\,,\quad u=0\,,\quad v=0\,,\quad w=0\,,\quad T=T_{\mathrm{g}}\,, \tag{29.31}$$

where u, v, and w are the velocity components in the r-, φ-, and z-directions, respectively, T is temperature, C is solute concentration, T_0 and C_0 are the reference values, H is solution height, and C_1 and C_2 are the concentrations on the growth and dissolution interfaces, respectively. Definitions of other symbols used here are given in Tables 29.3 and 29.4.

Table 29.3 Physical properties of the GaAs system [29.34]

Parameter	Value
Growth temperature T_g	800 °C
Current density J	10 A/cm^2
Peltier coefficient π	0.3 V
Solution electric conductivity σ_E	25 000 Ω^{-1} cm^{-1}
Solid (crystal and source) electric conductivity σ_S	40 Ω^{-1} cm^{-1}
Thermal diffusivity α	0.30 cm^2/s
Solutal diffusion coefficient D_C	4.0×10^{-5} cm^2/s
Solution kinematic viscosity ν	1.21×10^{-3} cm^2/s
Solution density ρ^L	5.63 g/cm^3
Solute electric mobility μ_E	0.027 cm^2/V s
Thermal expansion coefficient β_t	9.85×10^{-5} K^{-1}
Solutal expansion coefficient β_c	-8.4×10^{-2}
Thermal conductivities k	
Graphite	0.225 W/cm K
Ga-As solution and contact zone	0.526 W/cm K
GaAs substrate	0.082 W/cm K
Boron nitride: the r-direction	0.282 W/cm K
Boron nitride: the z-direction	0.440 W/cm K
Crystal radius	6.0 mm
Furnace radius	36.0 mm
Graphite height	19.0 mm
Solution height	6.0 mm
Substrate thickness	0.3 mm
Source thickness	5.0 mm
Contact zone height	2.0 mm

Table 29.4 Parameters of the GaAs system [29.35]

Parameter	Symbol	Value
Crystal radius	R_c	12.0 mm
Solution height	H	10.3 mm
Magnetic field intensity	B	0–12 kG
Nonuniformity coefficient	A	0–4
Thermal Grashof number	Gr_T	6.84×10^4
Solutal Grashof number	Gr_C	7.10×10^3
Hartmann number	Ha	0–871.5
Prandtl number	Pr	4.00×10^{-3}

29.5.2 Numerical Method

The commercial CFX software was used to solve the field equations. The computation mesh in the liquid is $120\times40\times80$ in the r-, φ-, and z-directions, respectively, which was demonstrated to be sufficient for an accurate and stable solution. Since the focus is on the flow field only, the evolution of the growth and dissolution interfaces is not included in the computations. The mass transport equation in the solution is solved simultaneously in order to take into account the influence of concentration field on the flow field (through the solutal Grashof number); however, the concentration field is not presented here for the sake of brevity. Transient terms are considered in the energy, mass transport, and momentum equations in order to account for possible unsteady flows and their influence. The simulation results are presented at $t = 1$ h since the flow field has fully developed by that time.

The required physical and growth parameters of the GaAs system are given in Tables 29.3 and 29.4. The Hartmann number is defined by

$$\mathrm{Ha} = BH\sqrt{\sigma_E/\rho_L\nu}\,. \tag{29.32}$$

The simulations are carried out for half of the cylindrical cell domain for computational efficiency. However, to ensure that the half-domain solution represents fully the flow structure of the full domain, a full-domain solution was carried out for a case for which asymmetry was assured (under a large magnetic field, $B = 4$ kG). Results demonstrated that the half-plane treatment is reliable.

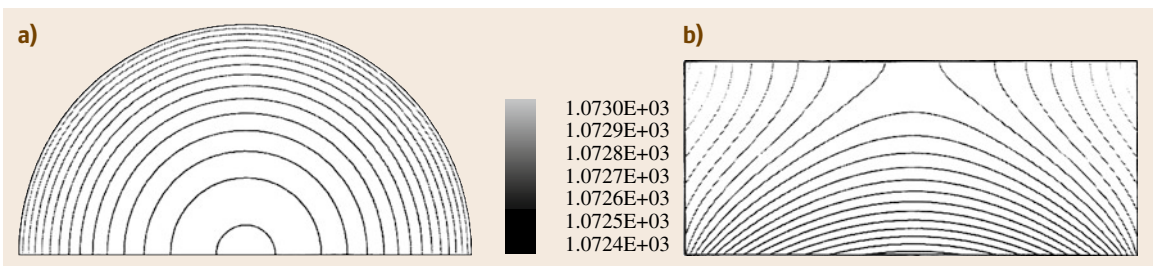


Fig. 29.16a,b Temperature distribution (K): **(a)** near the growth interface in the horizontal plane, **(b)** in the vertical plane at $\varphi = 0$ (after [29.35, 36])

Computed temperature distributions in the horizontal plane near the growth interface and also in the vertical plane at $\varphi = 0$ are given in Fig. 29.16. Temperature distributions agree with earlier 2-D solutions [29.34, 49, 50]. In addition, since the electric current is passing through the source in this setup, the computed isotherm patterns indicate that the shape of the growth interface will be single-humped (concave towards the crystal), as expected.

29.5.3 Effect of Magnetic Field Strength

For a better visualization, the simulation results for the flow field in the growth cell are presented in three distinct planes: the vertical (r–z) plane at $\varphi = 0$ that represents typical flow structures along the growth direction, the horizontal (r–φ) plane at the middle of the growth cell ($z = 5.15$ mm) where the changes in the flow field will be more prominent for radial and circum-

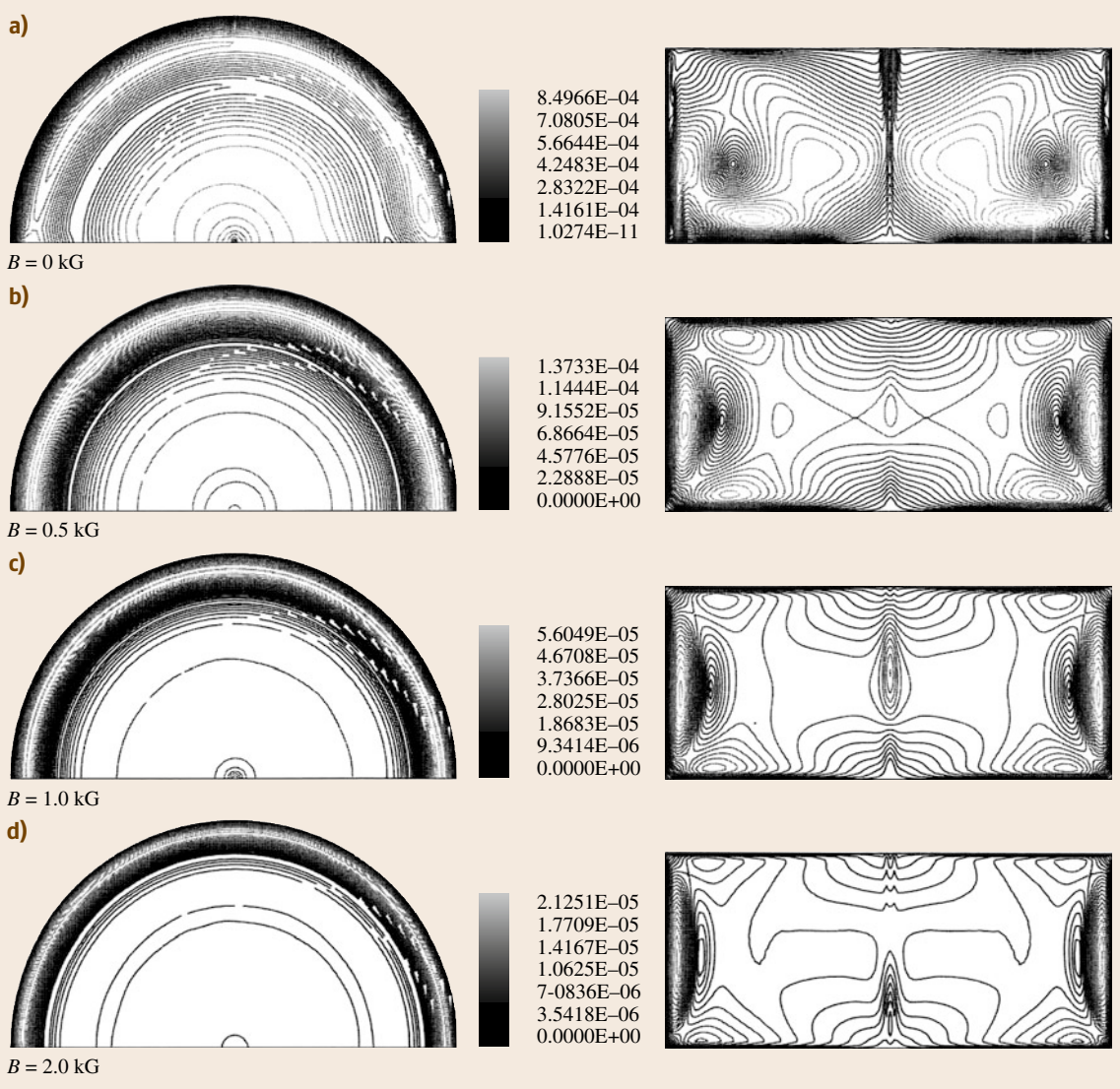


Fig. 29.17a–d Flow field in the horizontal plane in the middle of the solution zone (*left column*) and in the vertical plane at $\varphi = 0$ (*right column*) (after [29.35, 36])

ferential velocity components representing mixing, and finally the horizontal plane (r–φ) near the growth interface ($z = 0.4$ mm) where the flow field is often closely related to the quality of the grown crystals.

Figure 29.17 presents the simulation results for the flow field presented in the horizontal plane at the middle of the growth cell (left column) and in the vertical plane at $\varphi = 0$ (right column) for four levels of magnetic field strengths ($B = 0.0$, 0.5, 1.0, and 2.0 kG), while Fig. 29.18 shows the results for the flow field in the horizontal plane near the growth interface (left column) and in the vertical plane at $\varphi = 0$ (right column) for intermediate magnetic field intensities ($B = 2.5$ and 3.0 kG), and in the horizontal plane at the middle of the growth cell (left column) and in the vertical plane at $\varphi = 0$ (right column) for higher field levels ($B = 4.0$

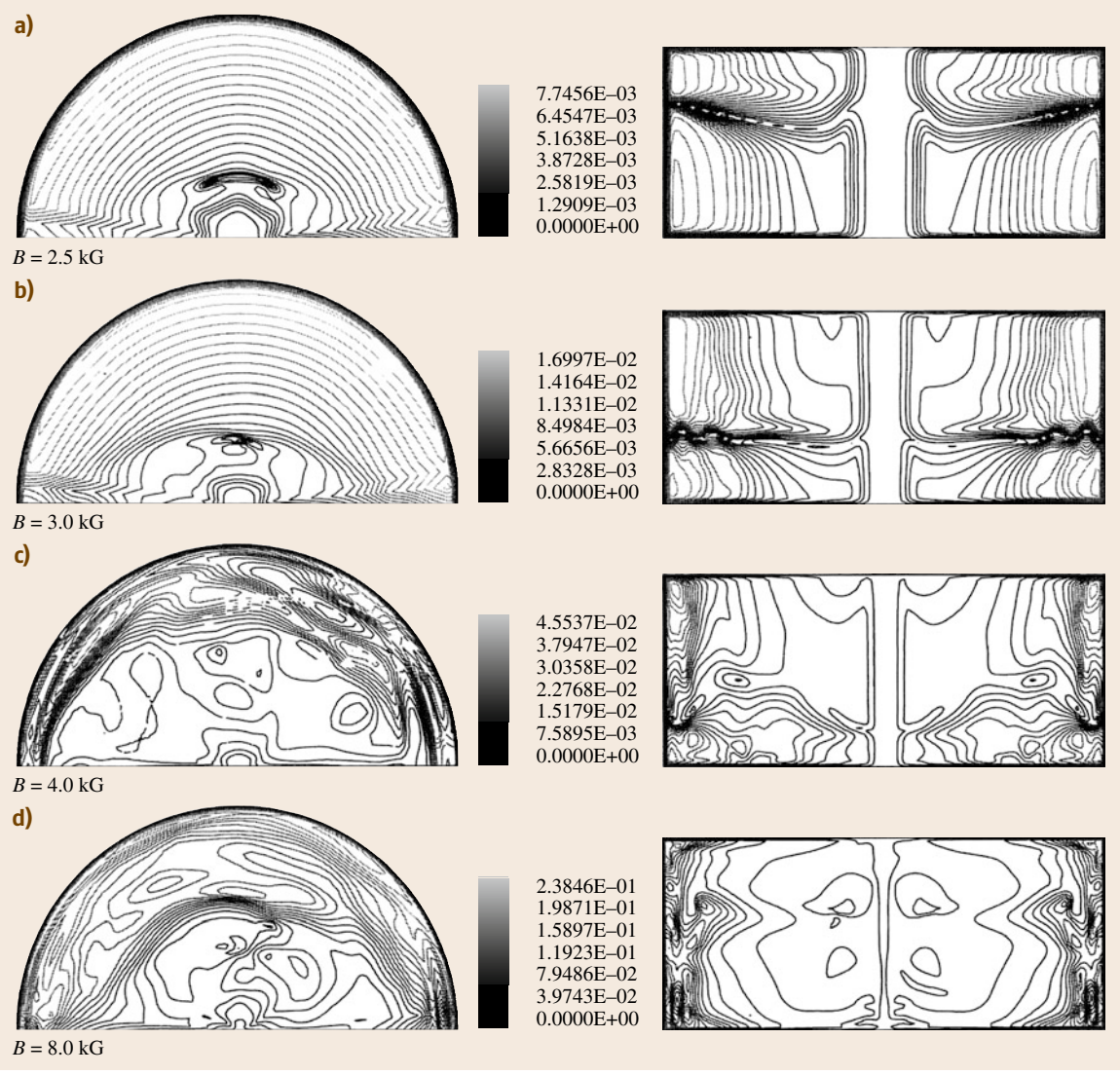


Fig. 29.18a–d Flow field in the horizontal plane near the growth interface at $z = 0.4$ mm (*left column*) and in the vertical plane at $\varphi = 0$ (*right column*) for **(a)** $B = 2.5$ kG, and **(b)** $B = 3.0$ kG, and in the horizontal plane at the middle of the growth cell (*left column*) and in the vertical plane (*right column*) for **(c)** $B = 4.0$ kG, and **(d)** $B = 8.0$ kG (after [29.35,36])

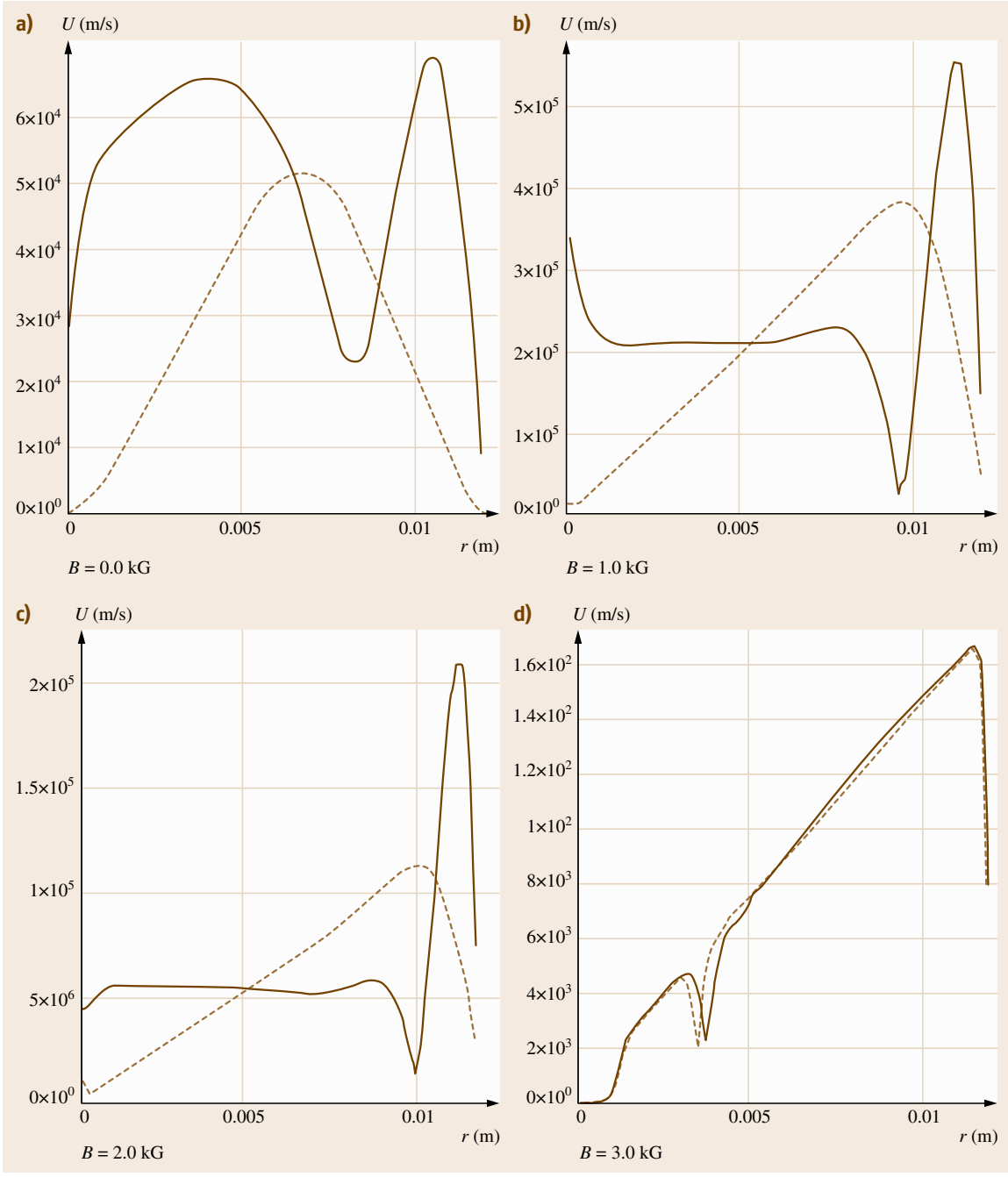


Fig. 29.19a–d Variation of the flow strength $U = (u^2 + v^2 + w^2)^{1/2}$ along the radial direction at $\varphi = \pi/2$ under various magnetic field strengths: **(a)** $B = 0.0$ kG, **(b)** $B = 1.0$ kG, **(c)** $B = 2.0$ kG, and **(d)** $B = 3.0$ kG. *Solid lines* at $z = 5.05$ mm (in the middle of the growth cell), and *dashed lines* at $z = 0.4$ mm (near the growth interface) (after [29.35, 36])

and 8.0 kG). Local magnitudes of the flow velocity $U = \sqrt{u^2 + v^2 + w^2}$ (which will be referred to as the *flow strength*) are computed in m/s, and the scales of flow strengths are shown in each figure.

The computed flow patterns in the vertical plane appear more complex than those of the 2-D simulations presented for the same LPEE growth system. This complexity can be attributed mainly to the inclusion of the contribution of the circumferential velocity component. In addition, the size of the growth crucible used in this study is twice that of the one used in [29.34,49,50]; this might have also contributed towards these differences. Such differences in the flow field of 2-D and 3-D models have also been observed in [29.39,52,53]. The strongest flow is seen in the lower part of the crucible cell, near the center of the half-domain along the r-direction. This is similar to what was observed in [29.34]. A weak flow cell forms just above the strongest flow cell. The flow in the horizontal planes is nearly homocentric. The flow in the horizontal plane near the growth interface is stronger in the middle region along the radial direction and becomes weaker and weaker near the growth cell wall or the axial center (Fig. 29.17a, left column). On the other hand, in the horizontal plane at the middle of the growth cell (Fig. 29.17a, right column) there are two maximum points for the flow intensity along the radial direction, with a relatively strong flow in the central region.

Note that the flow field in Fig. 29.17a is not strictly axisymmetric, especially in the middle of the growth cell, although the flow is stable. This result suggests that the present system is near the Grashof number that is a little bit lower than the critical Grashof number (2.5×10^5) under the same conditions (Prandtl number and geometry aspect ratio) given in [29.60] in the analysis of axisymmetry breaking of natural convection in a vertical Bridgman growth configuration.

Results are also summarized in Figs. 29.19 and 29.20 for various aspects of the flow field. Simulation results show three distinct characteristics depending on the level of applied magnetic field: (a) the *weak* magnetic field, with intensities from 0.0 to 2.0 kG, (b) the *intermediate* magnetic field, with levels from 2.0 to 3.0 kG, and (c) the *high* magnetic field, with intensities above 3.0 kG. Flow characteristics are quite different at each of these field levels.

Let us first focus on Fig. 29.17b–d, which represents results for magnetic field levels from 0.5 to 2.0 kG. In this category, the weak magnetic field level, an increase in the applied magnetic field strength results in significant reduction in the flow strength, which is, in general, desirable for a stable and controlled crystal growth. Flow cells are the strongest near the vertical wall and forms the so-called Hartmann layer [29.51–54]. At higher magnetic field strengths, the relative strength of these flow cells becomes increasingly stronger compared with the flow in the rest of the growth cell domain, and the Hartmann layer becomes thinner, which is in accordance with the scaling analysis given in [29.51, 52]. Furthermore, the strongest flow cells appearing in the lower part of the growth cell (Fig. 29.17a) move further down towards the growth interface with increasing strength, and new flow cells form near the dissolution interface. These strong flow cells form very visible boundary layers near the growth and dissolution interfaces (the so-called *end layers* [29.51, 52, 54]), and hence give rise to strong vertical velocity gradients in the vicinity of the growth interface, which may have an adverse effect on the crystal growth process.

One may state that the application of a magnetic field may not always be beneficial for the growth process [29.40, 54]. It was shown in [29.54] that the radial nonuniformity in vertical Bridgman is the most

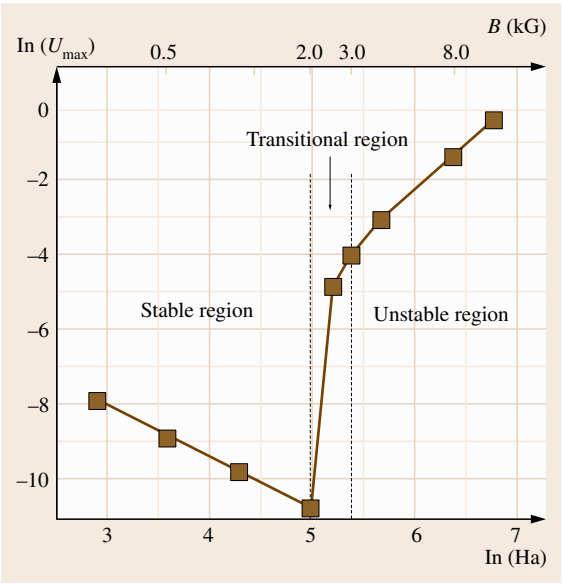


Fig. 29.20 Variation of the maximum velocity with magnetic field intensity: three distinct regions of stability of the flow field are obvious. The flow field is stable up to Ha = 150; in the region between Ha = 150 and 220 the flow is transitional; and above Ha = 220 the flow is unstable (after [29.35, 36])

significant at intermediate levels of magnetic field strength.

Flow fields become perfectly homocentric (and hence axisymmetric) with the increase of magnetic field strength, and hence the flow strength decreases. The flow velocities in the central region of the cell (in both the r- and z-directions) become more uniform at higher magnetic field levels, and form a core region of the uniform flow field, leading to a domain in which the flow is suppressed with the application of an applied magnetic field [29.51, 52, 54]. Indeed, an applied magnetic field at weak intensity levels can suppress fluid flows in a growth system, as shown in Fig. 29.17.

To the best of our knowledge, a magnetic field level higher than 8.0 kG is very strong for LPEE growth of thick crystals, due to the strong interaction between the applied magnetic and electric fields. Thus far, it was not possible to exceed the level of 4.5 kG in experiments, as higher magnetic field levels led to unsatisfactory growth [29.26]. The issue of LPEE growth under higher magnetic field levels and also the higher growth rates observed under an applied static magnetic field will be further discussed in the next section.

Figure 29.18 represents the simulation results for the intermediate magnetic intensity levels selected as $B = 2.5$, 3.0, 4.0, and 8.0 kG. In this case, flow strength increases with magnetic field intensity. This result has not been widely reported in the relevant literature. It is possible that it is numerical in nature, and can be fixed by some innovative numerical treatments [29.56], but it is also possibly physically, as was observed experimentally in the LPEE growth of GaAs in [29.27].

Some experimental and numerical studies indeed demonstrated the enhancement of heat transfer (and hence flow strength) in a melt under a stationary magnetic field [29.57–59]. As can be seen from Fig. 29.18, the flow patterns show dramatic changes, as two flow cells were formed in each half of the vertical plane, with the upper cells getting larger and lower cells getting smaller with increasing magnetic field intensity. In the vertical plane, some strong unidirectional flows appear in the middle region along the r-direction, and some with very weak intensity in the middle, forming a small cylindrical region where the flow is nearly stationary. The flow fields are no longer axisymmetric and homocentric. Such an axisymmetry breaking may be caused by the unsteadiness of the flow field according to [29.60].

The flow stability was also examined at $B = 2.5$ and 3.0 kG levels. Comparing the results at different times, it was found that the flow field is essentially stable in spite of some small changes with time. One can then speculate that the reason for the axisymmetry breaking of the flow field at these magnetic field levels may be physical and due to a very delicate balance between the buoyancy and magnetic forces, even though the flow still remains stable. The variation of the flow strength along the radial direction for $B = 0.0$, 1.0, 2.0, and 3.0 kG levels are shown in Fig. 29.19. Solid and dashed lines represent, respectively, the values at the middle and the lower regions of the solution zone. As can be seen, the flow strength fluctuates and shows differences in these regions. However, at higher magnetic field levels, the difference becomes less obvious. The variation of the flow strength in the radial direction near the growth interface is almost symmetric in the absence of a magnetic field. However, this symmetry is broken at higher magnetic field intensity levels.

Although higher magnetic field intensity levels, higher than 4.5 kG, appear not to be practical for LPEE growth of thick crystals [29.26], for the sake of completeness and also for the purpose of comparison with other studies, higher magnetic field intensity levels of $B = 4.0$ and 8.0 kG were also considered. The flow patterns become dramatically different and show large fluctuations with time, and temperature distributions show asymmetric behavior.

The maximum flow strength (the maximum magnitude of the local velocity vector) values $U_{\max}$ are presented in Fig. 29.20 for all magnetic field levels. As seen, the variation of the maximum flow strength shows three distinct regions of stability. The flow strength decreases with increasing magnetic field in the stable region (up to Ha = 150), but in the intermediate and unstable regions (between Ha = 150 and 220, and above Ha = 220, respectively) the flow strength increases. If one examines the logarithmic plot of $U_{\max}$ as a function of the Hartmann number Ha given in Fig. 29.20, we can see that within the stable region, the relationship between $U_{\max}$ and Ha obeys a power law of

$$U_{\max} \propto \mathrm{Ha}^{-5/4} , \tag{29.33}$$

which has been demonstrated by many authors [29.51–55], although the index of the power law is slightly different due to different system parameters and conditions. In the unstable region, this relationship becomes

$$U_{\max} \propto \mathrm{Ha}^{5/2} . \tag{29.34}$$

In the transitional (intermediate) region, on the other hand, as one expects, the change of the maximum strength ($U_{\max}$) of the flow field with the Hartmann

number is so dramatic that the power law is not suitable. The results given here for the intermediate and unstable regions were obtained for the first time in [29.35], and to the best of our knowledge, are not corroborated by anyone else in the literature.

29.5.4 Evolution of Interfaces

As mentioned earlier, the simulation results above were given at the end of a 1 h growth period ($t = 1.0$ h) and also for a *stationary* interface, for computational efficiency. This was sufficient for examining the flow field and the effect of a magnetic field on the flow structures. However, in order to draw meaningful conclusions for future experiments, and also examining the concentration fields in the solution and the evolution of interfaces, simulations were carried for a longer period of growth ($t = 20.0$ h) for three levels of magnetic fields, namely $B = 0.0$ kG (no applied magnetic field), $B = 1.0$ kG (the mid-field level in the stable region), and $B = 3.0$ kG (a field at the beginning of the unstable region), and also the evolution of growth and dissolution interfaces are included. Higher magnetic field levels are not considered since the reliability of computations for concentrations in the unstable region may be questionable.

Figure 29.21 summarizes the computed results for the flow field in the horizontal plane at the middle of the growth cell (left column) and in the vertical plane at $\varphi = 0$ (right column) for three levels of magnetic field intensities ($B = 0.0$, 1.0, and 3.0 kG). Flow strengths are computed in m/s, and the scales are shown in each figure. The inclusion of the evolution of growth and dissolution interfaces affected the flow patterns to a certain extent, although it is not significant in terms of magnitudes. The effect of the magnetic field in the stable region is obvious in suppressing the natural convection in the solution. At the $B = 3.0$ kG level, however, the flow patterns show signs of unstable flows.

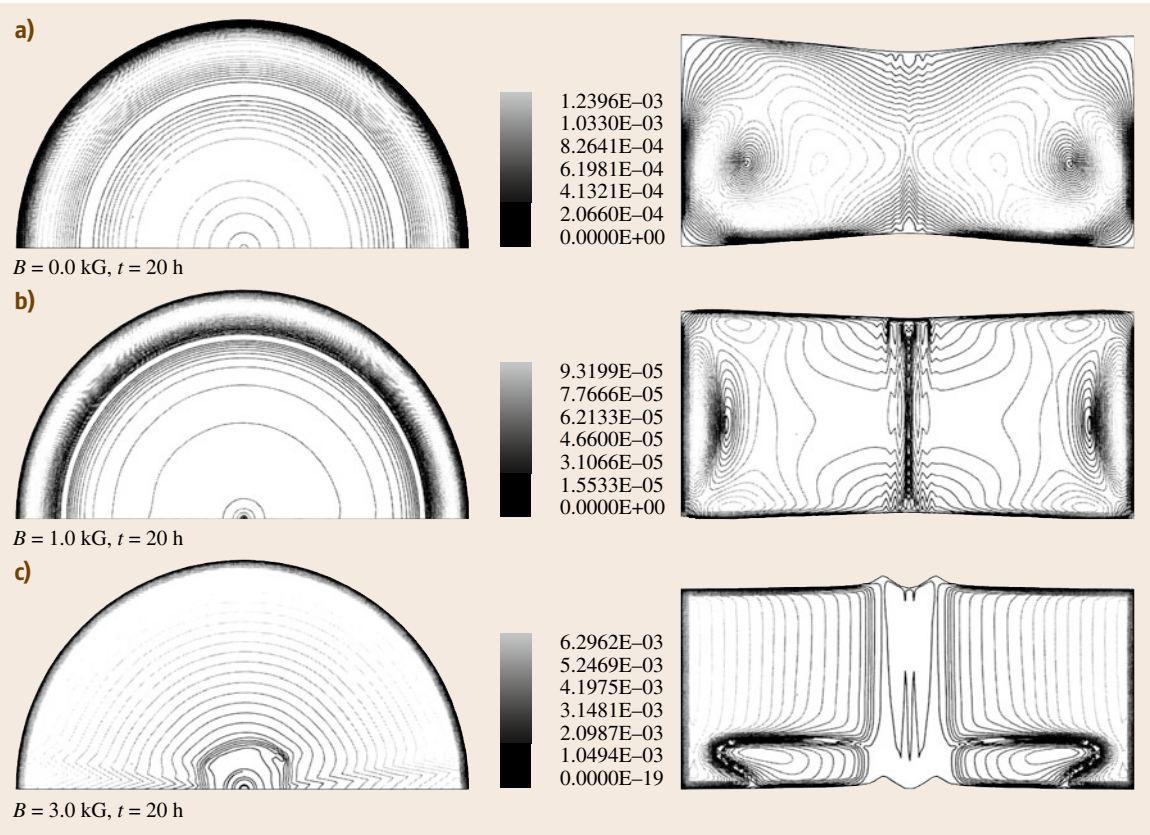


Fig. 29.21a–c Flow field in the horizontal plane in the middle of the growth cell (*left column*) and in the vertical plane at $\varphi = 0$ (*right column*) (after [29.35, 36])

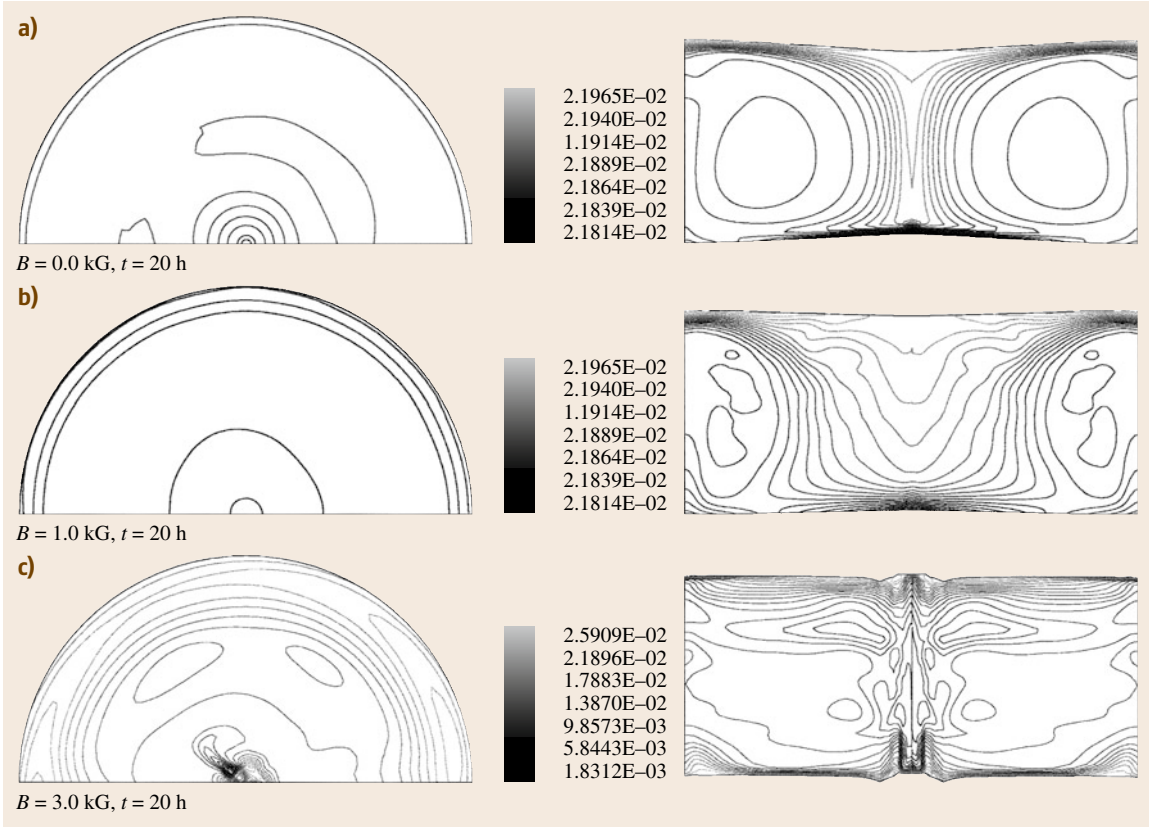


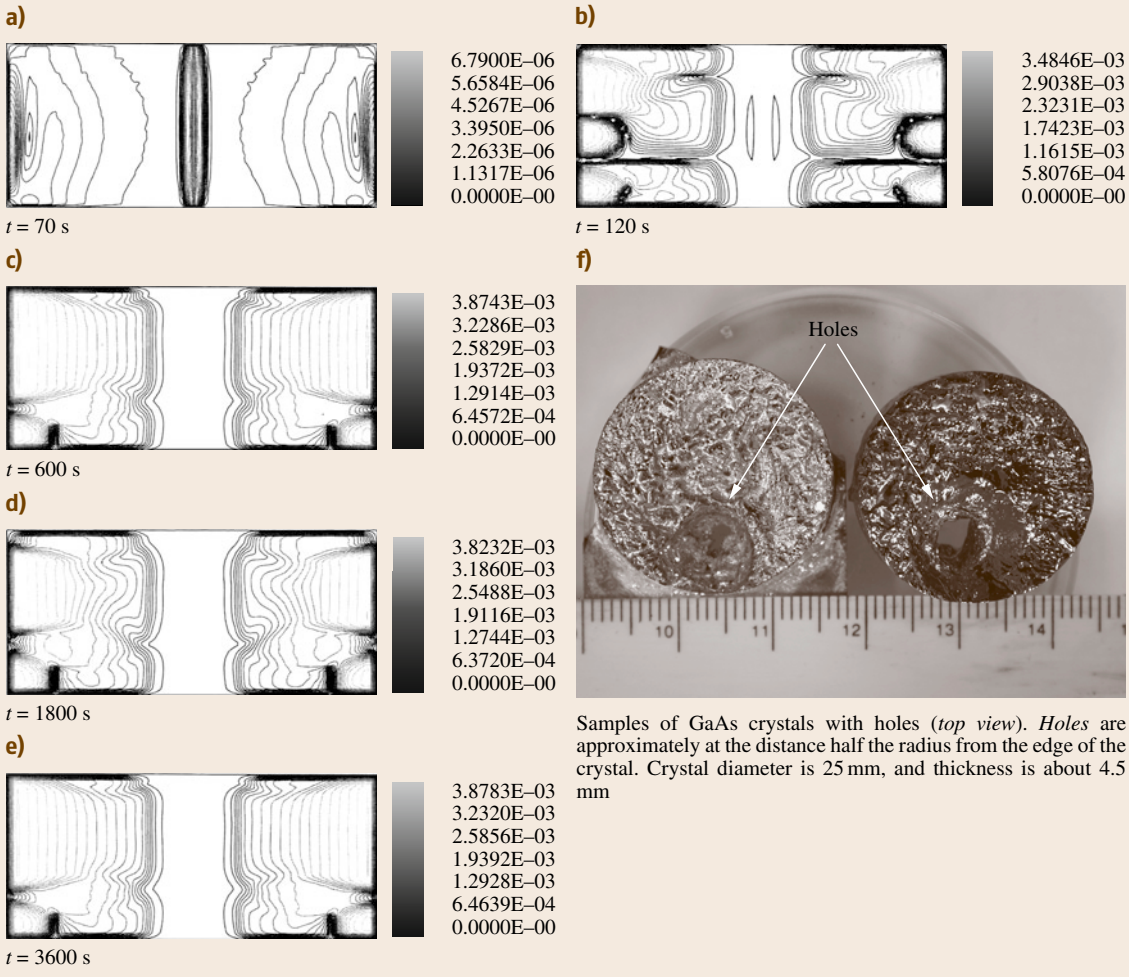
Fig. 29.22a–c Concentration distributions in the horizontal plane near the growth interface (*left column*) and in the vertical plane (*right column*) (after [29.35, 36])

The concentration distributions at the $B = 0.0$, 1.0, and 3.0 kG magnetic field levels are presented in Fig. 29.22 at $t = 20$ h. Comparing Fig. 29.22a,b one can see that the concentration gradients near the center of the growth interface and near the growth cell wall in the vicinity of the dissolution interface decrease with the increasing magnetic field levels. This is due to the reduction in flow strength, resulting in slower mass transfer towards the growth interface, consequently slowing down both the growth and dissolution rates. The concentration vortices in Fig. 29.22a nearly disappear in Fig. 29.22b. At the $B = 3.0$ kG level (Fig. 29.22c), concentration distributions exhibit significant changes compared with those of Fig. 29.22a,b. The concentration vortices near the center of the symmetric axis, in the vicinity of the growth interface, form strong concentration gradients and lead to fast growth rates in that region. Concentration layers are formed in the vicinity of both the growth and dissolution interfaces.

The computed growth rates obtained from simulations [29.35] under magnetic field do not predict the experimental growth rates [29.40]. We will discuss this issue later in detail.

29.5.5 Effect of High Electric and Magnetic Field Levels

The flow field was also numerically simulated under various electric current densities in [29.36]. The maximum flow strength shows the same trend at $J = 3$, 5, and 7 A/cm^2 electric field levels. The time evolution of the flow field was computed for various magnetic field intensity levels, and the one corresponding to $B = 4.0$ kG and $J = 7$ A/cm^2 is shown in Fig. 29.23. As can be seen, the flow patterns change with time and begin to become localized after 120 s of growth. This localization point is near the growth interface and is approximately at about a distance of a quarter of



Samples of GaAs crystals with holes (*top view*). *Holes* are approximately at the distance half the radius from the edge of the crystal. Crystal diameter is 25 mm, and thickness is about 4.5 mm

Fig. 29.23a–f Time evolution of the flow field (isostrength contours). The flow field is localized near the growth interface at about a distance of half of the crystal radius where approximately holes are observed (after [29.39]), samples from [29.27]

the diameter of the crystal from the edge. This point is almost at the locations where the holes (or damages) were observed in crystals grown under high fields in [29.27]. Simulations are shown in Fig. 29.23 for the liquid solution zone, but it is possible that similar flow patterns can also be computed in the liquid contact zone below the seed crystal. The flow patterns computed for $B = 3.0\,\mathrm{kG}$ and $J = 5\,\mathrm{A/cm^2}$ show similar patterns [29.36]. The numerical simulations are in qualitative agreement with experiments, but do not agree quantitatively on the critical value of the magnetic field. Numerical simulations predict a lower value (just over 2 kG) than the maximum experimental value of 4.5 kG. Crystals were grown in [29.26] at field values up to $B = 4.5\,\mathrm{kG}$, and $J = 7\,\mathrm{A/cm^2}$. However, experiments at higher fields failed (two samples of such crystals are shown in Fig. 29.23f).

Three-dimensional numerical simulation results have shown that magnetic field intensities up to 2.0 kG suppress the flow structures in the solution, and the flow structures are stable and get weaker with the increasing magnetic field level. These levels of magnetic field are beneficial in suppressing the natural convection. However, field intensities higher than 2.0 kG change the flow patterns significantly, and at intensities higher than 3.0 kG the flow structures become unstable.

As mentioned earlier, the LPEE growth system developed in [29.26] allowed the growth of a large number of GaAs and $Ga_{0.96}In_{0.04}As$ single crystals of thicknesses up to 9 mm. It was possible to apply electric current densities of 3, 5, and 7 A/cm^2, and the corresponding growth rates in these experiments with no magnetic field were, respectively, about 0.57, 0.75, and 1.25 mm/day. Growth interfaces were very flat, and the growth experiments were reproducible in terms of crystal thickness and growth rate. Experiments at higher electric current intensities were not successful. Experiments at 3, 5, and 7 A/cm^2 electric current densities were repeated under various applied static magnetic field levels. Results showed that LPEE experiments at the 4.5 kG and lower magnetic field levels were successful, but those under higher magnetic field levels were not.

In fact the experimental study of the LPEE growth of GaAs and $Ga_{0.96}In_{0.04}As$ single crystals supported qualitatively the results of the three-dimensional numerical simulations [29.26]. It seems that the 4.5 kG field intensity level is a maximum (*critical*) value above which the growth is not stable. This experimental *critical* magnetic field level is higher than that predicted from the numerical simulations performed under the same growth conditions, which was somewhere between 2.0 and 3.0 kG [29.35]. Considering the complexity of the LPEE growth process, this is a good qualitative agreement.

29.6 High Growth Rates in LPEE: Electromagnetic Mobility

The LPEE experiments conducted in [29.26] yielded a very significant result that was not predicted from the modeling studies conducted earlier: the experimental LPEE growth rates under magnetic field were much higher than the predicted values. For instance, the growth rate at the 4.5 kG magnetic field level (at $J = 3\,A/cm^2$) was about 6.1 mm/day, which is about 12 times higher than that with no magnetic field. Experiments performed at $B = 1.0$ and 2.0 kG field levels (at $J = 3\,A/cm^2$) were also successful, and the growth rates were also higher: 1.62 and 2.35 mm/day, respectively. One more interesting observation of the LPEE experiments was that the direction of the applied magnetic field, either up or down, was not relevant. The growth rate was almost the same, being about 5–6% less when the magnetic field was in the direction of the applied electric field. As predicted from the three-dimensional models, at higher magnetic field levels (even with the $J = 3\,A/cm^2$ electric current density level), and higher electric current density levels ($J = 10\,A/cm^2$ or higher), experiments did not lead to successful growth, but showed very interesting outcomes [29.26, 36].

Although very thick crystals were grown, even up to 9 mm thickness, the growth processes were unstable, and led to uneven growth (Figs. 29.23 and 29.24). From visual inspection of the grown crystals, the adverse effect of natural convection was obvious, causing either one-sided growth or leading to holes in the grown crystals. It was considered that such growth (one-sided and with holes) is because of the strong and unstable convection in the liquid zones (solution and contact zones) due to the strong interaction between the applied magnetic field and the applied electric current. Such predictions were also confirmed qualitatively by the numerical simulations carried out by considering field nonuniformities in [29.35], and also by using a newly defined electromagnetic mobility in [29.40, 61, 62] and a new model in [29.63]. The simulated flow structures show the possibility of causing such nonuniform growth of crystals.

The contribution of electromigration under magnetic field was obtained through a nonlinear model in [29.63] (for the binary GaAs system, in the absence of the Soret effect) as

$$\boldsymbol{i} = \rho_L D_C \nabla C + \rho_L \left(D_{EC} + D_{ECB} B\right) C \boldsymbol{E} \,. \qquad (29.35)$$

The second term in (29.35) represents the contribution of the applied electric current density to mass transport under the effect of a static external magnetic field. This term represents electromigration in the mass transport equation. Its coefficient, which is called the *total mobility* [29.61–63], is written in the following form for convenience

$$\mu_T \equiv D_{EC} + D_{ECB} B \equiv \mu_E + \mu_B B \,, \qquad (29.36)$$

where the material constant μ_E (a second-order material coefficient) is the classical electric mobility of the solute (As) in the liquid solution (Ga/As solution) due to the applied electric current in the absence of an applied magnetic field. The constant μ_B is a third-order

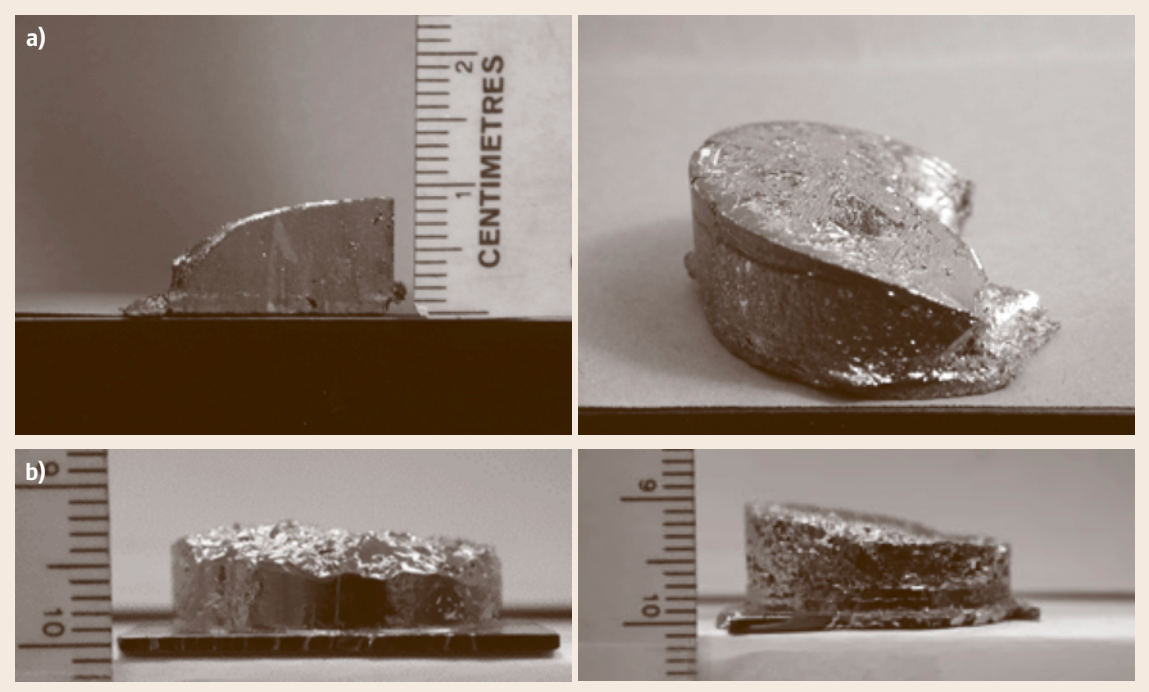


Fig. 29.24a,b Samples of LPEE-grown GaAs crystals under high fields: grown under **(a)** $B = 20\,\mathrm{kG}$ and $J = 3\,\mathrm{A/cm^2}$, and **(b)** $B = 0$, and $J = 10\,\mathrm{A/cm^2}$ (after [29.26])

material coefficient that represents the contribution of the applied magnetic field intensity to the electromigration of species. It is zero (or insignificant) in the absence of an applied electric current. This term is called the electromagnetic mobility of solute [29.63]. Its value is determined using the data of [29.26]. The mass transport equation then becomes

$$(\mu_\mathrm{E} + \mu_\mathrm{B} B)\,(\boldsymbol{E}\cdot\nabla C) + D_\mathrm{C}\nabla^2 C = \frac{\partial C}{\partial t} + \boldsymbol{v}\cdot\nabla C\,. \tag{29.37}$$

29.6.1 Estimation of the Electromagnetic Mobility Value

Experiments show that the growth rate is proportional to the applied electric current density, and we have evaluated the value of μ_E in the Ga/As solution in the absence of an applied magnetic field. The numerical simulations based on this value verify the experimental growth rates at all three electric current levels ($J = 3$, 5, and $7\,\mathrm{A/cm^2}$). Of course, diffusion (the second term in (29.37)) and also natural convection (the last term on the right-hand side of (29.37)) contribute to the growth rate. However, as shown many times numerically, in LPEE the contribution of the first term (electromigration) is dominant, and the growth rate can be assumed approximately proportional to this term.

Experiments also show that the growth rate increases significantly in the presence of a static magnetic field, and is also proportional to the field intensity level as long as the field level is below a critical value, above which the growth is not stable [29.27]. Numerical values of the mobilities are calculated using the results of a large number of experiments in [29.26] in which the magnetic field vector B was applied both upward and downward. The growth rates in these experiments were almost the same whether B was up or down. In other words the mass transport due to electromigration was only dependent on the magnetic field intensity but not on its direction. This is also in compliance with the defined constitutive equations in [29.63]. Using the measured growth rates, the mobility values were computed (Table 29.5).

The dimensionless mobility is defined as

$$\mu = \frac{\mu_\mathrm{T}}{\mu_\mathrm{E}} = 1 + \frac{\mu_\mathrm{B}}{\mu_\mathrm{E}} \cong 1 + 2B\,, \tag{29.38}$$

Table 29.5 Numerical values of mobilities [29.26, 61, 63]

Experimental values				
Magnetic field (kG)	0.0	1.0	2.0	4.5
Electric current density (A/cm^2)	3.0	3.0	3.0	3.0
Growth rate (mm/day)	0.50	1.62	2.35	6.10
Computed values				
Electric mobility constant μ_E ($m^2/V\,s$)	0.7×10^{-5}	0.7×10^{-5}	0.7×10^{-5}	0.7×10^{-5}
Total mobility $\mu_T=\mu_E+\mu_B B$ ($m^2/V\,s$)	0.7×10^{-5}	2.3×10^{-5}	3.4×10^{-5}	7.1×10^{-5}
Electromagnetic mobility $\mu_B B$ ($m^2/V\,s$)	0.0	1.6×10^{-5}	2.7×10^{-5}	6.4×10^{-5}
Electromagnetic mobility constant μ_B ($m^2/V\,s\,kG$)	–	1.4×10^{-5}	1.4×10^{-5}	1.4×10^{-5}
Dimensionless mobility $\mu=\mu_T/\mu_B\cong 1+2B$	1	3	5	10

and is plotted in Fig. 29.25. As seen, the total mobility is almost linearly dependent on the magnetic field intensity, of course, within the limit of experimental measurements.

The first term in the mass transport equation reads explicitly

$$(\mu_E+\mu_B B)\,(\boldsymbol{E}+\boldsymbol{v}\times\boldsymbol{B})\cdot\nabla C\,, \tag{29.39}$$

where the term $(\boldsymbol{v}\times\boldsymbol{B})\cdot\nabla C$ is the contribution of the applied magnetic field due to the motion of the fluid particles (coupling term). Its contribution was found to be very small compared with that of $\boldsymbol{E}\cdot\nabla C$ (on the order of 3% based on a maximum velocity of 0.01 m/s and a 10 kG field level [29.35, 39]). Therefore, its contribution can be neglected in the model for computer simulations. Then, the electromigration term in the mass transport equation is written as

$$(\mu_E+\mu_B B)\,\boldsymbol{E}\cdot\nabla C=\mu_T\boldsymbol{E}\cdot\nabla C\,, \tag{29.40}$$

and the growth rate is computed by

$$V_n^g=\frac{\rho_L}{\rho_S}\left(D_C\frac{\partial C}{\partial n}+\mu_T C E_n\right)\frac{1}{C_S-C}\,, \tag{29.41}$$

or simply by, for the purpose of evaluating the mobility constants,

$$V^g=\frac{\rho_L}{\rho_S}\left(D_C\frac{\partial C}{\partial n}+\mu_T C E_z\right)\frac{1}{C_S-C}\,. \tag{29.42}$$

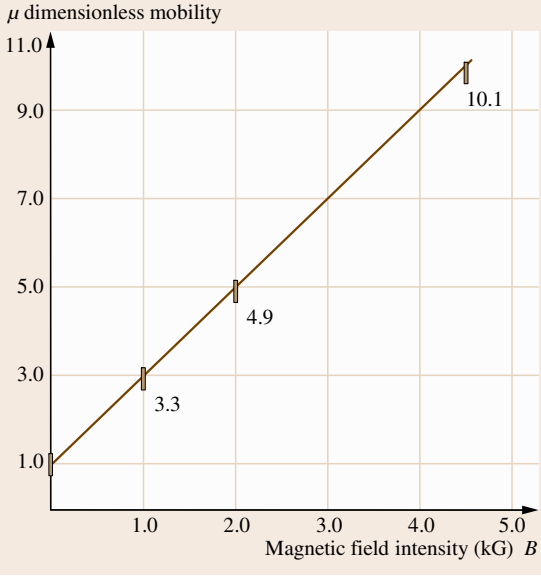


Fig. 29.25 Dependence of the total As mobility on magnetic field intensity (after [29.61, 63])

29.6.2 Simulations of High Growth Rates in a GaAs System

Simulations presented in the previous sections were repeated using the total mobility values in the mass transport equation (29.37). In earlier simulations only the electric mobility μ_E was used. A summary of the growth rates from these numerical simulations is presented in Fig. 29.26a. The values under no magnetic field are the experimental growth rates and are used to compute the value of μ_E. Naturally, they are coincident with the computed values. As seen, the growth rate decreases first with the magnetic field level and then increases with the magnetic field above the critical value. This pattern is similar to the pattern of experimental growth rates under various magnetic field levels, and also agrees with the numerical simulation results.

In order to predict the high experimental growth rates under an applied magnetic field, the mass transport equation in (29.24) was replaced with (29.37), and

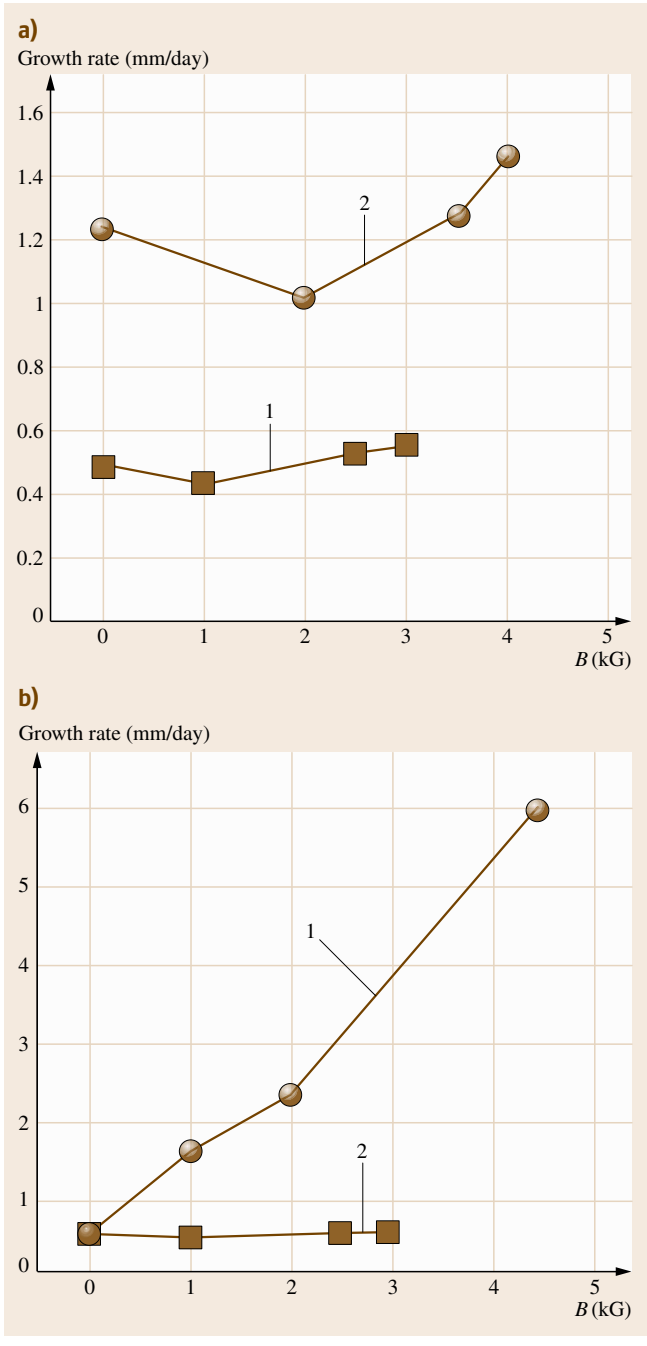


Fig. 29.26 **(a)** Computed growth rates are presented versus applied magnetic field with the use of constant electric mobility μ_E. *Squares* represent the values at $J = 3\,A/cm^2$ and *circles* denote for the values at $J = 7\,A/cm^2$. **(b)** Growth rates computed using the total mobility, $\mu_T = \mu_E + \mu_B B$, are represented by *full circles*. These values are in agreement with the experimental growth rates (*hollow circles*, invisible as coincident with the *full circles*). The growth rates under no magnetic field are also shown (*squares*) for comparison (after [29.36]) ▶

then the 3-D simulations were repeated using the total electromagnetic mobility $\mu_T = \mu_E + \mu_B B$ values given in Table 29.5. The growth rates computed using the total mobility μ_T are presented in Fig. 29.26b (full circles), and agree with those of experiments. The growth rates using only the electric mobility μ_E are also given in Fig. 29.26b for comparison (empty squares). For the sake of completeness and for comparison, the experimental growth rates under a magnetic field are also presented in Fig. 29.26b (note that the full and empty circles are coincident).

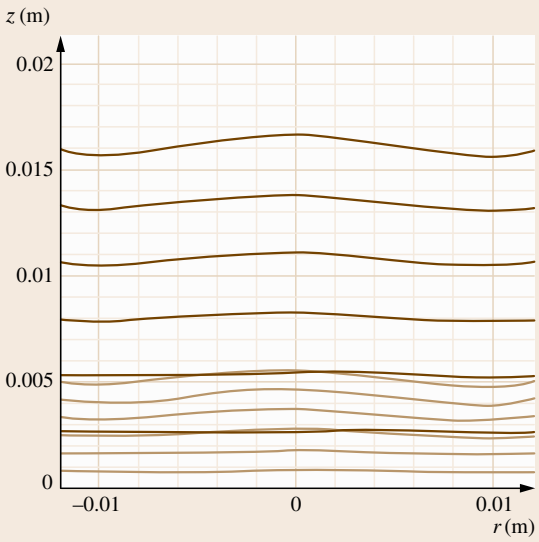


Fig. 29.27 Evolution of the computed growth interface using the total mobility $\mu_T = \mu_E + \mu_B B$ (*thick lines*) and using only the electric mobility μ_E (*thin lines*). Time increment between lines is 40 h. The shapes of the computed interfaces are in excellent agreement with those of experiments with and without magnetic field (see sample GaAs crystals in Fig. 29.11). It must be noted that the numerical simulations were let go until 15 mm growth thickness just to show that the interface remains flat. Thicknesses of the grown GaAs and GaInAs crystals were less than this value

The above results show that the inclusion of the new total mobility that includes the contributions of both

the applied electric current and also its interaction with the applied magnetic field provide much more accurate predictions for growth rates in LPEE under a magnetic field.

When only the electric mobility is used in simulations, not only the growth rates but also the shapes of the evolving growth interface were not predicted accurately. However, when the total mobility was introduced, in addition to better predictions for growth rates, the growth interface shapes were also much closer to those of experiments.

In Fig. 29.27 the evolution of computed growth interface is presented. As seen, the computed interfaces are flatter and agree with the shapes of the interfaces of the crystals grown in [29.26], which are almost perfectly flat.

These results show that the introduction of a bulk constitutive coefficient representing the total electromagnetic mobility due to electromigration under magnetic fields in LPEE (the nonlinear model for LPEE under magnetic field in [29.63]) is a step in the right direction. It is by no means complete, since the interaction of electric and magnetic fields must also affect mass transport at the growth interface. Therefore, in addition to this model in the bulk (solution), a closer look may also be needed at various surface phenomena under the combined effect of applied electric and magnetic fields, in order to obtain better predictions from modeling.

References

29.1 M. Kumagawa, A.F. Witt, M. Lichtensteiger, H.C. Gatos: Current-controlled growth and dopant modulation in liquid-phase epitaxy, J. Electrochem. Soc. **120**, 583/584 (1973)

29.2 J.J. Daniele: Peltier-induced LPE and composition stabilization of GaAlAs, Appl. Phys. Lett. **27**, 373 (1975)

29.3 J.J. Daniele: Experiments showing absence of electromigration of As and Al in Peltier LPE of GaAs and $Ga_{1-x}Al_xAs$, J. Electrochem. Soc. **124**, 1143 (1977)

29.4 V.A. Gevorkyan, L.V. Golubev, S.G. Petrosyan, Y.A. Shik, Y.V. Shmatsev: Sov. Phys. Tech. Phys. **22**, 750–755 (1977)

29.5 L. Jastrzebski, H.C. Gatos, A.F. Witt: Electromigration in current-controlled LPE, J. Electrochem. Soc. **123**, 1121 (1976)

29.6 L. Jastrzebski, H.C. Gatos: Current-controlled growth, segregation and amphoteric behavior of Si in GaAs from Si-doped solutions, J. Cryst. Growth **42**, 309–314 (1977)

29.7 L. Jastrzebski, H.C. Gatos, A.F. Witt: Current-induced solution growth of garnet layers, J. Electrochem. Soc. **124**, 633 (1977)

29.8 L. Jastrzebski, Y. Imamura, H.C. Gatos: Thickness uniformity of GaAs layers grown by electroepitaxy, J. Electrochem. Soc. **125**, 1140 (1978)

29.9 L. Jastrzebski, J. Lagowski, H.C. Gatos, A.F. Witt: Determination of carrier concentration distribution in semiconductors by IR absorption – Si, J. Appl. Phys. **49**, 5909 (1978)

29.10 T. Bryskiewicz: Peltier-induced growth kinetics of liquid-phase epitaxial GaAs, J. Cryst. Growth **43**, 567–571 (1978)

29.11 A.F. Witt, H.C. Gatos, M. Lichtensteiger, C.J. Herman: Crystal-growth and segregation under zero gravity – Ge, J. Eletrochem. Soc. **125**, 1832 (1978)

29.12 T. Bryskiewicz, A. Laferriere: Growth of alloy substrates by liquid-phase electroepitaxy – theoretical considerations, J. Cryst. Growth **129**, 429–442 (1993)

29.13 Y.H. Lo, R. Bhat, P.S.D. Lin, T.P. Lee: Long-wavelength optoelectronic integrated circuit transmitter, Proc. SPIE **1582**, 60–70 (1992)

29.14 W.P. Hong, G.K. Chang, R. Bhat, C. Nguyen, H.P. Lee, L. Wong, K.R. Runge: InP-based MSM-HEMT receiver OEICs for long-wavelength light wave systems, Proc. SPIE **1582**, 134–144 (1992)

29.15 W.J. Schaff, P.J. Tasker, M.C. Foisy, L.F. Eastman: Device applications of stained layer epitaxy. In: *Semiconductors and Semimetals*, Vol. 33, ed. by T.P. Pearsall (Academic, New York 1991) pp. 73–133

29.16 L.F. Eastman: In: *Optoelectronic Materials and Devices Concepts*, ed. by M. Razeghi (SPIE Optical Engineering Press, Bellingham 1991) p. 41

29.17 A. Andaspaeva, A.N. Baranov, A. Guseinov, A.N. Imenkov, L.M. Litvak: Sov. Tech. Phys. Lett. **14**, 377 (1988)

29.18 H.K. Choi, S.J. Eglash: High-efficiency high-power GaInAsSb-AlGaAsSb double-heterostructure lasers emitting at 2.3 μm, IEEE J. Quantum Electron. **27**, 1555 (1991)

29.19 S.J. Eglash, H.K. Choi: MBE growth, material properties, and performance of GaSb-based 2.2 μm diode-lasers, Inst. Phys. Conf. Ser. **120**(10), 487 (1992)

29.20 T. Bryskiewicz, C.F. Boucher Jr., J. Lagowski, H.C. Gatos: Bulk GaAs crystal-growth by liquid-phase electroepitaxy, J. Cryst. Growth **82**, 279–288 (1987)

29.21 S. Dannefear, P. Mascher, D. Kerr: Annealing of grown-in defects in GaAs, Proc. MRS **104**, 471 (1978)

29.22 C.F. Boucher Jr., O. Ueda, T. Bryskiewicz, J. Lagowski, H.C. Gatos: Elimination of disloca-

tions in bulk GaAs crystals grown by liquid-phase electroepitaxy, J. Appl. Phys. **61**, 359–364 (1987)
29.23 T. Bryskiewicz, M. Bugajski, J. Lagowski, H.C. Gatos: Growth and characterization of high quality LPEE GaAs bulk crystals, J. Cryst. Growth **85**, 136–141 (1987)
29.24 T. Bryskiewicz, M. Bugajski, B. Bryskiewicz, J. Lagowski, H.C. Gatos: LPEE growth an characterization of In(V)Ga/As(V) crystal, Proc. Inst. Phys. Ser. **91**(3), 259 (1988)
29.25 B. Bryskiewicz, T. Bryskiewicz, E. Jiran: Internal strain and dislocations in $In_xGa_{1-x}As$ crystals grown by liquid phase electroepitaxy, J. Electron. Mater. **24**, 203 (1995)
29.26 H. Sheibani, S. Dost, S. Sakai, B. Lent: Growth of bulk single crystals under applied magnetic field by liquid phase electroepitaxy, J. Cryst. Growth, **258**(3-4), 283–295 (2003)
29.27 H. Sheibani, Y.C. Liu, S. Sakai, B. Lent, S. Dost: The effect of applied magnetic field on the growth mechanisms of liquid phase electroepitaxy, Int. J. Eng. Sci. **41**, 401–415 (2003)
29.28 G. Bischopink, K.W. Benz: THM growth of $Al_xGa_{1-x}Sb$ bulk crystals, J. Cryst. Growth **128**, 466–470 (1993)
29.29 J.J. Daniele, A.J. Hebling: Peltier-induced liquid-phase epitaxy and compositional control of mm-thick layers of (Al,Ga)As, J. Appl. Phys. **52**, 4325–4327 (1981)
29.30 Z.R. Zytkiewicz: Influence of convection on the composition profiles of thick GaAlAs layers grown by liquid phase electroepitaxy, J. Cryst. Growth **131**, 426–430 (1993)
29.31 Z.R. Zytkiewicz: Liquid phase electroepitaxial growth of thick and compositionally uniform AlGaAs layers on GaAs substrates, J. Cryst. Growth **146**, 283–286 (1995)
29.32 Z.R. Zytkiewicz: Joule effect as a barrier for unrestricted growth of bulk crystals by liquid phase electroepitaxy, J. Cryst. Growth **172**, 259–268 (1997)
29.33 C. Takenaka, K. Nakajima: Effect of electric-current on the LPE growth of InP, J. Cryst. Growth **108**, 519 (1991)
29.34 N. Djilali, Z. Qin, S. Dost: Role of thermosolutal convection in liquid phase electroepitaxial growth of gallium arsenide, J. Cryst. Growth **149**, 153–166 (1995)
29.35 Y.C. Liu, Y. Okano, S. Dost: The effect of applied magnetic field on flow structures in liquid phase electroepitaxy – A three-dimensional simulation model, J. Cryst. Growth **244**, 12–26 (2002)
29.36 Y.C. Liu, S. Dost, H. Sheibani: A three dimensional numerical simulation for the transport structures in liquid phase electroepitaxy under applied magnetic field, Int. J. Trans. Phenom. **6**, 51–62 (2004)
29.37 R.W. Wilcox: Influence of convection on the growth of crystals from solution, J. Cryst. Growth **65**, 133 (1983)
29.38 S. Ostrach: Fluid mechanics in crystal growth – The 1982 Freeman scholar lecture, J. Fluids Eng. **105**, 5–20 (1983)
29.39 S. Dost, Y.C. Liu, B. Lent: A numerical simulation study for the effect of applied magnetic field in liquid phase electroepitaxy, J. Cryst. Growth **240**, 39–51 (2002)
29.40 S. Dost, B. Lent, H. Sheibani, Y.C. Liu: Recent developments in liquid phase electroepitaxial growth of bulk crystals under magnetic field, C. r. Mec. **332**(5/6), 413–428 (2004)
29.41 S. Dost, B. Lent: *Single Crystal Growth of Semiconductors from Metallic Solutions* (Elsevier, Amsterdam 2007)
29.42 M. Yildiz, S. Dost, B. Lent: Growth of bulk SiGe single crystals by liquid phase diffusion, J. Cryst. Growth **280**(1-2), 151–160 (2005)
29.43 M. Yildiz, S. Dost: A continuum model for the liquid phase diffusion growth of bulk SiGe single crystals, Int. J. Eng. Sci. **43**, 1059–1080 (2005)
29.44 E. Yildiz, S. Dost, M. Yildiz: A numerical simulation study for the effect of magnetic fields in liquid phase diffusion growth of SiGe single crystals, J. Cryst. Growth **291**, 497–511 (2006)
29.45 M. Yildiz, S. Dost, B. Lent: Evolution of the growth interface in liquid phase diffusion growth of bulk SiGe single crystals, Cryst. Res. Technol. **41**(3), 211–216 (2006)
29.46 S. Dost, Z. Qin: A model for liquid phase electroepitaxial growth of ternary alloy semiconductors – 1. Theory, Int. J. Electromagn. Mech. **7**(2), 109–128 (1996)
29.47 Y. Imamura, L. Jastrzebski, H.C. Gatos: Defect structure and electronic charateristics of GaAs layers grow by electroepitaxy and thermal LPE, J. Electrochem. **126**(8), 1381–1385 (1979)
29.48 Z. Qin, S. Dost, N. Djilali, B. Tabarrok: A finite element model for liquid phase electroepitaxy, Int. J. Numer. Methods Eng. **38**(23), 3949–3968 (1995)
29.49 Z. Qin, S. Dost, N. Djilali, B. Tabarrok: A model for liquid phase electroepitaxy under an external magnetic field – 2. Application, J. Cryst. Growth **153**, 131–139 (1995)
29.50 Z. Qin, S. Dost: A model for liquid phase electroepitaxial growth of ternary alloy semiconductors – 2. Application, Int. J. Appl. Electromagn. Mech. **7**(2), 129–142 (1996)
29.51 H. Ben Hadid, D. Henry: Numerical study of convection in the horizontal Bridgman configuration under the action of a constant magnetic field. Part 1. Two-dimensional flow, J. Fluid Mech. **333**, 23–56 (1996)
29.52 H. Ben Hadid, D. Henry: Numerical study of convection in the horizontal Bridgman configuration under the action of a constant magnetic field. Part 2. Three-dimensional flow, J. Fluid Mech. **333**, 57–83 (1996)

29.53 H. Ben Hadid, S. Vaux, S. Kaddeche: Three-dimensional flow transitions under a rotating magnetic field, J. Cryst. Growth **230**, 57–62 (2001)

29.54 D.H. Kim, P.M. Adornato, R.A. Brown: Effect of vertical magnetic field on convection and segregation in vertical Bridgman crystal growth, J. Cryst. Growth **89**, 339 (1988)

29.55 L. Davoust, M.D. Cowley, R. Moreau, R. Bolcato: Buoyancy-driven convection with a uniform magnetic field – Part 2. Experimental investigation, J. Fluid Mech. **400**, 59 (1999)

29.56 V. Kumar, S. Dost, F. Durst: Numerical modeling of crystal growth under strong magnetic fields: An application to the travelling heater method, Appl. Math. Modell. **31**(3), 589–605 (2006)

29.57 T. Tagawa, H. Ozoe: Enhancement of heat transfer rate by application of a static magnetic field during natural convection of liquid metal in a cube, J. Heat Transf. **119**, 265 (1997)

29.58 T. Tagawa, H. Ozoe: Enhanced heat transfer rate measured for natural convection in liquid gallium in a cubical enclosure under a static magnetic field, J. Heat Transf. **120**, 1027 (1998)

29.59 K. Terashima, J. Nishio, S. Washizuka, M. Watanabe: Magnetic field effect on residual impurity concentrations for LEC GaAs crystal growth, J. Cryst. Growth **84**, 247 (1987)

29.60 A.Y. Gelfgat, P.Z. Bar-Yoseph, A. Solan: Axisymmetry breaking instabilities of natural convection in a vertical Bridgman growth configuration, J. Cryst. Growth **220**, 316 (2000)

29.61 S. Dost, H. Sheibani, Y.C. Liu, B. Lent: Recent developments in modelling of liquid phase electroepitaxy under applied magnetic field, Cryst. Res. Technol. **40**(4/5), 313 (2005)

29.62 S. Dost, H. Sheibani, Y.C. Liu, B. Lent: On the high growth rates in electroepitaxial growth of bulk semiconductor crystals in magnetic field, J. Cryst. Growth **275**(1-2), e1–e6 (2005)

29.63 S. Dost, H. Sheibani: A mathematical model for solution growth of bulk crystals under magnetic field, Philos. Mag. **85**(33-35), 4331–4351 (2005)

29.64 K. Nakajima: Liquid phase epitaxial growth of very thick $In_{1-x}Ga_xAs$ layers with uniform composition by source-current-controlled method, J. Appl. Phys. **61**(9), 4626 (1987)

29.65 K. Nakajima, S. Yamazaki, I. Umebu: A new growth method using source current control to supply solute elements-demonstration of $In_{1-x}Ga_xAs$ case, Jpn. J. Appl. Phys. **23**(1), L26–L28 (1984)

29.66 K. Nakajima, S. Yamazaki: A new method to supply solute elements into growth solutions – Demonstration by liquid phase epitaxial growth of $In_{1-x}Ga_xAs$, J. Electrochem. Soc. **132**, 904 (1985)

29.67 K. Nakajima, S. Yamazaki: Growth of very thick $In_{1-x}Ga_xAs$ layers by source-current-controlled method, J. Cryst. Growth **74**, 39–47 (1986)

29.68 K. Nakajima: Calculation of composition variation of $In_{1-x}Ga_xAs$ ternary crystals for diffusion and electromigration limited growth from a temperature graded solution with source material, J. Cryst. Growth. **110**, 781–794 (1991)

29.69 K. Nakajima: Calculation of stresses in $In_{0.12}Ga_{0.88}As$ ternary bulk crystals with compositionally graded $In_{1-x}Ga_xAs$ layers on GaAs seeds, J. Cryst. Growth **113**, 477–484 (1991)

29.70 S. Dost, N. Djilali, Z. Qin: A two-dimensional diffusion model for liquid-phase electroepitaxial growth of GaAs, J. Cryst. Growth **143**(3/4), 141–154 (1994)

29.71 S. Dost, H.A. Erbay: A continuum model for liquid phase electroepitaxy, Int. J. Eng. Sci. **33**(10), 1385–1402 (1995)

29.72 S. Dost, Z. Qin: A model for liquid-phase electroepitaxy under an external magnetic field. Part 1. Theory, J. Cryst. Growth **153**, 123–130 (1995)

29.73 S. Dost, Z. Qin: A numerical simulation model for liquid phase electroepitaxial growth of GaInAs, J. Cryst. Growth **187**, 51–64 (1998)

29.74 A. Okamoto, S. Isozumi, J. Lagowski, H.C. Gatos: In situ monitoring of liquid phase electroepitaxial growth, J. Electrochem. Soc. **129**, 2095–2098 (1982)

30. Epitaxial Lateral Overgrowth of Semiconductors

Zbigniew R. Zytkiewicz

The state of the art and recent developments of lateral overgrowth of compound semiconductors are reviewed. First we focus on the mechanism of epitaxial lateral overgrowth (ELO) from the liquid phase, highlighting the phenomena that are crucial for growing high-quality layers with large aspect ratio. Epitaxy from the liquid phase has been chosen since the equilibrium growth techniques such as liquid-phase epitaxy (LPE) are the most suitable for lateral overgrowth. We then present numerous examples for which the defect filtration in the ELO procedure is very efficient and leads to significant progress in the development of high-performance semiconductor devices made of lattice-mismatched structures. Structural perfection of seams that appear when layers grown from neighboring seeds merge is also discussed. Next, we concentrate on strain commonly found in various ELO structures and arising due to the interaction of ELO layers with the mask. Its origin, and possible ways of its control, are presented. Then we show that the thermal strain in lattice-mismatched ELO structures can be relaxed by additional tilting of ELO wings while still preserving their high quality. Finally, recent progresses in the lateral overgrowth of semiconductors, including new mask materials and liquid-phase electroepitaxial growth on substrates coated by electrically conductive masks, are presented. New versions of the ELO technique from solution and from the vapor (growth from ridges and pendeo-epitaxy) are described and compared with standard ELO. A wide range of semiconductors, including III–V compounds grown from solution and vapor-grown GaN, are used to illustrate phenomena discussed. Very often, the similar behavior of various ELO structures reveals that the phenomena presented are not related to a specific group of compounds or their growth techniques, but have a much more general nature.

30.1 **Overview** 1000

30.2 **Mechanism of Epitaxial Lateral Overgrowth from the Liquid Phase** 1002
30.2.1 Choice of Substrate Geometry for Growth of ELO Layers 1004
30.2.2 Optimization of Liquid-Phase Lateral Overgrowth Procedure 1007

30.3 **Dislocations in ELO Layers** 1011
30.3.1 Filtration of Substrate Dislocations in ELO 1011
30.3.2 Structural Perfection of Coalescence Front in Fully Overgrown ELO Structures 1014

30.4 **Strain in ELO Layers** 1016
30.4.1 Mask-Induced Strain in Homoepitaxial ELO Layers 1017
30.4.2 Thermal Strain in ELO Layers 1024

30.5 **Recent Progress in Lateral Overgrowth of Semiconductor Structures** 1026
30.5.1 Developments in Liquid-Phase ELO Growth 1027
30.5.2 New Concepts of ELO Growth 1030

30.6 **Concluding Remarks** 1034

References 1035

30.1 Overview

Modern micro- and optoelectronic devices consist of thin layers grown epitaxially on a substrate. These layers must be of high crystallographic quality. Very often however, there are no substrates available for lattice-matched epitaxy, and dislocations are generated at the layer–substrate interface to relax the lattice-mismatch strain. Segments of these dislocations thread to the surface of the epilayer, and then to the next-grown layers of the structure, deteriorating parameters of the device and leading to its fast degradation. To reduce the density of threading dislocations in lattice-mismatched heteroepitaxial structures a buffer layer with a graded or abrupt composition profile is usually deposited on an available substrate to obtain a layer with the required lattice parameter value. Strained layer superlattices are often inserted into the buffer to bend threading dislocations and partially prevent them from propagating to the surface [30.1]. In situ or ex situ annealing of buffers has also been employed to induce mutual reactions between threading dislocations that eventually lead to their annihilation [30.2]. In addition, for heterostructures with large lattice mismatch, buffer growth is usually initiated at relatively low temperature, and is then continued at a higher temperature (so-called two-step growth) [30.3]. All these sophisticated methods of buffer layer engineering have been proven to reduce significantly density of defects in lattice-mismatched heterostructures. However, despite the progress made, the best modern buffers still contain dislocations with a density of $\approx 10^6\,\mathrm{cm}^{-2}$, which is often too high for device applications. Therefore, the epitaxial lateral overgrowth (ELO) technique has been developed to block dislocations threading from the substrate or substrate/buffer structure to the next-grown epitaxial layers.

Figure 30.1 schematically illustrates the principle of epitaxial lateral overgrowth. First the substrate or substrate with a suitable buffer is covered by a thin amorphous masking film. Dielectrics, such as SiO_2 or Si_3N_4, are commonly used to mask the substrate, but other materials such as tungsten [30.4, 5], zirconium nitride [30.6] or graphite [30.7] have also been successfully applied for the ELO of III–V semiconductors. In the next step the substrate is patterned by conventional photolithography and etching to form on its whole area a grating of mask-free seeding windows of width W separated by masked areas of width M (Fig. 30.1a). The ELO growth starts by a selective-area epitaxy, i. e., nucleation takes place exclusively in narrow windows (Fig. 30.1b). When the layer becomes thicker than the masking film the growth proceeds also in a lateral direction over the mask. Finally, a new epitaxial layer fully covers the patterned substrate if the growth time is sufficiently long for coalescence of ELO stripes grown from neighboring seeds. The main advantage of the ELO process is that dislocations can thread from the substrate into the epitaxial layer through the very narrow windows, so respective etch pits should be visible only in the part of the ELO layer grown vertically from the seed (Fig. 30.1b). Since there is no continuity of crystal lattice at the interface between monocrystalline substrate and the amorphous mask, substrate dislocations cannot propagate into the mask and must vanish at the substrate surface. In this way the mask efficiently blocks dislocations and their density in the laterally grown parts (wings) of the layer should be significantly reduced. Indeed, there are many experimental evidences showing that the density of dislocations in the wings area is orders of magnitude lower than that observed in standard planar epilayers (Sect. 30.3.1). Thus the ELO process

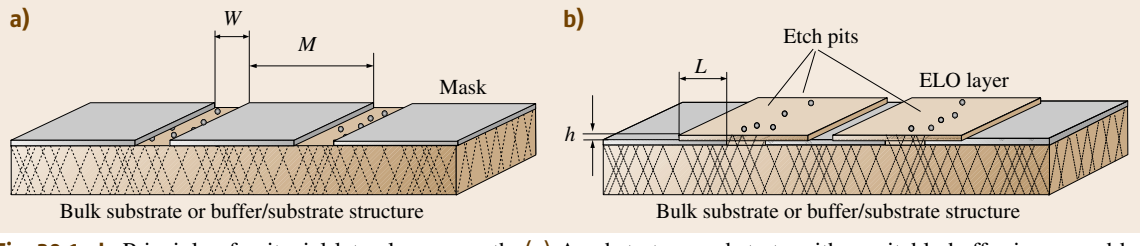


Fig. 30.1a,b Principle of epitaxial lateral overgrowth. **(a)** A substrate or substrate with a suitable buffer is covered by a thin amorphous masking film. Seeding windows of width W separated by masked areas of width M are opened up in the mask. **(b)** The ELO growth begins exclusively inside the seeding windows. Then the growth proceeds laterally over the mask. Note that substrate dislocations, marked by *dotted lines*, propagate to the ELO layer through the openings in the mask, only. L and h denote the width of the laterally grown part (wing) of the layer and its thickness, respectively

is a powerful method to grow epilayers with a low density of dislocations on heavily dislocated substrates such as the relaxed buffers deposited on lattice-mismatched substrates. When combined with well-developed methods of buffer layers engineering the ELO technique offers the possibility of producing high-quality substrates with adjustable value of lattice constant required by modern electronics. If ternary buffers are used the value of the lattice parameter might be controlled by choosing the proper composition of the buffer. Next, homoepitaxial lattice matched lateral overgrowth would be employed to remove the buffer dislocations.

A more detailed discussion on the dislocation filtration processes in ELO together with some examples will be presented in Sect. 30.3. However, it must be mentioned here that, according to the ELO concept presented in Fig. 30.1, the number of dislocations propagating into the layer should decrease as the width W of the seeding window decreases. In particular, the growth of a completely dislocation-free layer might be possible if the window width is smaller than the average distance between dislocations in the substrate. This means that for the density of dislocations of $\approx 10^{10}\,\mathrm{cm}^{-2}$ the condition $W \ll 100\,\mathrm{nm}$ should be fulfilled. In principle, windows a few nanometers wide should be sufficient for the transfer of information on the lattice parameter from the substrate to the epilayer, i. e., for epitaxial growth. Due to technical reasons, however, windows a few micrometers wide are usually applied. Note also that in many epitaxial systems dislocations threading through the seeding windows are inclined at some angle to the epilayer surface, so the width of dislocated area on the surface increases as the layer grows thicker [30.8]. Thus, the ELO layers should be as wide and as thin as possible in order to maximize dislocation-free wing area available for devices. In other words, the largest value of the aspect (width/thickness) ratio AR, defined in this work by the equation $\mathrm{AR} = (2L + W)/h$, is required.

Besides the use for defect filtration during the heteroepitaxy of lattice-mismatched systems, the ELO technique in its homoepitaxial version has also found practical applications in the production of silicon-on-insulator structures [30.9], metal-oxide-semiconductor (MOS) transistors [30.10], field-effect transistors [30.11], solar cells [30.12], pressure sensors [30.13], and for three-dimensional device integration [30.14]. If metallic masks are applied they may be further used as buried electrical contacts to devices produced by ELO [30.11]. Finally, if the windows opened up in the insulating mask are so narrow that electrical conductance of ELO–substrate connections becomes small, the ELO process can be used for growing layers electrically separated from the substrate. This might be the way to obtain electrically insulated epilayers of semiconductors for which lattice-matched semi-insulating substrates are not available (e.g., GaSb).

To the best of our knowledge, the ELO process was initiated in 1980 when *McClelland* and his coworkers applied vapor-phase epitaxy (VPE) to grow GaAs epitaxial layers on masked-GaAs substrate [30.15]. Since the ELO layer was easily separable from its host substrate by cleaving, the technique was designed as an efficient way of producing GaAs epitaxial films on reusable substrates. *Tsaur* et al. were probably the first to report (in 1982) an efficient reduction of dislocation density in GaAs layers grown by VPE lateral overgrowth on Ge-coated Si substrates [30.16]. Then the ELO technique was used by *Jastrzebski*'s [30.9] and *Bauser*'s [30.17] groups to produce silicon-on-insulator structures by VPE and liquid-phase epitaxy (LPE), respectively. Finally, the efficient filtration of substrate dislocations has been reported in a series of papers on the growth of SiGe, GaAs, GaP, InP, InGaAs, and InGaP ELO layers on lattice-mismatched substrates [30.18–22]. However, the most spectacular recent achievement of the ELO technique was the breakthrough in the development of long-lifetime GaN/InGaN blue lasers, partly being due to the high efficiency of defects filtration during the lateral growth of GaN on sapphire [30.23].

Nowadays, a large body of ELO research concentrates on the lateral overgrowth of GaN epilayers on sapphire or SiC substrates using metalorganic VPE (MOVPE) or hydride VPE (HVPE) techniques. This is a result of the market demand for low-dislocation-density GaN substrates. GaN ELO activity is so dominant that sometimes the technique is called ELOG, which stands for epitaxial lateral overgrowth of gallium nitride. Apparently this is not correct since the principle of the ELO process is much more general and is not limited to either a specific group of materials or to a single epitaxial growth technique. In parallel, a great deal of increase in research on the lateral overgrowth of zincblende III–V epilayers on various substrates is observed.

The aim of this chapter is to present recent developments in the lateral overgrowth of compound semiconductors and review its present state. The chapter is organized as follows. In Sect. 30.2 we will focus on the mechanism of the ELO growth of zincblende III–V compounds from the liquid phase, highlighting the phenomena that are crucial for growing high-quality

layers with large aspect ratio. Epitaxy from the liquid phase has been chosen since, as will be discussed in that section, the equilibrium growth techniques such as LPE are the most suitable for lateral overgrowth. The issue of defect filtration in the ELO procedure is widely addressed in Sect. 30.3.1. Numerous examples which are instrumental in showing that the defect filtration mechanism schematically shown in Fig. 30.1 is very efficient and leading to significant progresses in the development of high-performance semiconductor devices made of lattice mismatched structures, are presented. In Sect. 30.3.2 structural perfection of seams that appear when layers grown from neighboring seeds merge in fully overgrown ELO structures will be discussed. Strain in epitaxial layers may lead to the generation of new defects if it is too large. Therefore, the issue of strain in ELO layers will be extensively addressed in Sect. 30.4. First, we will concentrate on strain arising due to interaction of ELO layers with the mask underneath. This phenomenon leads to tilting of ELO wings and is commonly found in various ELO structures. The origin of mask-induced wing tilting and possible ways of its control will be presented. Then, the thermal strain in lattice-mismatched ELO structures will be discussed (Sect. 30.4.2). We will show that in ELO structures thermal strain can be relaxed by additional tilting of ELO wings while still preserving their high quality. Finally, in Sect. 30.5 the recent progress made in the lateral overgrowth of semiconductor structures is presented. New developments in the ELO growth from the liquid phase include new mask materials and liquid-phase electroepitaxial growth on substrates coated by electrically conductive masks. Then, the new versions of ELO technique from solution and from the vapor phase, namely growth from ridges and pendeo-epitaxy, will be discussed and compared with the standard ELO. Since the author's intention was to present the ELO growth of a wide range of compound semiconductors, both the zincblende III–V compounds grown from the liquid solution as well as the vapor-grown GaN will be used to illustrate the phenomena discussed. In particular, in Sects. 30.3 and 30.4 it will be shown that a quite similar behavior is observed in various ELO systems despite the large differences in their properties and the variety of growth techniques applied. This, in turn, clearly indicates that the phenomena presented are not related to a specific group of compounds or their growth techniques, but have much more general nature.

30.2 Mechanism of Epitaxial Lateral Overgrowth from the Liquid Phase

An efficient ELO procedure requires a large growth rate anisotropy, i. e., growth conditions are necessary at which the lateral growth of the epilayer is much faster than that in the direction normal to the substrate. This can be achieved by taking advantage of the natural growth anisotropy of various crystal faces.

Figure 30.2 schematically shows growth rate versus interface supersaturation for three types of crystal faces. On a perfect singular face, atoms can be incorporated into solid in the form of two-dimensional nuclei only. If the face is singular but imperfect, the surface imperfections (e.g., dislocations) supply steps necessary for its growth. Due to the limited rate of the surface processes involved, larger surface supersaturation is required in these two growth modes to get a notable growth velocity. On the other hand, atoms can be added to an atomically rough crystal face in a random way and the growth rate of such a face varies linearly with the interface supersaturation. As discussed by *Nishinaga* [30.18] the basic idea of ELO lies in fundamental dissimilarities between those growth modes. If a slowly grown facet covers the upper plane of ELO while the sidewalls are rough, then for low supersaturation, growth rates in ver-

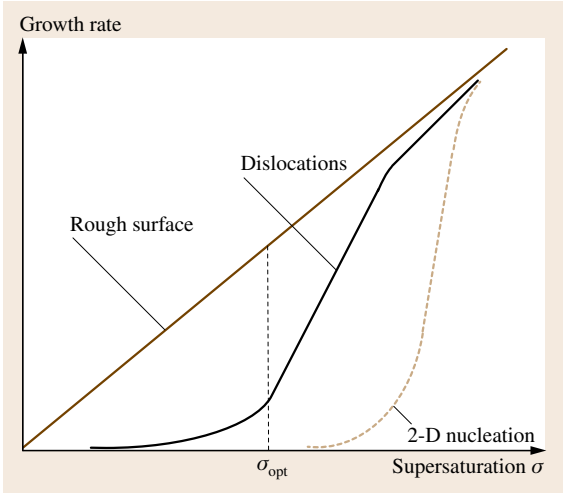


Fig. 30.2 Growth rate versus surface supersaturation for various crystal faces: perfect singular (*brown dashed line*), imperfect singular (*thin central line*), and atomically rough face (*solid line*). *Dashed line* schematically marks the surface supersaturation σ_{opt} that is optimal for large growth anisotropy in ELO

tical and horizontal directions differ significantly (see the vertical dashed line in Fig. 30.2 that schematically marks the optimal supersaturation σ_{opt}). Consequently, a large anisotropy of shape of ELO layers can be obtained.

Figure 30.2 indicates that there are two conditions necessary for successful lateral overgrowth. First, a proper geometry of the substrate must be chosen to keep the upper surface of the layer singular while the sidewalls of ELO remain atomically rough. Next, the growth conditions must be carefully adjusted for precise control of the surface supersaturation. These issues will be discussed in detail in Sects. 30.2.1 and 30.2.2. To illustrate that the mechanism of ELO growth presented above works in practice and to study the relative importance of various physical phenomena occurring during growth, we have conducted numerical simulations of ELO growth by LPE [30.24].

Figure 30.3 illustrates conceptually the solute movement in the liquid solution occurring during an ELO growth by LPE. Due to symmetry, only half of the substrate and liquid zone is shown. During LPE growth, the temperature of the system is slowly lowered to supersaturate the liquid solution. Since there is no nucleation on the mask, solute species diffuse exclusively towards the seeding area and are then incorporated into the growing ELO layer. Sidewall of the ELO layer is atomically rough, so there is no barrier there for incorporation of arriving species into the solid. Therefore, the solute concentration in the liquid zone near the side ELO face is equal to C_{eq}, the equilibrium concentration, a value that is determined by the phase diagram and actual temperature (Fig. 30.3). However, the upper ELO layer is faceted and the surface solute concentration there C_{in} is larger than the equilibrium concentration. This gives rise to a horizontal solute concentration gradient and so-called near-surface diffusion [30.25] in the liquid of solute species from upper ELO surface to its sidewall. It is obvious that the presence of near-surface diffusion significantly enhances lateral growth of ELO layers.

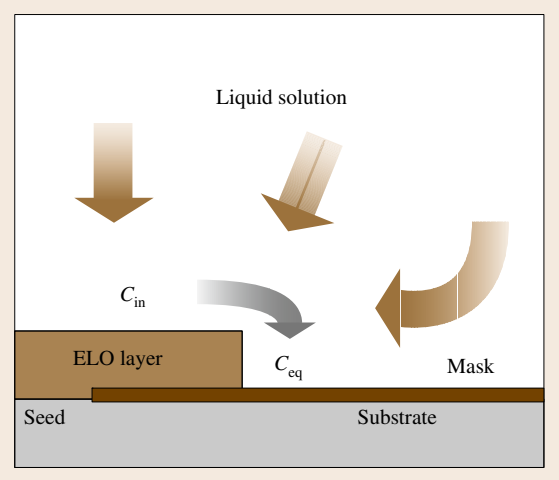


Fig. 30.3 Solute flow during LPE growth of ELO layer. The *arrow* from C_{in} to C_{eq} marks diffusion of solute in the liquid phase from upper surface of ELO to its sidewall (near-surface diffusion)

In the model [30.24], we consider a computational domain such as that shown in Fig. 30.3 and compute the change in the solute concentration field in the liquid zone as the system is cooled to force epitaxial growth on masked substrate. As discussed before, we assume a linear surface kinetic law to calculate the solute concentration C_{in} in the vicinity of the upper ELO surface while the solid–liquid phase equilibrium condition is used at the sidewall of the ELO layer. Moreover, the possible contribution of the Gibbs–Thomson effect is also considered to take into account the dependence of local equilibrium solute concentration on the surface curvature of the growing layer. Finally, the evolution of the solid–liquid interface shape with time is obtained using the growth velocity calculated from the solute concentration field in the vicinity of the growing layer. The results of calculations are discussed below, taking the GaAs/GaAs ELO system as an example due to its technological importance and the availability of the required input data.

Figure 30.4a shows the computed distribution of As concentration in the Ga/As solution after 2.5 h of ELO growth with an initial growth temperature of 650 °C and a cooling rate of 0.5 °C/min In the far field, isoconcentration contours are circular since at this length scale ELO crystal looks like a point located in the lower left corner of the computational domain. Close to the crystal, however, the As concentration distribution shows additional features. In particular, a solute concentration gradient appears along the upper ELO surface, leading to near-surface diffusion of As towards the ELO sidewall (compare Fig. 30.3) and enhancing lateral growth rate. Furthermore, the As concentration gradient close to sidewall of ELO is much larger than that at the upper face. This indicates that the layer grows faster laterally than in the vertical direction.

Figure 30.4b presents the cross section of GaAs ELO layer calculated for the same input parameters. As seen, the shape of the upper ELO surface is very flat, which closely agrees with experimental observa-

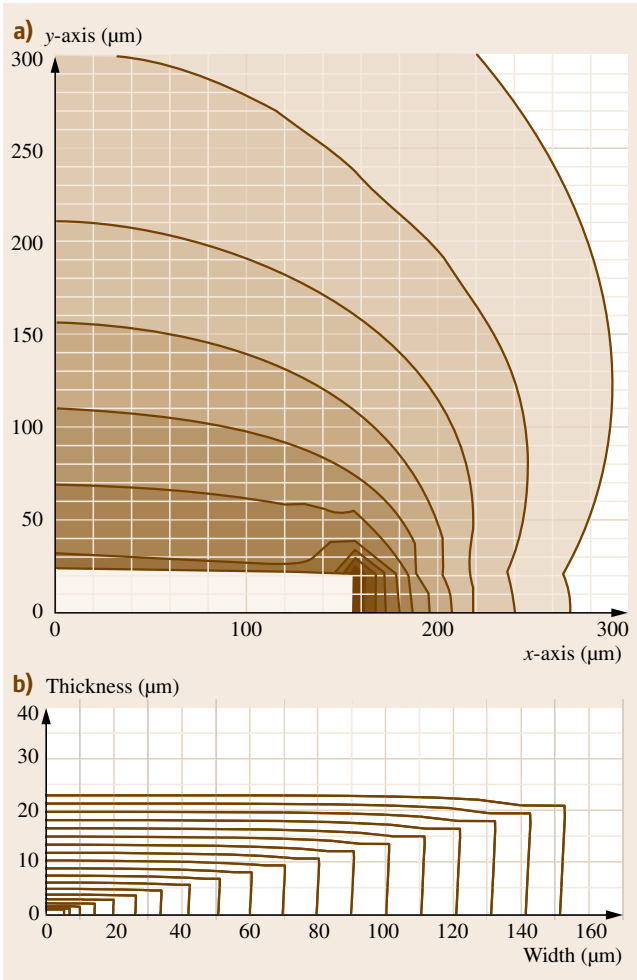


Fig. 30.4a,b Computed distribution of As concentration in the Ga-As solution **(a)** and cross section of GaAs ELO layer **(b)** after 2.5 h LPE growth with initial growth temperature of 650 °C and cooling rate of 0.5 °C/min

tions. Moreover, the lateral growth rate is much larger than that in the vertical direction, resulting in a thin and wide layer. A similar shape of layer was obtained by *Yan* et al. in their two-dimensional simulations of InP ELO growth by LPE [30.26]. However, in our case the prediction of a flat top face is a *result* of surface kinetics and diffusion processes included in the model, while in their approach the vertical growth rate is *assumed* to be the same along the whole upper surface. It is also noteworthy to mention that our model provides a good prediction for the dependence of ELO aspect ratio on surface kinetic coefficient and cooling rate [30.24]. Therefore, it can be successfully used to optimize the LPE growth procedure. In addition, we have shown that a spiky growth starts from the ELO corner, which is not observed in real experiments, if the dependence of solid–liquid equilibrium on surface curvature is not taken into account. This demonstrates the importance of Gibbs–Thomson effect in growth of ELO layers.

Having presented the mechanism of lateral overgrowth it is now worth to discuss which epitaxial growth technique is most suitable for ELO growth. As was mentioned earlier the surface supersaturation must be kept low to grow ELO layers with large aspect ratio. Otherwise, two-dimensional nucleation takes place on the upper ELO surface leading to higher vertical growth rate and consequently less growth anisotropy. Therefore the equilibrium growth techniques such as LPE should be chosen, if possible, for lateral overgrowth. Indeed, as will be shown later, ELO layers with aspect ratio as large as 130 can be grown from a liquid phase. Solution or melt growth of group III nitrides is extremely complicated due to low solubility of nitrogen in liquid metals (Sect. 30.5.2). Therefore, MOVPE or HVPE is commonly used nowadays to grow ELO structures of these compounds. Then, however, large supersaturation at the growing face makes control of growth anisotropy difficult, so lateral structures of GaN with aspect ratio up to 6 are usually obtained [30.27]. Molecular beam epitaxy (MBE) growth of ELO layers is even more complicated. The ELO technique requires that growth proceeds from the seeding areas only while deposition of polycrystalline material on the mask is hard to avoid during MBE [30.28, 29]. Sophisticated systems with molecular beams oriented at low angle to the substrate, low growth rate, and precise temperature control [30.30, 31] or special substrate preparation and growth procedures [30.32] are necessary to get the growth started selectively from the seeds. Even then, however, MBE-grown ELO layers are usually very narrow, which makes their application in devices production very difficult.

30.2.1 Choice of Substrate Geometry for Growth of ELO Layers

As noted above, large anisotropy of ELO growth rate can be obtained if the upper surface of the layer is facetted while the sidewalls of ELO remain atomically rough. In practice, for zincblende III–V semiconductors this can be achieved using (111)- or (100)-oriented substrate and by twisting the line openings in the mask by

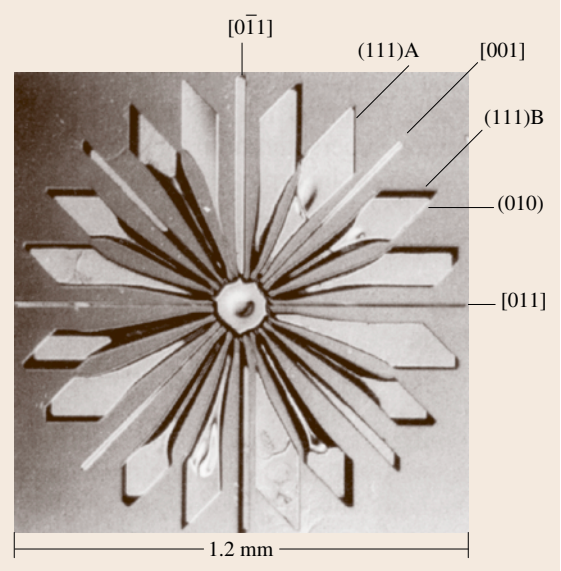


Fig. 30.5 Microphotograph of a GaAs ELO layer grown by LPE on a (100) GaAs substrate with a star-like pattern of seeds cut in the SiO_2 mask

some degrees off from the main crystallographic directions. Usually, optimal orientations of seeding lines in the substrate plane are determined experimentally by growing an ELO layer on a masked substrate with a star-like pattern of the seeds and by studying lateral growth rate versus seed orientation.

Figure 30.5 shows a plane view of a GaAs epilayer grown by LPE on a (100) GaAs substrate with a star-like pattern of openings cut in the SiO_2 mask [30.33]. The growth was forced by cooling the system from the temperature of 750 °C with a rate of 0.2 °C/min for 50 min. From Fig. 30.5 the dependence of lateral growth rate on the orientation of the seeds can be easily found. The width of the ELO is the largest when the stripe is ±15 and ±30° off-oriented from the [011] direction or its equivalent orientations, while it is the smallest for the [011] and [001] directions. From simple geometrical considerations the facets on the side walls of ELO layers can be identified as the slowly grown {111} and {100} planes. Their formation proceeded much faster when the stripe was oriented in the ⟨011⟩ and ⟨001⟩ directions, which slowed down the lateral development of the ELO layer. Therefore, on the (100) substrate the line seeds oriented 15, 30, 60 or 75° off from the ⟨011⟩ direction should be used to obtain a large value of the lateral-to-normal growth rate ratio. Note that the angular dependence of the lateral growth rate is centrosymmetric. As will be discussed later this is no longer true if slightly misoriented substrate is used. Note also that the width of ELO stripes shown in Fig. 30.5 increases with the distance r from the star center. As shown by *Zhang* and *Nishinaga* [30.34] this is due to the fact that the effective sink area for the growth units, and thus also the surface supersaturation, varies with r. For lateral overgrowth of GaSb on (100)-oriented GaSb substrates by LPE we have found the same dependence of ELO lateral growth rate on seed orientation. Results shown in Fig. 30.5 are in agreement with those published by *Zhang* and *Nishinaga* [30.34] and by *Naritsuka* and *Nishinaga* [30.35] for liquid-phase homoepitaxial lateral overgrowth of GaAs and InP, respectively. Similar angular dependence of lateral overgrowth has also been reported for ELO growth of GaAs by MOVPE [30.36] and for InP by VPE [30.37]. For (111)-oriented substrates analogous dependences can be found in [30.38, 39] for LPE ELO growth of GaAs and GaP, respectively.

In the case discussed thus far, the vertical growth of ELO was possible due to surface steps supplied by dislocations present in the substrates. However, sometimes dislocation-free substrates are used, for example, in silicon homoepitaxy. They must be slightly misoriented to provide steps necessary for crystal growth if uncontrolled two-dimensional nucleation is to be avoided. Then, the size and shape of ELO layer becomes dependent on the angle between seeding line and the substrate miscut directions. This issue has been discussed in detail by *Bergmann* et al. [30.40] and then by *Bergmann* [30.41] for lateral overgrowth of silicon on perfect Si substrates. On off-oriented imperfect substrates additional restriction applies for optimal direction of seed orientation. *Sakawa* and *Nishinaga* [30.42] have shown that, in such case, from many equivalent seed directions that are optimal on the (100) plane (Fig. 30.5), the one that should be chosen is that for which the density of misorientation steps inside the seeding area is the smallest.

The discussion above indicates that the presence of dislocations significantly affects the ELO growth mode. To present this issue in more detail Fig. 30.6 shows a sketch of the ELO layer grown from a line window on a misoriented, dislocation-free substrate. Direction of seeding line perpendicular to substrate miscut direction is assumed. As long as two-dimensional nucleation does not take place, the layer grows by flow of steps according to the misorientation direction (i. e., to the right side). Then the ELO growth is self-limited, i. e., it stops as soon as all the steps reach the edge of the layer

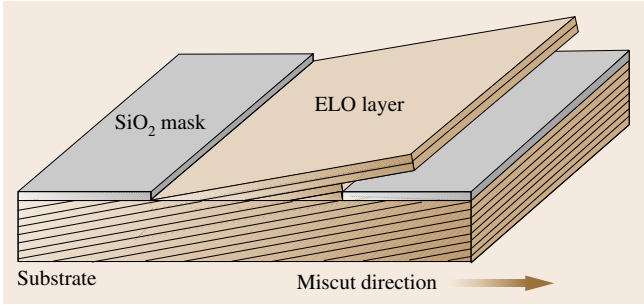


Fig. 30.6 Schematic illustration of ELO growth on misoriented, dislocation-free substrates; the seeding line is set perpendicular to the substrate miscut direction

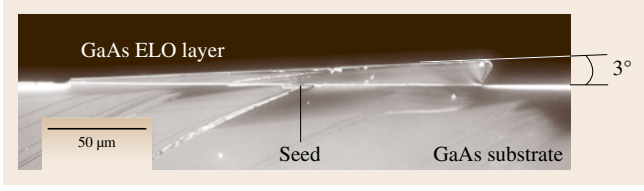


Fig. 30.7 Cross section of a GaAs ELO layer grown on a GaAs substrate with the surface misoriented by 3° from the (100) plane; substrate is miscut in the right-hand side direction



Fig. 30.8 Scanning electron microscopy image of Si ELO layers grown by LPE on a SiO_2-masked Si substrate. Perfectly grown, defect-free layers are visible in the *foreground*; irregular-shaped defective layers are seen in the *background* (courtesy of Banhart)

and the sidewall is covered by a slowly grown facet. In particular, this means that no growth should take place to the left of the seeding line. Indeed, such behavior has been observed during LPE growth of silicon ELO layers on perfect silicon substrates [30.41]. Note also in Fig. 30.6 that, for small substrate misorientation angles, substrate steps are partly in the shade of the mask edge and the ELO layer cannot get out of the seeding area if the mask is too thick. This prediction is also in agreement with experimental observations [30.40].

On the other hand, the situation is quite different if substrate dislocations contribute to the ELO growth. Figure 30.7 shows a cross section of a GaAs ELO layer grown on a GaAs substrate with the surface miscut by 3° from the (100) plane [30.43]. The seeds were oriented 15° off from the $[0\bar{1}1]$ direction for fast lateral overgrowth, and were nearly perpendicular to the direction of substrate miscut. It is seen that, similar to the case shown in Fig. 30.6, the thickness of the ELO layer is not uniform. The reason for this is the same as before: substrate steps lead to the layer growth to the right-hand side. This time, however, growth to the left of the window (i. e., in the direction opposite to substrate miscut axis) is also observed. Such growth can be explained by considering dislocations (with a density of $1.5 \times 10^3\,\mathrm{cm}^{-2}$) present in our GaAs substrate. These dislocations provided steps that made an additional contribution to epitaxial growth and led to continuous growth of the layer to both sides of the seeding line. It is noteworthy that the surface of the layer is inclined to the substrate surface by an angle of $\approx 3°$ (Fig. 30.7). Simple geometrical considerations of sample geometry show that this is the angle at which the (100) plane intersects with the substrate plane in the $(1\bar{1}0)$ cleavage section. This indicates that, despite substrate misorientation, the upper surface of the ELO layer forms the exact (100) plane, which is in agreement with our explanation.

As the final example of the essential role played by dislocations in the growth of ELO layers, Fig. 30.8 shows a scanning electron microscopy (SEM) image of silicon ELO layers grown by LPE on SiO_2-masked Si substrate [30.44]. Perfectly grown flat ELO layers of thickness around 2–3 µm are seen in the foreground. Studies by transmission electron microscopy (TEM) showed that they were entirely free of crystallographic defects. However, as shown in the background of the figure, some ELO layers of thickness exceeding 20 µm and quite different shape were also found on the same substrate wafer. They showed deep grooves and were about twice as wide as perfectly grown layers. TEM analysis revealed that those layers contained a regular arrangement of dislocations generated during ELO growth in the vicinity of SiO_2 mask edge. The origin of these dislocations is not fully clarified but it is believed that their formation is influenced by stresses caused

by thermal oxidation of substrate wafer, the shape of the oxide edge, and nonuniform supersaturation of liquid solution volume. Once dislocations are generated they govern further development of the growth process. Instead of flow of steps supplied by substrate misorientation as illustrated in Fig. 30.6, dislocations become the dominant growth step source that leads to significant enhancement of vertical growth rate and worse surface morphology.

30.2.2 Optimization of Liquid-Phase Lateral Overgrowth Procedure

Having presented the effect of substrate geometry on the ELO growth process, in this section we will show examples of how parameters of an LPE procedure should be adjusted to optimize supersaturation near the crystal faces and obtain high growth anisotropy. In LPE, surface supersaturation reflects the relative magnitude of the solute supply from bulk of the liquid phase and solute consumption at the surface of growing crystal. Thus, the main parameters controlling these processes are the LPE growth temperature, the initial supercooling of the liquid phase, and the cooling rate.

Figure 30.9 shows lateral $V_{lat} = L/t$ and vertical $V_{ver} = h/t$ growth rates versus LPE growth temperature T_0 for nominally undoped GaAs ELO layers grown on SiO_2-masked GaAs substrates. Figure 30.10 presents values of the aspect ratio (AR) for the same layers [30.43]. As seen, for high LPE temperature the vertical ELO growth rate is large since some thermal roughening of the upper crystal face takes place. Moreover, surface kinetic processes are very fast, which leads to an additional increase in the vertical growth rate. This results in ELO layers with a small value of the aspect ratio. On the other hand, for very low growth temperatures the sidewalls cannot be considered as ideally rough. Most probably surface kinetic processes start to play a role there and lateral growth rate decreases, which leads again to a decrease of the aspect ratio. Thus, there is an optimum growth temperature T_{opt} at which ELO layers have the largest aspect ratio. As shown by *Yan* et al. [30.25] for InP ELO layers T_{opt} corresponds to the temperature at which the interface supersaturation at the upper ELO face is the smallest. For other growth temperatures surface supersaturation is larger than σ_{opt} (compare Fig. 30.2). It is obvious that the optimum ELO growth temperature depends on many parameters such as the slope of the liquidus curve of the phase diagram, the geometry of the LPE system, etc., so it must be determined experimentally for each particular case under study. Values of T_{opt} equal to 500, 580, and 530 °C have been reported for LPE growth of InP/InP [30.25], GaAs/GaAs [30.43], and GaAs/Si [30.45] ELO systems, respectively. For homoepitaxial GaSb ELO layers grown by LPE the optimal growth temperature $T_{opt} \leq 350$ °C has been found [30.46].

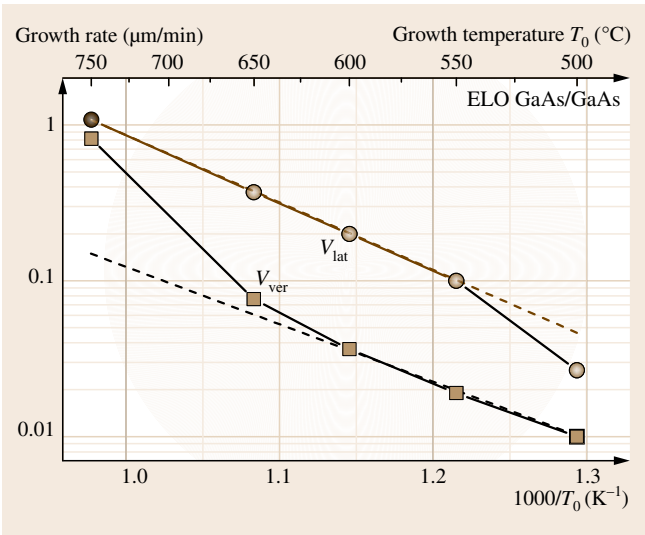


Fig. 30.9 Lateral V_{lat} (*dots*) and vertical V_{ver} (*squares*) growth rates versus LPE growth temperature T_0 for undoped GaAs ELO layers grown on GaAs substrates masked by SiO_2 film; seeding windows and masked area were 10 and 100 μm wide, respectively. The *dashed lines* guide the eye along the $V \propto \exp(-1/T_0)$ dependence

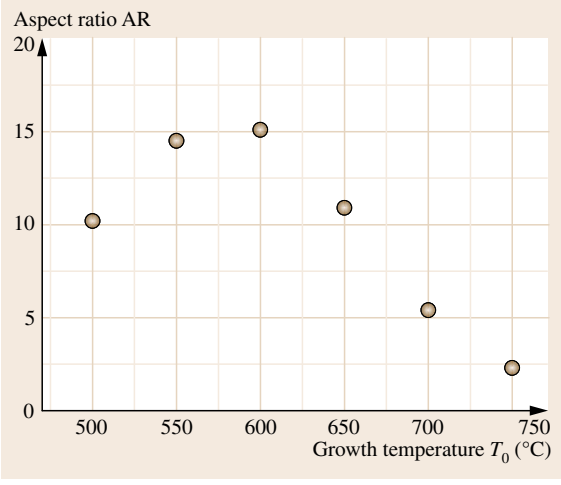


Fig. 30.10 Aspect ratio (AR) of undoped GaAs ELO layers versus LPE growth temperature T_0

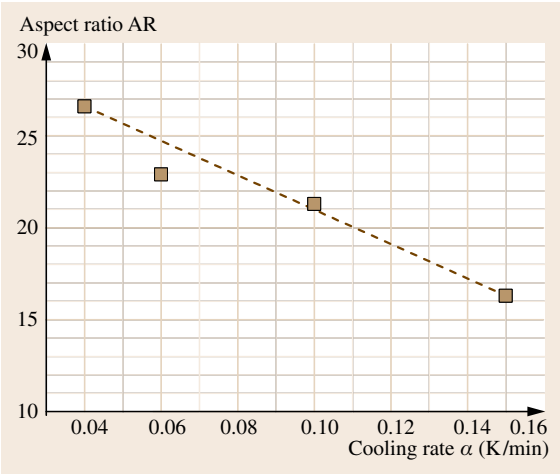


Fig. 30.11 Aspect ratio of Si-doped GaAs ELO layers grown at $T_0 = 575\,°C$ versus cooling rate; silicon concentration in the Ga-As solution was [Si] = 0.5 at. %

To keep surface supersaturation at a low value, the cooling rate during LPE should be as low as possible. This is illustrated in Fig. 30.11, which shows the aspect ratio of Si-doped GaAs ELO layers versus cooling rate α [30.43]. The main idea behind that is to supply solute to the surface of the growing layer slowly enough that it can be transported by near-surface diffusion from upper to side walls of ELO and be incorporated there without any significant increase in surface supersaturation at the upper face. Similar dependence of aspect ratio on cooling rate has been found for InP [30.26] and GaSb [30.46] ELO systems. Also numerical simulations of ELO growth by LPE predict AR versus α dependence similar to that shown in Fig. 30.11 [30.24, 26].

Optimal choice of initial solution supercooling during ELO growth by LPE requires some additional effects to be taken into account. It is apparent that supercooling of the solution should be as small as possible. Then, the initial supersaturation of bulk of the liquid, and consequently that at the ELO surface, can be kept low. Indeed, we have found experimentally that, for the LPE growth of GaSb ELO structures, smaller values of initial melt supercooling lead to larger values of aspect ratio [30.46]. There are ELO systems, however, which require some initial supercooling of the liquid solution to initiate growth. Otherwise, the Gibbs–Thomson effect hinders the layer from growing out from the opening in the mask. In particular, this is the case for LPE growth of Si ELO layers on defect-free silicon substrates. Then the ELO layer grows only by flow of steps supplied by substrate miscut as shown in Fig. 30.6. Sidewall of the layer beginning to grow laterally is strongly curved, and due to the Gibbs–Thomson effect, it requires higher equilibrium solute concentration than the planar face. Thus, instead of growing laterally the layer is dissolved and cannot get out of the opening in the mask as long as the liquid phase is not supersaturated sufficiently [30.47].

The situation is different if dislocations enhancing vertical ELO growth are present in the substrate. To illustrate this point Fig. 30.12a shows a cross section of silicon-doped GaAs ELO layer grown on GaAs substrate by LPE without any initial supersaturation of the liquid solution [30.48]. As we have already reported [30.49], by proper chemical etching, boundaries between vertically and laterally grown parts of the layer can be revealed, allowing analysis of the temporal development of ELO shape. These boundaries are sketched in Fig. 30.12b by dashed black lines, while dashed brown lines schematically mark their shape that would be observed without the contribution of the Gibbs–Thomson effect. Note that the Gibbs–Thomson

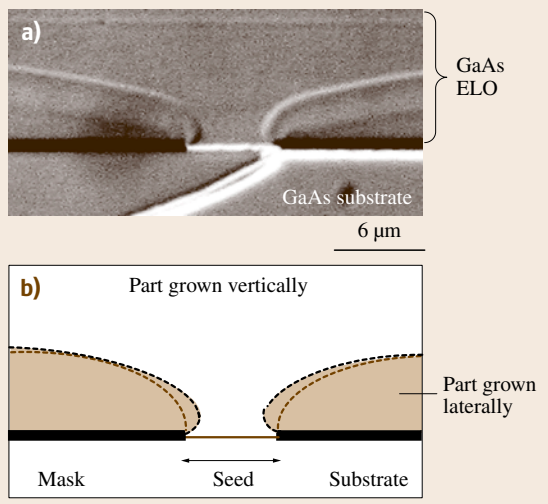


Fig. 30.12a,b Microphotograph **(a)** and schematic drawing **(b)** of the cross section of Si-doped GaAs ELO layer grown on GaAs substrate by LPE without initial supersaturation of the liquid solution. Boundaries between vertically and laterally grown parts of the layer are marked by *black dashed line* while their shape that would be observed without contribution of the Gibbs–Thomson effect are drawn by the *brown dashed lines* in **(b)**. Note dissolution in the lateral direction induced by the Gibbs–Thomson effect at the beginning of growth

effect induces *dissolution* in the lateral direction instead of *growth* at the beginning of epitaxy. This is clearly visible as initial narrowing of the vertically grown part of ELO. However, during continuous cooling of the system, steps supplied by substrate dislocations still allow for vertical growth of the layer in the middle of the seed despite slow dissolution induced by the Gibbs–Thomson effect taking place in the direction parallel to the substrate. As soon as the layer grows thicker, curvature of the solid–liquid interface at the ELO sidewall decreases. Consequently, the contribution of the Gibbs–Thomson effect decreases and lateral overgrowth along the mask begins. We have proved that this explanation is correct by showing that the magnitude of lateral dissolution caused by the Gibbs–Thomson effect can be reduced if slightly supersaturated liquid solution is used to grow the layer [30.33]. This example again illustrates the important role of dislocations in the growth of ELO layers.

Doping has been found to be a useful way to reduce the vertical growth rate, leading to thin and wide ELO layers. Figure 30.13a shows an image of a GaAs ELO layer grown from a nominally undoped solution by LPE on SiO_2-masked GaAs substrate at 750 °C [30.49]. Since the growth temperature was much higher than the optimal temperature, the cross section of the layer has

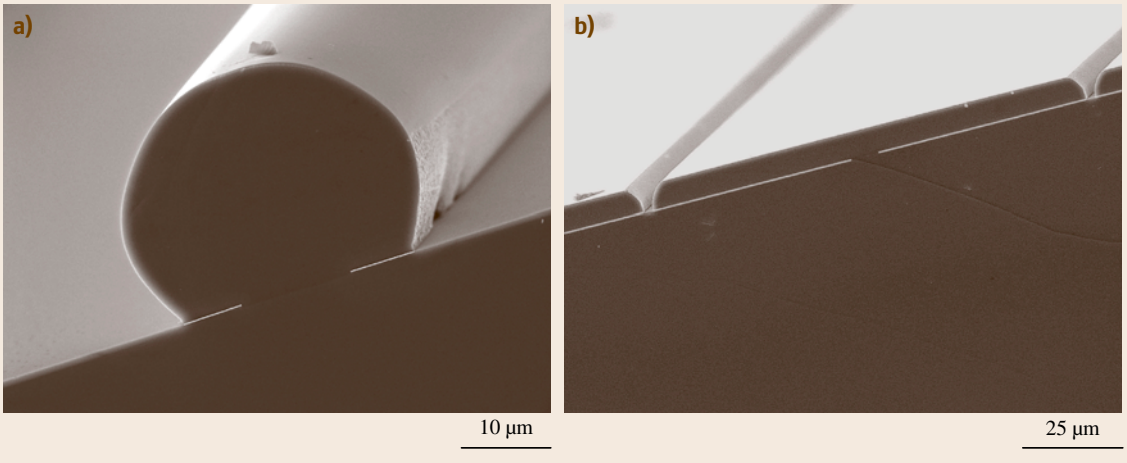


Fig. 30.13a,b Scanning electron microscopy images of undoped **(a)** and Si-doped **(b)** GaAs ELO layers grown at 750 °C on SiO_2-masked GaAs substrates. Besides doping, all the other growth parameters were the same for both layers

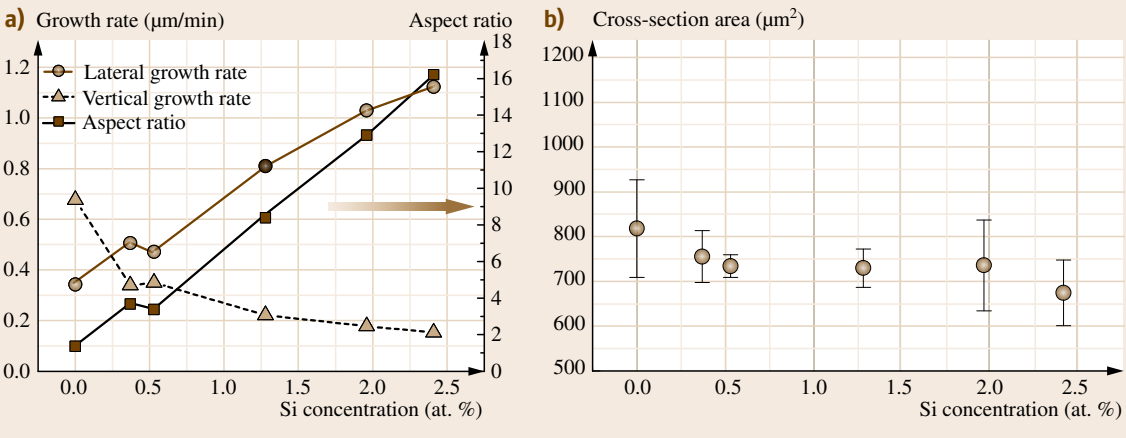


Fig. 30.14 **(a)** Lateral and vertical growth rates, and aspect ratio of GaAs ELO layers grown on GaAs substrates at 750 °C versus Si concentration [Si] added to the Ga-As liquid solution; **(b)** the cross sectional area of the same layers versus [Si]

a circular shape and the aspect ration is small (compare Fig. 30.10). A noticeable improvement of ELO shape has been achieved by adding 2.4 at. % Si to the Ga/As solution (Fig. 30.13b). As a result, the upper surface of the layer becomes very flat and the aspect ratio increases.

In order to study the effect in more detail, Fig. 30.14a shows a plot of lateral and vertical growth rates, and of the ELO aspect ratio of Si-doped GaAs layers versus the silicon concentration in the liquid phase. As seen, the vertical growth rate continuously decreases while the lateral one increases for higher dopant concentration in the liquid solution. As a result the ELO aspect ratio increases significantly with the doping level. It is important to emphasize that, whereas ELO layers become thinner and wider owing to the Si doping, within the limits of experimental error their cross sectional areas, being a measure of the amount of material deposited on the substrate, do not change (Fig. 30.14b). This means that the presence of Si has a negligible influence on the phase equilibrium, and a possible change of the liquid-zone supersaturation with doping cannot explain the effect presented. Thus, other processes affecting the redistribution of the growth units on the ELO surfaces should be considered.

The phenomenon of crystal habit modification by impurities is well known [30.50]. It might be attributed to hindering of the smooth step flow on the crystal face [30.51,52]. The possible mechanism involves steps that are forced to squeeze between immobile impurity atoms tightly adsorbed on the surface. Thus, the steps have a local large curvature, which results in a decrease of their flow velocity. When applied to ELO growth this model leads us to the following interpretation of the data shown in Fig. 30.14: silicon atoms, when introduced to the liquid solution, are adsorbed on the upper ELO surface and render the steps there less operative, reducing in this way the vertical growth rate. On the other hand, the doping should have no direct influence on the random incorporation of growth units on atomically rough ELO faces. Since solute atoms cannot be easily incorporated onto the upper surface they are efficiently transported to the sidewalls by the near-surface diffusion. In that way the lateral growth rate increases at the expense of vertical one, while the total amount of material deposited on the substrate does not change.

It is worth mentioning that we have found a similar effect for Sn and Te dopants, although the reduction in the GaAs vertical growth rate caused by these impurities was much smaller than that for Si [30.49]. The reason is still unclear, but different values of the surface mobility of Si, Sn, and Te might be responsible. We have also observed the same increase of ELO aspect ratio by Si dopant during LPE growth of Si-doped GaSb ELO layers [30.46]. Change of ELO shape with doping similar to that shown in Fig. 30.13 has been observed during ELO growth of GaAs on GaAs-coated Si substrates by LPE [30.8]. *Kim* and *Lee* reported that selenium is even more efficient than silicon in reducing vertical growth rate of ELO GaAs [30.53], but our experiments do not confirm their conclusion. It is important to point out that, in principle, similar phenomenon should be expected in other epitaxial systems in which the upper ELO face grows by the flow of surface steps. Indeed, a strong influence of magnesium doping on the growth habit of the GaN ELO layers grown by MOVPE on sapphire was found recently [30.54]. The effect of Si doping was negligible in that case. This shows that the phenomenon of vertical growth rate reduction by doping is not a specific attribute of ELO layers grown by LPE and has much more general nature, although its magnitude depends on growth conditions and the most effective impurity must be found for each particular system under study. Basic research is still needed to obtain a deeper insight into the effect of impurities on the growing crystal and to indicate which best suits the growth of thin and wide ELO layers.

Finally it should be noted that in some systems the value of the ELO aspect ratio is much more sensitive to the presence of dopants in the liquid solution than to the LPE growth temperature. One may try to increase the aspect ratio by growing ELO layers at a relatively low optimal temperature T_{opt} as discussed earlier. However, the dopant solubility at this temperature might be too low to increase further the aspect ratio. Alternatively, some increase in the growth temperature over T_{opt} with a simultaneous increase of doping might be a reasonable way to grow thin and wide ELO layers. This procedure can give satisfactory results only if the aspect ratio increase due to higher doping is larger that its decrease caused by higher LPE growth temperature. We have found that this is the case for Si-doped GaSb ELO layers [30.46]. However, care is advised when trying this approach since heavy doping may lead to a worse surface morphology of the layers.

To summarize this section, Fig. 30.15 presents a scanning electron microscopy cross section image of a GaAs ELO layer grown in our laboratory by LPE on SiO_2-masked GaAs substrate. The layer was doped with silicon and the LPE growth conditions were carefully adjusted according to the rules discussed above to get

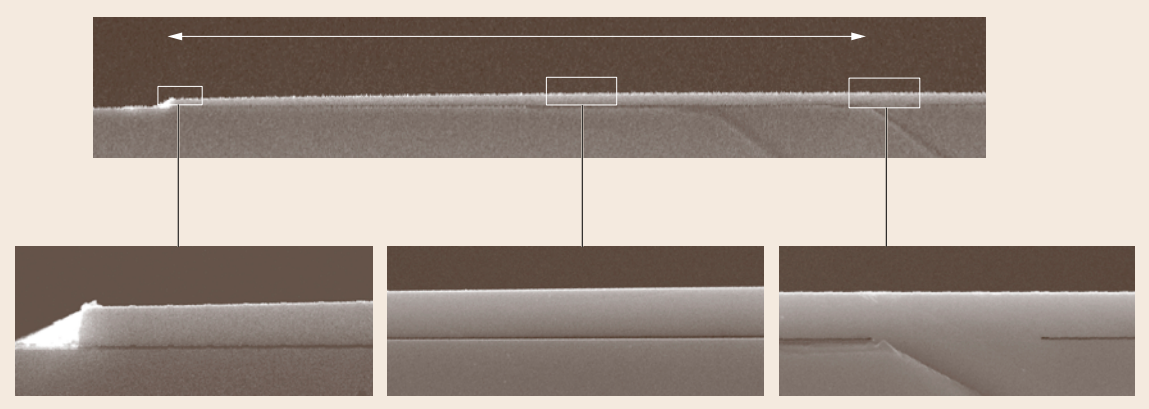


Fig. 30.15 Cross-section of Si-doped GaAs ELO layer grown by LPE on SiO_2-masked GaAs substrate. The thickness of the layer in its central part is 2.8 μm. The wing (marked by the *white arrow*) is 172 μm wide, yielding an aspect ratio of 126

the largest value of aspect ratio. The layer is 2.8 μm thick while its wings are 172 μm wide, which gives an aspect ratio as large as 126. Note that the thickness of the layer continuously decreases from the center to the edge. This is evidence of bending of the layer towards the mask as will be discussed in Sect. 30.4.1.

30.3 Dislocations in ELO Layers

As discussed in Sect. 30.1 the unique advantage of growing epitaxial layers by the ELO technique is that substrate dislocations are efficiently filtered during the growth, so low-dislocation-density epitaxial layers can be obtained on heavily dislocated substrates. In the following subsections this issue will be addressed in more detail. First, we will focus on the distribution of dislocations in lattice-mismatched ELO layers, showing examples where the process illustrated schematically in Fig. 30.1 is really observed in laboratory practice. In Sect. 30.3.2, the structural perfection of seams that appear when layers grown from neighboring seeds merge in fully overgrown ELO structure will be discussed.

30.3.1 Filtration of Substrate Dislocations in ELO

Figure 30.16a shows a cross section of a GaAs ELO layer grown by LPE on a silicon substrate. The structure consists of Si(100) substrate with 2 μm-thick GaAs buffer grown by MBE and coated with the SiO_2 masking film. Width and thickness of the layer are 85 and 11 μm, respectively. More growth details can be found elsewhere [30.55]. Figure 30.16b shows a plane view of the same layer with etch pits revealed by etching in molten KOH. It is noteworthy that the density of etch pits on the buffer surface is very high ($\approx 10^8$ cm^{-2}). On the contrary, wings of the layer are nearly dislocation free, and only the dislocations threading through the seeding window from the buffer are seen on the upper ELO surface. In agreement with the model presented in Fig. 30.1 these dislocations are confined in a narrow area above the seed. The same behavior is seen in Fig. 30.17, which presents a cross section (Fig. 30.17a) and plane view (Fig. 30.17b) of GaSb ELO layer grown by LPE on GaAs substrate coated by MBE-grown GaSb buffer [30.43]. These examples show that substrate defect filtration during ELO procedure is very efficient, and low-dislocation-density epilayers can be obtained by this technique despite high dislocation density in the seeding area.

Figure 30.18a presents a cross sectional TEM image of a GaAs ELO layer grown by LPE on a GaAs-coated Si substrate [30.56]. Note the large density of dislocations at the bottom of the figure, i. e., close to the GaAs/Si interface. Density of these defects decreases with thickness of the buffer and finally all of them are blocked by the SiO_2 film. Thus, no dislocation is

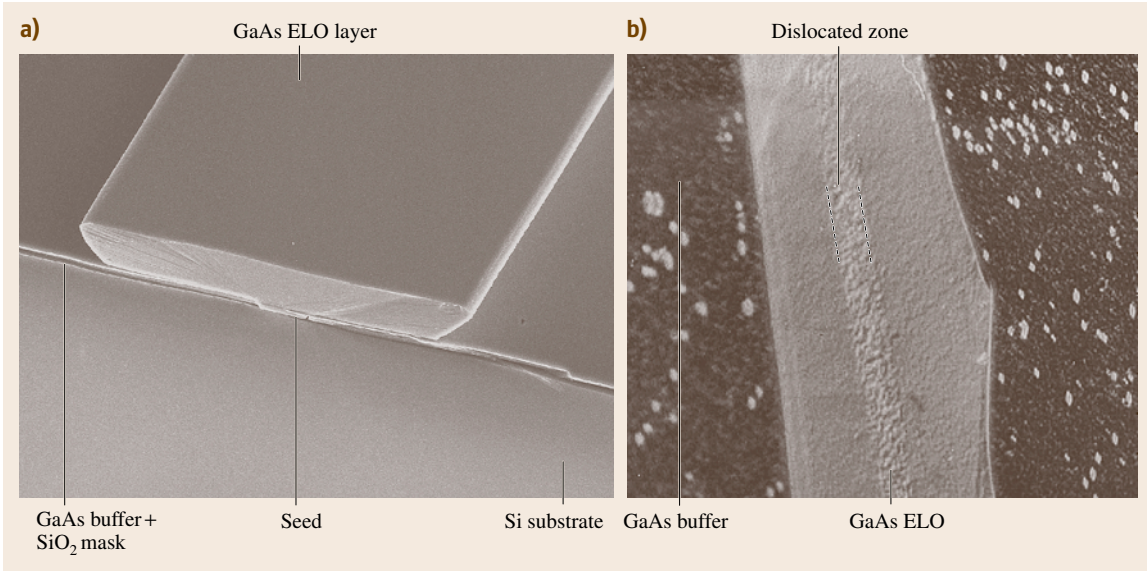


Fig. 30.16a,b SEM image of LPE-grown GaAs on Si ELO structure **(a)** and its plane view after etching in KOH to reveal the etch pit distribution **(b)**. The layer is 85 μm wide

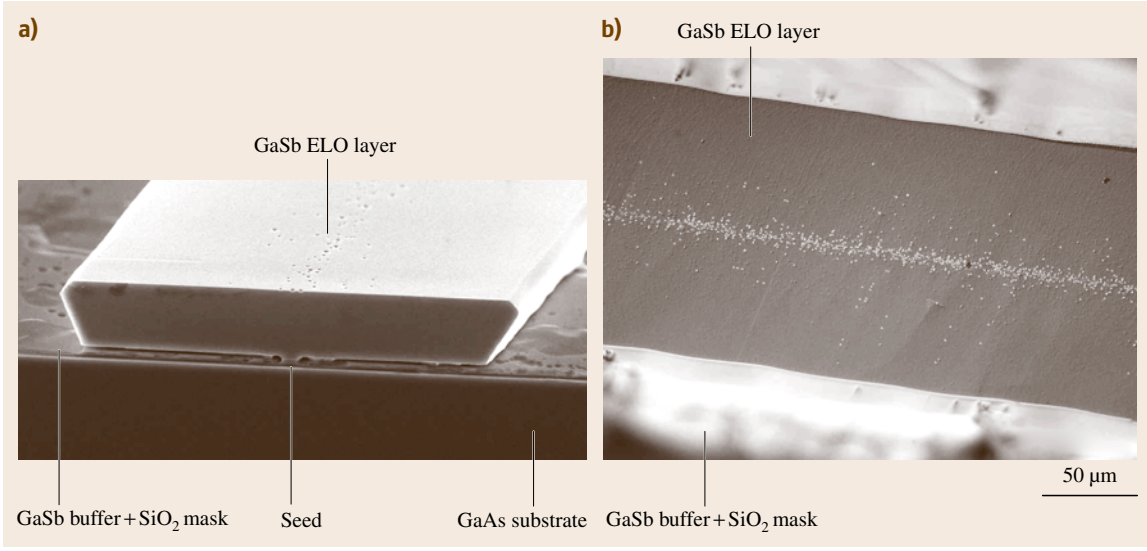


Fig. 30.17a,b Cross-section **(a)** and plane view **(b)** of a GaSb ELO layer grown by LPE on a GaAs substrate coated by 2 μm-thick GaSb buffer and SiO_2 mask; in **(b)** etch pits are revealed by chemical etching

seen in the GaAs wing area above the mask (compare Fig. 30.18b), which agrees with results shown in Figs. 30.16 and 30.17. Dislocations propagate to the ELO layer only through the opening in the mask. They are aligned on {111} planes, so the width of the defective area on the layer surface increases with the thickness of the layer. This is exactly the behavior postulated in Sect. 30.1 when explaining the ELO concept with the use of Fig. 30.1.

As the last example of an efficient filtration of substrate dislocations in ELO, Fig. 30.19 shows the distribution of dislocations analyzed by cross sectional

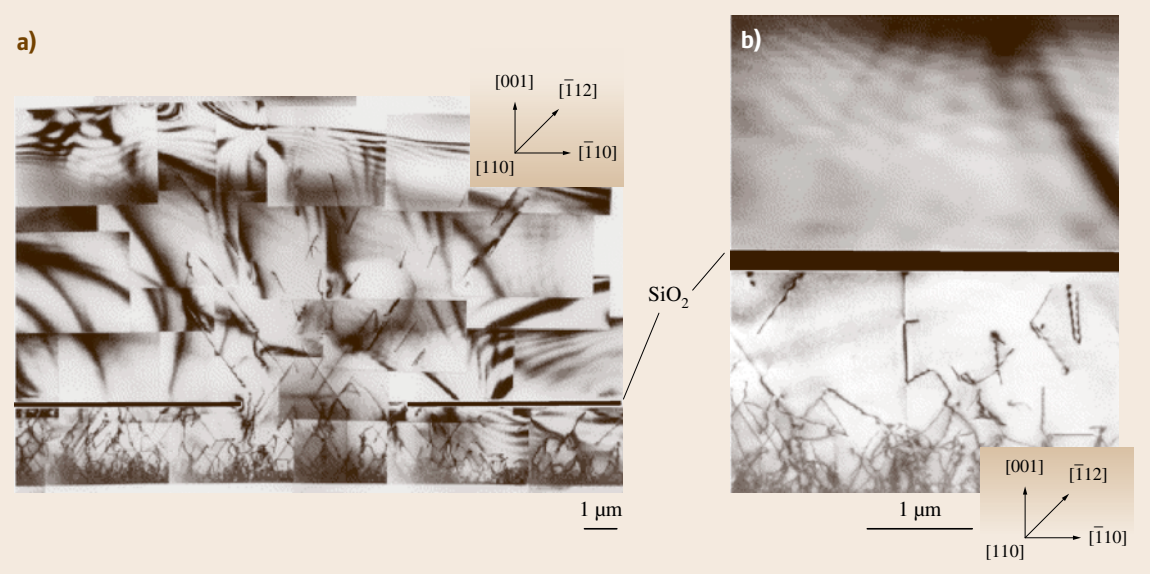


Fig. 30.18 (a) TEM cross sectional image of a GaAs ELO layer grown by LPE on a GaAs-coated Si substrate; (b) larger magnification of the ELO GaAs/SiO_2/GaAs buffer area; the mask position is marked by *solid black lines* for better visibility (courtesy of Tamura)

TEM in 140 µm-thick GaN ELO layers grown by HVPE on sapphire substrates [30.57]. The efficient blocking of buffer dislocations by the 1 µm-wide mask is clearly visible. Note also that some of the dislocations threading to the layer through the opening in the SiO_2 mask change their direction close to the mask edge and propagate parallel to the mask surface without reaching the upper ELO face. *Sakai* et al. claim that this phenomenon, instead of mechanical blocking of dislocations by the mask, gives the main contribution to the reduction of threading dislocation density in HVPE GaN films [30.58]. The mechanism of bending of threading dislocations in GaN ELO layers is not yet fully understood, although there are suggestions that it may be closely related to the change of the growth front direction and appearance of crystal facets at the very beginning of lateral overgrowth [30.58]. Probably, stress present during growth of the layers may play a role as well. It is worth mentioning that, to the best of our knowledge, there are no reports on similar behavior of dislocations in ELO structures of zincblende III–V materials. Actually, no evidence of dislocation bending in GaAs ELO layer grown on Si substrate is seen in Fig. 30.18. Some of the dislocations threading from the GaN buffer propagate through the GaN ELO layer without any change of direction, creating a defect zone on the surface above the seed, similar to those shown in Figs. 30.16b and 30.17b. In GaN these dislocations are aligned parallel to the *c*-axis, i. e., perpendicular to the substrate surface. Consequently, the width of the de-

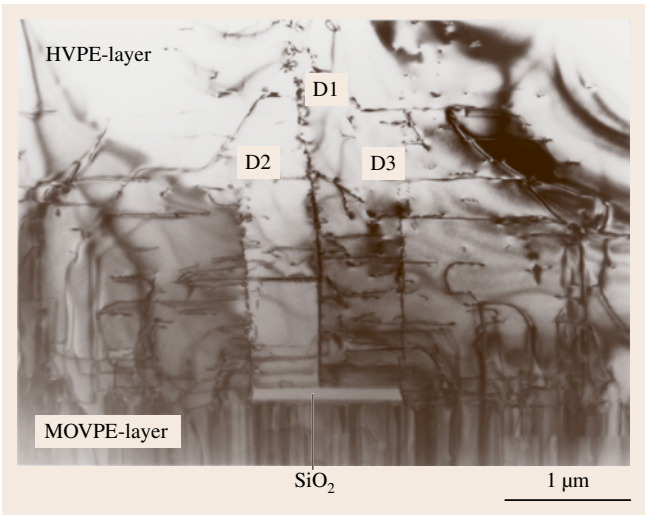


Fig. 30.19 TEM cross sectional image of a GaN ELO layer grown by HVPE on a GaN-coated sapphire substrate. Note that some dislocations threading from the buffer bend in the vicinity of the SiO_2 mask edge and propagate along the mask surface to ELO side-wall (courtesy of Sakai)

fect zone on the upper ELO surface does not change significantly with layer thickness. Again, this behavior is quite different from that observed in ELO structures of zincblende compounds (compare Fig. 30.18a). Note also that two additional types of defects, denoted by D1 and D2, aligned along the [0001] direction, are visible in Fig. 30.19: the D1 defect originates approximately from the center of the SiO_2 mask while the D2 defect originates from both edges of the mask. Their origin will be discussed in the following sections.

Finally, let us mention that high crystallographic perfection of ELO wings results in their better optical and electrical properties than those of planar buffers. In particular, we have shown that cathodoluminescence intensity is much higher in the wing area of the GaAs on Si ELO structure than that from the material grown vertically above the seed, and even higher than that from the MBE-grown GaAs buffer [30.20]. This should be expected, as dislocations are known to behave like centers of effective nonradiative recombination of excited carriers. A similar distribution of cathodoluminescence intensity has also been observed across surfaces of GaN on sapphire [30.59] and SiGe on Si [30.60] ELO structures. The advantage of using ELO layers for devices was first demonstrated by *Nakamura* et al., who reported in 1997 continuous-wave (CW) room-temperature operation of a blue, InGaN/GaN laser diode with a lifetime longer than 1000 h when deposited on GaN/sapphire ELO substrate [30.23]. Since then the lifetime of these diodes has been significantly increased, so they are commercially available at present. This remarkable improvement was due to a significant reduction of the threshold current density of the laser diodes fabricated on low-dislocation-density GaN ELO substrates. Similar behavior has been reported for GaAs-based lasers grown on GaAs/Si ELO substrate [30.61]. *Kozodoy* et al. have shown that the use of lateral overgrowth to eliminate dislocations leads to better electrical properties of GaN p-n junctions [30.62]. Reverse-bias leakage current was reduced by three orders of magnitude for diodes located on low-dislocation-density ELO wings. This has led to improved performance of light-emitting diodes [30.63,64]. Moreover, large reduction of dark current and sharper cutoff have been found for AlGaN solar-blind photodetectors fabricated on ELO substrates [30.65]. These examples provide the best evidence that the application of ELO technology led to significant progresses in the development of high-performance semiconductor devices made of lattice-mismatched epitaxial structures.

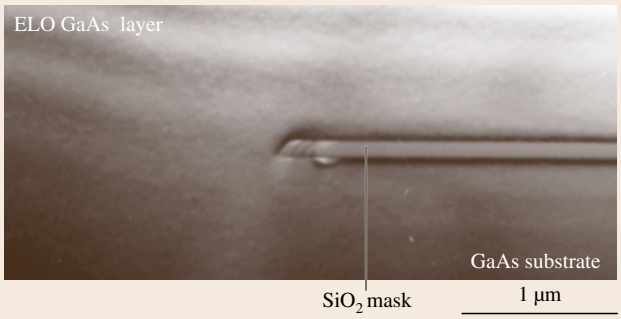


Fig. 30.20 TEM image of GaAs ELO/GaAs substrate interface at the edge of the SiO_2 mask. Note that no dislocation is seen above the mask edge

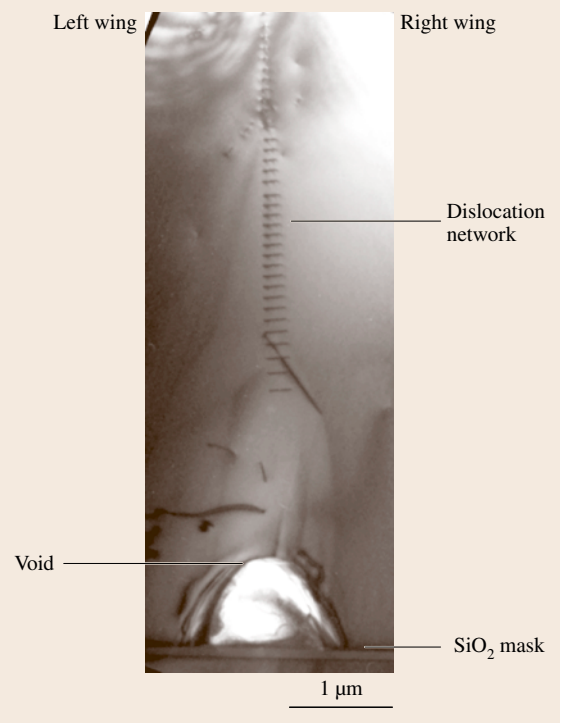
30.3.2 Structural Perfection of Coalescence Front in Fully Overgrown ELO Structures

In order to ensure high quality of ELO layers special care must be taken to avoid the generation of new defects during lateral overgrowth. The areas above the mask edges and fronts of coalescence of layers grown from neighboring seeds are the most sensitive since stress is the largest there. Therefore, factors influencing the structural perfection of ELO layers at these particular points will be briefly discussed.

Figure 30.20 shows a TEM image of the seeding area of the GaAs ELO layer grown by LPE on a SiO_2-masked GaAs substrate [30.66]. It is seen that the layer sticks closely to the SiO_2 mask. Contrast in the ELO image is very homogeneous, indicating that the overgrowth at the seed level is nearly stress free. Note that no dislocation is visible around the mask edge. As mentioned earlier, regular arrangements of dislocations generated during LPE growth in the vicinity of SiO_2 mask edge are sometimes observed in Si ELO layers [30.44]. Their creation leads to the growth of defective layers such as those shown in Fig. 30.8. It seems, however, that in this particular case defects originated due to specific LPE growth conditions used and that their generation can be completely avoided if the growth parameters are carefully adjusted. Indeed, the same group of authors has reported LPE growth of completely defect-free Si ELO layers [30.40]. A quite different behavior is observed in GaN ELO structures, in which dislocations above the mask edge are commonly found (see the D2 defect in Fig. 30.19). They are related to tilting of ELO wings as will be discussed in Sect. 30.4.1.

Coalescence fronts (seams) of ELO layers grown from adjacent seeding windows in fully overgrown ELO structures are other points where large mechanical stresses might be present. Thus, nucleation of dislocations is likely to occur there. Figure 30.21 presents a front of partial coalescence of two GaAs ELO layers grown from neighboring seeds by LPE on SiO_2-masked GaAs substrate [30.66]. From geometrical considerations the sidewalls of ELO layers were identified as close to the {232} planes. When approaching each other these planes separated mask area from the solution, thus the growth just above the mask terminated and a small void remained at the front of coalescence. Then the upper hollow filled up and the continuous ELO layer with a flat upper surface was finally obtained.

Creation of voids can significantly deteriorate the perfection of ELO layers. During their liquid-phase growth the solution is likely to be trapped inside the void. Since the thermal expansion coefficients of metallic inclusion and surrounding crystal are usually different this may lead to stresses that extend from the inclusion into the epitaxial layer and eventually to the formation of dislocations. This issue was extensively studied by *Nagel* et al. [30.67] and *Banhart* et al. [30.68] for LPE growth of Si ELO layers on thermally oxidized



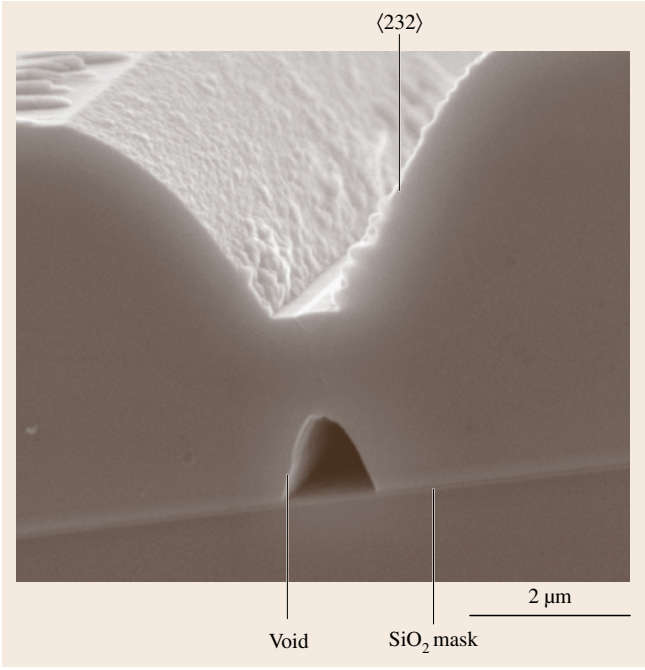


Fig. 30.21 SEM image of the front of partial coalescence of two GaAs ELO layers grown by LPE on SiO_2-masked GaAs substrate. Note the presence of inclined ⟨232⟩ sidewalls of ELO

Si substrates. They found that shape, size, orientation, and arrangements of seeding windows must be carefully designed for high quality of coalescence front. In particular, they have shown that 150 µm-long defect-free seams can be achieved in silicon lateral overgrowth by LPE if advancing layers start to join at only one point. Then the process should proceed in a zipper-like way in the direction of the seeding line. It is important that the growth solution could flow away along the window direction from the space between advancing growth fronts. When the merging layers join at two or more different points of the seam and coalescence proceeds inwards, solute inclusion and related dislocations are observed at the last point of coalescence. This usually happens if parallel seeding windows are used. Therefore, the application of nonparallel seeds has been advised to facilitate the zipper-like mechanism. This approach has been successfully employed

Fig. 30.22 TEM image of the front of coalescence of two GaAs ELO layers grown on SiO_2-masked GaAs substrate. Note the void at the mask surface and a well-organized network of dislocations separated by a distance of about 100 nm ◀

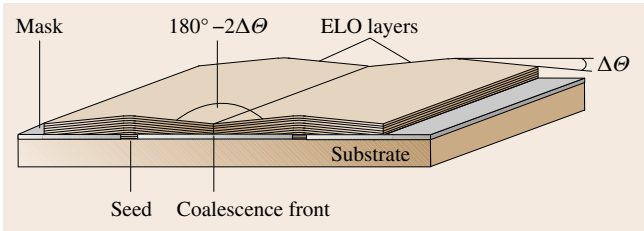


Fig. 30.23 Schematic drawing of merged ELO layers with their laterally grown parts tilted towards the mask. Since wings are tilted in opposite directions, the misorientation of the lattice planes at the seam is $2\Delta\Theta$, where $\Delta\Theta$ is the tilt angle of each wing

for defect-free coalescence of GaAs/GaAs [30.69] and InP/InP [30.70] ELO layers by LPE.

Figure 30.22 shows a TEM image of a coalescence front of two GaAs ELO layers grown by LPE on a SiO_2-masked GaAs substrate [30.66]. A void at the mask surface, similar to that seen in Fig. 30.21, is clearly visible. Probably, it was originally filled with gallium solvent that was later removed during the ion milling of the specimen. This might explain why some single dislocations surrounding the void are seen in its vicinity. The most important feature, however, is a well-organized set of dislocations starting at a distance of about 1 μm from the void and creating a low-angle grain boundary at the seam plane. These dislocations are parallel to the ⟨010⟩ direction with a Burgers vector of $\frac{1}{2}$[110], both included in the (001) plane that is parallel to the surface of the substrate. Creation of the grain boundary at the seam can be explained as follows: very often laterally grown sections of ELO layers are tilted towards the mask with the tilt axis aligned parallel to the seeding window direction. This situation is schematically illustrated in Fig. 30.23 where the wing tilt angle is marked by $\Delta\Theta$. The phenomenon of wing tilting will be extensively discussed in Sect. 30.4.1. However, it must be noted here that it creates significant problems for coalescence of neighboring ELO layers as they tilt in opposite direction. Since the sense of bending must be reversed at the seam the crystal lattice is heavily stressed in this region and dislocations appear as soon as the critical shear stress is reached. As discussed by *Banhart* et al. [30.68] for LPE growth of Si ELO layers on thermally oxidized Si substrates the stress at the seam increases with the thickness of merging layers and the value of the tilt angle $\Delta\Theta$. They have found defect-free coalescence for Si ELO layers thinner than 3–4 μm and tilted less than 0.1°. For larger thicknesses and/or values of the tilt angle $\Delta\Theta$ generation of a low-angle grain boundary similar to that shown in Fig. 30.22 was observed.

The dislocation network created in our GaAs ELO layers is similar to the D1 defect found at the center of the mask in GaN ELO layers (Fig. 30.19). *Sakai* et al. claim that this grain boundary consists of dislocations that had propagated laterally in the GaN wing area and had changed their direction again in the seam area [30.57]. On the contrary, the density of threading dislocations in GaAs ELO layers is very low. Therefore, in our case new dislocations must have been generated and rearranged into the low-angle grain boundary at the seam to accommodate the mutual tilt of merging wings. A similar behavior has been reported by *Shih* et al. for fully overgrown Si ELO layers grown by chemical vapor deposition on SiO_2-masked Si substrates [30.71]. The result they found, i.e., that the number of seam defects increases with increasing buried oxide width, can be explained by larger tilt angles $\Delta\Theta$ for wider overgrowths.

The geometry of the dislocation network shown in Fig. 30.22 can be used to measure the relative misorientation of merging ELO lattice planes [30.66]. The value of the tilt angle $\Delta\Theta \approx 0.1°$ obtained in this way is nearly ten times smaller than that determined for the GaN on sapphire ELO layer shown in Fig. 30.19. Possible reasons and consequences of that will be discussed in the next section.

30.4 Strain in ELO Layers

All examples presented thus far prove that ELO layers are of much higher quality than the reference planar structures. Even if new dislocations are created at the fronts of coalescence the overall density of dislocations is still significantly reduced. However, ELO layers are not free of strain. In particular, there is a question about strain induced by the mask itself and/or its possible interaction with the overgrown epitaxial layer. Moreover, the lattice mismatch and thermal strain induced by different thermal expansion coefficients of various parts of ELO structure may result in large deformations of the layers. Although in real heteroepitaxial ELO structures all these phenomena take place simultaneously, for clarity of presentation we will first focus on the

problem of interaction of ELO layers with the mask underneath (Sect. 30.4.1). Then, the issue of thermal strain in ELO structures will be addressed (Sect. 30.4.2). It is important to point out that ELO samples have a specific geometry. They consist of separated or coalesced monocrystalline stripes grown from line seeds cut in the mask, so different properties in the directions parallel and perpendicular to seeding lines may be expected. Therefore, before proceeding further a brief introduction will be given, presenting how the x-ray diffraction (XRD) technique is used to study a strain field in ELO layers.

Figure 30.24 shows the geometry commonly used for analysis of ELO layers by XRD. First, x-ray diffraction measurements are taken for the sample position in which the scattering plane (defined by the incident and diffracted wavevectors) is perpendicular to the seeding line direction. This corresponds to the axis of sample rotation during the ω scan being parallel to the seeds ($\varphi = 0°$). Next, the sample is rotated around the substrate normal and the measurements are repeated for the x-ray scattering plane being parallel to the seeds (i. e., for $\varphi = 90°$). Reciprocal space maps and/or x-ray diffraction curves (so-called *rocking curves*) are recorded for each sample positions.

Usually ELO layers are much narrower than the x-ray beam used in the standard x-ray diffraction experiments. Therefore, the diffraction pattern obtained contains information integrated over many stripes illuminated by the beam. To analyze strain distribution in a more detail we have developed a XRD technique in which the sample is moved in small steps and the x-ray beam, being much narrower than the ELO width, illuminates in sequence various parts of a single ELO stripe. For each position of the beam the diffraction pattern originating from a precisely defined area of the stripe is recorded. This allows us to analyze strain field *locally* inside a particular ELO stripe chosen. As will be shown later, this procedure allows the most characteristic features of the strain field in the samples to be determined.

30.4.1 Mask-Induced Strain in Homoepitaxial ELO Layers

Figure 30.25 shows x-ray rocking curves of GaAs ELO layer grown by LPE on SiO_2-masked GaAs substrate [30.7]. The layer was 10 µm thick and 90 µm wide. Both, the as-grown sample and that cut from the same wafer and etched to remove selectively the SiO_2 mask have been studied. The rocking curves presented in Fig. 30.25a,b have been measured with the $\varphi = 0°$ and $\varphi = 90°$ sample orientation, respectively. As can be seen from Fig. 30.25a the rocking curve of the as-grown

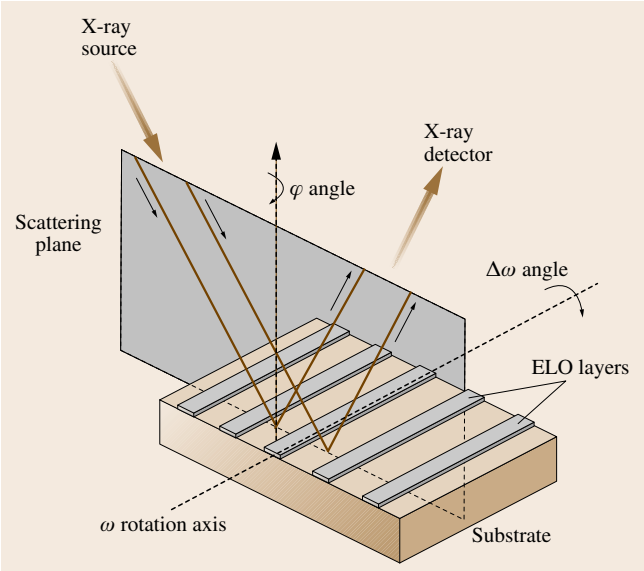


Fig. 30.24 Sketch of the geometry used for x-ray diffraction studies of ELO samples. The scattering plane is defined by the incident and diffracted vectors. φ is the angle between the direction of the seeds and the axis of sample rotation during the ω scan; $\varphi = 0°$ for the geometry shown

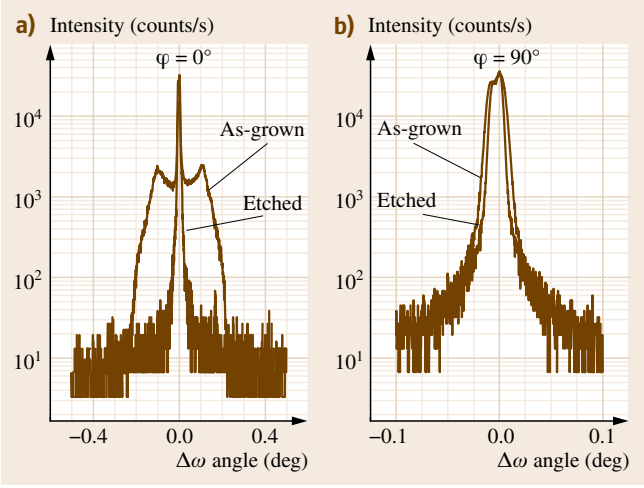


Fig. 30.25a,b X-ray rocking curves of (400) CuK_{α_1} reflection from the as-grown and etched GaAs ELO layer on the SiO_2-covered GaAs substrate measured with the ω-axis parallel **(a)** and perpendicular **(b)** to the seeding lines

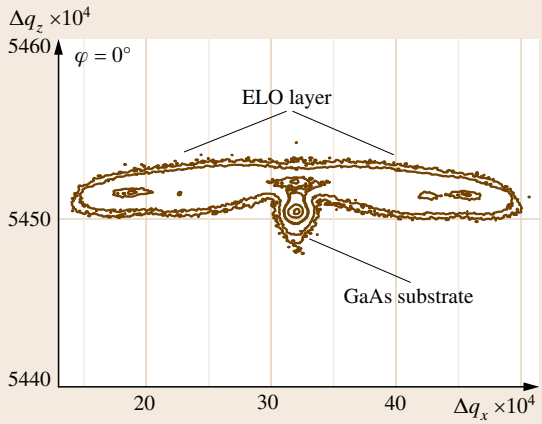


Fig. 30.26 Reciprocal-space map of (400) CuK_{α_1} reflection from the as-grown GaAs ELO layer on a SiO_2-covered GaAs substrate with the ω-axis parallel to the seeding lines. Both axes are marked in units of $\lambda/2d$, where λ is the wavelength of the x-ray radiation and d is the GaAs lattice spacing

sample is very broad. However, it becomes much narrower after the SiO_2 mask has been removed. On the contrary, for $\varphi = 90°$ the rocking curve is quite narrow and etching causes only a slight change of its shape (Fig. 30.25b). It is worth mentioning that during rocking curve measurements the x-ray diffraction is sensitive mainly to the distortion of the crystal planes in the scattering plane. Therefore, the large width of the rocking curve shown in Fig. 30.25a indicates that a significant deformation of the as-grown ELO takes place in the cross section plane perpendicular to the seeds.

Figure 30.26 shows the reciprocal space map of the same ELO layer. This map has been constructed for the same sample geometry as that used to measure the rocking curves presented in Fig. 30.25a. The distribution of the diffracted x-ray intensity proves that the reason for the large width of the rocking curve is the ELO stripe deformation in the direction perpendicular to the seeding line. This deformation is elastic since a narrow well-concentrated diffraction picture has been observed on the reciprocal lattice map if this is taken for the sample with the SiO_2 removed [30.72]. There are also slightly different positions of the ELO and substrate peaks visible in Fig. 30.26 that indicates some lattice misfit ($\approx 7 \times 10^{-5}$) between the heavily Si-doped ELO layer and undoped GaAs substrate [30.7].

When developing our ELO growth technology, occasionally we found GaAs ELO layers with an air gap between their laterally grown parts and the mask [30.73]. A cross section of such a layer is shown in Fig. 30.27a. Note the wings hanging over the substrate without any contact with the mask surface. Note also that the (400) CuK_{α_1} rocking curves of the layer are very narrow for both positions of the scattering plane (Fig. 30.27b), which means that deformation of the ELO wings, if any, is negligible. Comparison of the data shown in Figs. 30.25a and 30.27b proves that deformation of GaAs ELO wings must be due to their interaction with the mask surface.

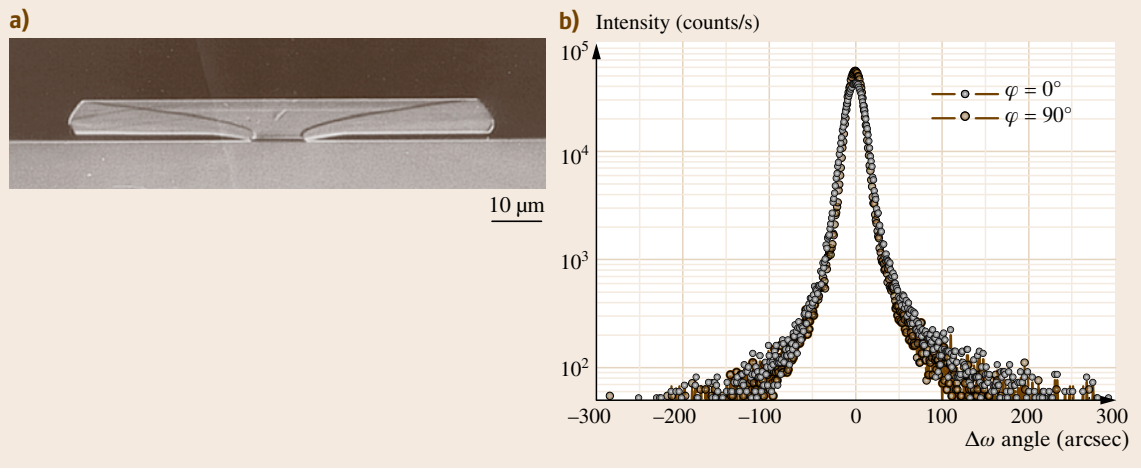


Fig. 30.27a,b Cross-section image **(a)** and x-ray rocking curves of (400) CuK_{α_1} reflection **(b)** of a GaAs ELO layer on a SiO_2-masked GaAs substrate. Note the air gap between the mask and the ELO wings

The results presented above allow us to present a picture of deformation of the ELO layer that is schematically illustrated in Fig. 30.28: During the growth the layer is bent towards the SiO_2 mask in the direction perpendicular to the seeding line (Fig. 30.28a). When the SiO_2 mask is removed by selective etching the ELO stripe behaves like a released spring and the wings tilt disappears (Fig. 30.28b). This corresponds to a narrowing of the curve shown in Fig. 30.25a. From a comparison of the rocking curves shown in Fig. 30.25b it can be concluded that similar deformation does not occur along the ELO stripe.

In the model presented in Fig. 30.28a it is postulated that the ELO wings are tilted towards the mask. Note, however, that the direction of tilt, upward or downward, cannot be directly determined from the shape of the rocking curve shown in Fig. 30.25a. Therefore, we have applied a spatially resolved XRD (SRXRD) technique in which the ELO stripe is moved in small steps, so the x-ray beam, being much narrower than the ELO width, illuminates in sequence various parts of the layer (see the geometry of the experiment sketched in Fig. 30.29c). For each position of the beam on the sample a local x-ray rocking curve originating from a precisely defined area of the stripe is recorded. The results of such a procedure are presented in Fig. 30.29b, which shows the set of x-ray rocking curves measured locally at various parts of a 304 μm-wide and 16.8 μm-thick GaAs ELO layer grown on SiO_2-masked GaAs substrate [30.74]. The standard rocking curve measured for the same sample with a 500 μm-wide x-ray beam is included in Fig. 30.29a for comparison. The diffraction plane was perpendicular to the seeds in these experiments. For the positions $x = -184\,\mu m$ and $x = 230\,\mu m$ the beam was out of the ELO area. Therefore, only narrow substrate peaks are visible. As soon as the beam approached the ELO stripe ($x = -138\,\mu m$) a strong side-maximum, apparently due to radiation diffracted by the ELO edge, appeared in the curve. For $x = 0\,\mu m$ only the middle part of the stripe was illuminated by x-rays. As a result, the side-maxima are missing whereas the intensity of the substrate peak is significantly reduced. When the beam starts to move out of the stripe surface ($x = 138\,\mu m$) the intensity in the central part of the curve decreased again, the side-maximum due to the edge reflection remained, while the intensity of the substrate peak increased. The data shown in Fig. 30.29b explains the origin of all the features visible in the standard rocking curve presented in Fig. 30.29a. In particular, this proves that the side-maxima are due to diffraction from the edges of the

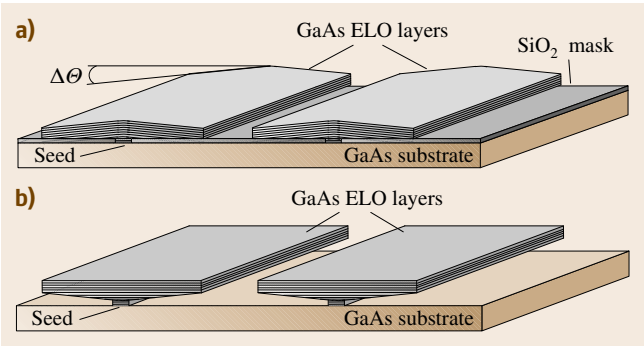


Fig. 30.28a,b Schematic drawing of the GaAs on GaAs ELO cross section. Bent layers are shown in the as-grown state **(a)**, deformation of ELO disappears when the SiO_2 mask is removed by etching **(b)**. $\Delta\Theta$ is the tilt of (100) crystal planes between the edge and the central part of ELO

ELO stripe. Therefore, the misorientation of ELO crystal planes must be the largest there. Consequently, half of the angular separation of the side-peaks on the standard rocking curve can be used as a good measure of the maximum tilt angle $\Delta\Theta$ of ELO lattice planes (compare Fig. 30.28a). Moreover, using a simple geometry of the diffraction system and analyzing the order in which the side-maxima appear while the beam moves across the stripe the direction of ELO tilt can be unambiguously determined. For example, the x-ray rocking curves of Fig. 30.29b show that the left wing diffracts x-rays at larger Bragg angle than the right one. This is clear evidence that the wings are tilted towards the mask [30.74]. It is worth mentioning that, improving the technique of spatially resolved x-ray diffraction presented above, we have employed it successfully for analysis of ELO layers using an x-ray beam as narrow as 20 μm [30.66]. Subsequently, even a 10 μm-wide beam moved in steps as small as 3 μm was applied, which allowed us to create x-ray rocking curve maps of bent ELO stripes [30.75]. Despite this progress the technique cannot be easily applied for GaN ELO layers. Since their widths are usually much smaller, electron diffraction [30.76] or synchrotron x-ray diffraction [30.77] must be used to determine the local magnitude and direction of the GaN wing tilt.

Additional evidence of the downward tilt of ELO wings comes from our studies of similar GaAs ELO samples by synchrotron x-ray topography (SXRT) [30.78, 79]. The basic idea of the experiment was to analyze how a white-beam synchrotron x-ray radiation collimated with the use of very narrow slits reflects back from the surface of the GaAs ELO sam-

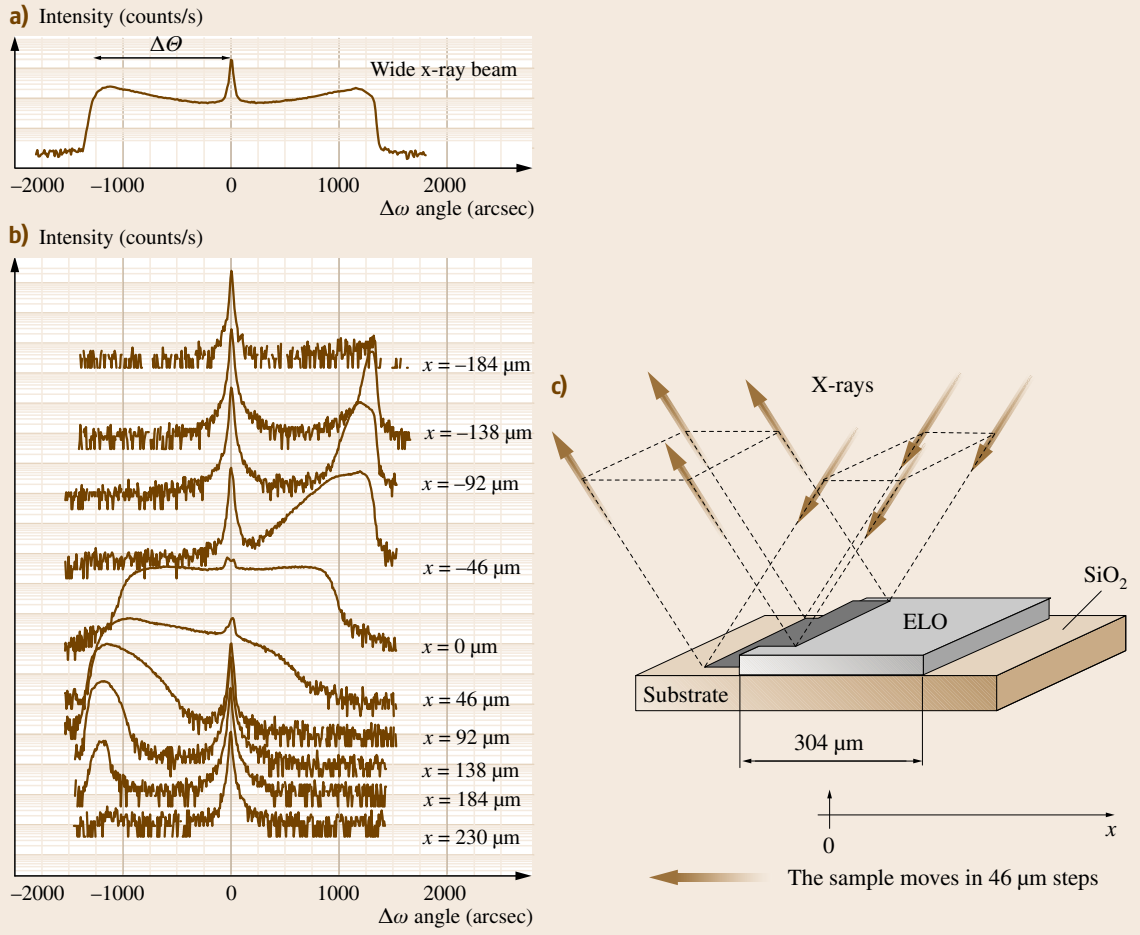


Fig. 30.29a–c X-ray rocking curves of a 304 μm-wide as-grown GaAs ELO layer on a GaAs substrate measured with the scattering plane perpendicular to the seeding lines with the use of a 500 μm-wide x-ray beam **(a)** and 90 μm-wide x-ray beam scanned across the ELO stripe in 46 μm steps **(b)**. x denotes the position of the beam on the ELO. The $x = -184$ and 230 μm curves correspond to reflections from the substrate outside the ELO stripe. **(c)** Schematic geometry of the experiment

ple. More details on experimental conditions can be found in the source publications. The topograph obtained is shown in Fig. 30.30a where "w," "s," and "e" mark images of window seed, substrate, and ELO layer areas, respectively. The shape of the ELO layer image directly indicates a downward tilt of the GaAs ELO wings. Additionally, the image was simulated numerically (Fig. 30.30b) to gain information on the distribution of the (100) lattice plane tilt angle versus position across the single ELO stripe. This data allowed us to calculate the shape of (100) lattice planes in the sample as shown in Fig. 30.30c together with the experimental profile of ELO layer thickness determined by surface stylus profilometer. As seen, the numerical results agree very well with the experimental results.

It should be noted in Fig. 30.30c that the curvature of the (100) lattice planes is the largest in the central part of the ELO stripe. This finding has allowed us to postulate that bending of the ELO stripes starts at the very beginning of growth when the laterally overgrown parts are thin and flexible. Then, the bent crystal planes might be reproduced during subsequent growth, although retaining their shape [30.78]. As will be shown later, this is exactly the behavior observed experimentally.

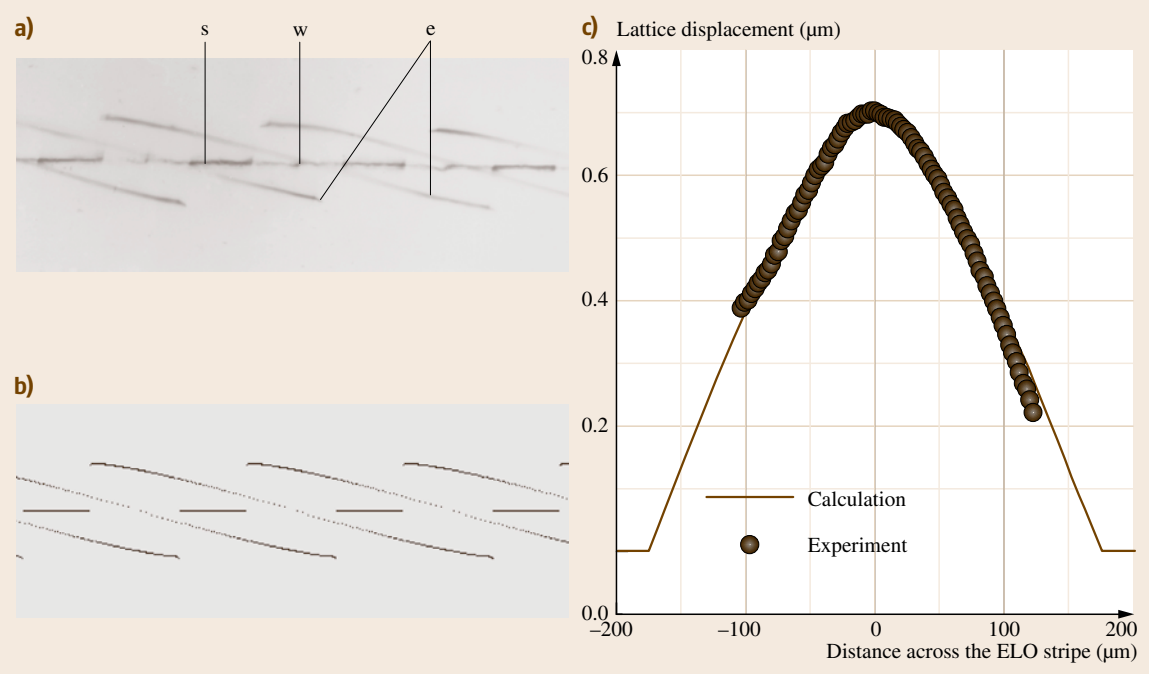


Fig. 30.30a–c 206 back-reflection section synchrotron x-ray topograph measured **(a)** and calculated **(b)** for GaAs ELO layer grown by LPE on SiO_2-masked GaAs substrate; "w," "s," and "e" mark topographic images of window seed, substrate and ELO layer, respectively; calculated (*solid line*) and measured (*dots*) displacements of ELO lattice planes across the single ELO stripe are shown in **(c)**

Note also that the data shown in Fig. 30.30c indicates a continuous decrease of layer thickness from the center of the layer to its edge, being due to the wing tilt. Indeed, this effect is observed experimentally as, for example, is clearly seen in the cross section image presented in Fig. 30.15. Its magnitude, namely the thickness change of $\approx 0.9\,\mu m$ on a $\approx 170\,\mu m$-wide wing, also agrees with data shown in Fig. 30.30c. A similar phenomenon has been reported for LPE-grown Si ELO layers. Since the upper ELO surface was a singular, atomically smooth face, tilting of the wings should lead to the presence of monoatomic steps at the Si/SiO_2 interface where silicon crystal planes terminate (see the sketch in Fig. 30.28a). Indeed, analysis of Si ELO layers by high-resolution transmission electron microscopy revealed the presence of such steps at the interface between the epitaxial layer and the SiO_2 film [30.17].

The tilting of ELO wings towards the mask has also been found for many other ELO systems, including Si layers grown by LPE on oxidized Si substrates [30.68] and GaN ELO layers grown by HVPE or MOVPE on sapphire or SiC substrates coated by SiO_2 or Si_3N_4 masks [30.57, 80–82]. *Fini* and coworkers used x-ray diffraction to observe the time evolution of lattice planes bending in GaN on sapphire ELO layers [30.83]. The scans in reciprocal space through the $(10\bar{1}3)$ diffraction peak of GaN obtained in situ, during growth of the structures by MOVPE, are shown for various growth times in Fig. 30.31a. Prior to epitaxy only the central peak, due to the diffraction from the GaN buffer, is visible. Emergence of the side-maximum is evident for the growth times $\approx 100\,s$, and by $\approx 300\,s$, distinct side-peaks have evolved, indicating wing tilt of $\approx 0.9°$. Postgrowth SEM studies of layer cross sections have shown that a significant lateral overgrowth started just at that moment. The wing peak narrows during subsequent growth while its position changes slightly, reaching a tilt of $\approx 1.19°$ after 3600 s of growth. The data collected during this elegant experiment are direct evidence that wing tilt must be due to interaction of the wing with the mask and that it starts at the very beginning of lateral overgrowth, in agreement with our earlier suggestion presented above when discussing the results shown in Fig. 30.30c [30.78].

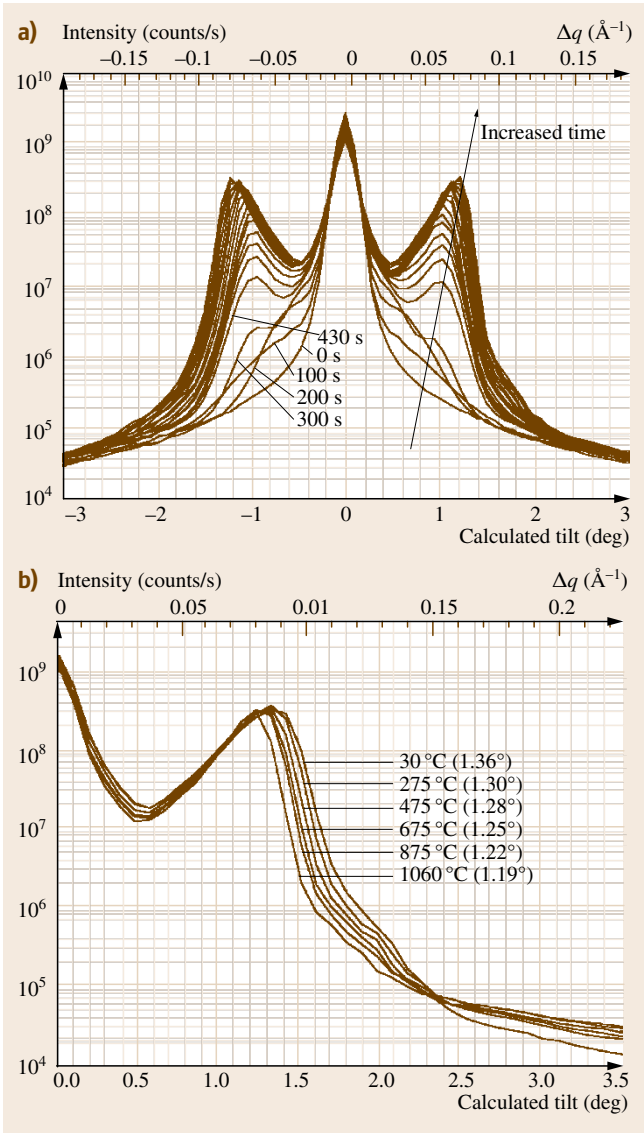


Fig. 30.31a,b Scans in reciprocal space trough the $(10\bar{1}3)$ x-ray diffraction peak obtained for various growth times during MOVPE growth of GaN on sapphire ELO structures **(a)** and on cooling to room temperature **(b)**. The scans in **(b)** are shown over half of the range measured (courtesy of Fini)

The difference between the GaN and GaAs or Si ELO structures is the magnitude of wing tilt and the mechanism of its relaxation. As discussed earlier, tilt angles $\Delta\Theta \approx 0.1°$ are commonly found in LPE-grown GaAs ELO layers (Fig. 30.25a). Thus, the bending strain at the seed–mask edges is small enough to be accommodated elastically. Generation of dislocations in the layer does not take place and deformation of the wings disappears completely when the mask is removed by postgrowth etching. Consequently, no defects above the mask edge (compare Fig. 30.20) and narrow x-ray rocking curve of the sample with the mask removed (Fig. 30.25a) are observed. On the contrary, tilt angles up to $\approx 1.5°$ are reported in literature for GaN ELO layers (Fig. 30.31a and [30.57, 80]). So large tilts cannot be accommodated elastically and the strain above the mask edge partially relaxes via the creation of a low-angle grain boundary. This relaxation process is energetically favorable if the ELO structure contains dislocations threading from the buffer. In such a case, the grain boundary can be easily formed from bent threading dislocations that pile up near the mask–seed edges. Indeed, as shown already in Fig. 30.19, cross sectional TEM studies of GaN on sapphire ELO structures have revealed additional low-angle tilt boundaries (the D2 defects) above the edges of the SiO_2 mask. Consequently, when the mask is etched away after the growth, similarly to our approach for GaAs/GaAs layers, only the elastic component of the bending strain is released. The plastic deformation of the crystal lattice above the mask edge remains, and as a result, only partial narrowing of the x-ray rocking curve takes place. This is exactly the behavior observed experimentally [30.81].

All the arguments presented up to now indicate that downward tilt of ELO wings originates from an interaction between laterally grown material and the mask. The nature of this interaction is still not fully clarified. *Kohler* with coworkers have suggested that an attractive van der Waals force starts to be active during LPE growth when the mutual distance between Si ELO wing and the upper surface of the SiO_2 mask falls below a certain value [30.84]. The surface tension of the solution should also be taken into account as a source of wing tilting during ELO growth from the liquid phase [30.47]. However, the tilt of GaN ELO wings seems to be too large to be related to adhesion or weak van der Waals forces. Therefore, a chemical reaction of laterally overgrown GaN with the SiO_2 underneath and/or densification of the oxide being due to a high temperature and chemically aggressive conditions (e.g., presence of ammonia) during vapor phase epitaxy of GaN have been considered as possible reasons of the tilt [30.83].

As discussed in Sect. 30.3.2, tilting of ELO wings towards the mask creates a significant problem for coalescence of neighboring stripes as they tilt in the opposite direction (Fig. 30.23). Therefore, there is no

doubt that for ELO layers of high crystallographic quality tilting of their wings should be reduced as much as possible.

We have proposed the following recipe for reduction of the mask-induced bending of the ELO layers: The initial vertical growth rate of ELO must be increased to start the lateral overgrowth at some microscopic distance from the upper surface of the mask [30.74]. Then the chance of ELO wings capture by attractive force and their interaction with the mask should be reduced. We have found that this recipe works efficiently in practice. In particular, tilting of the wings has been efficiently tailored by controlling the ratio of vertical to lateral growth rates at the beginning of ELO growth. This has been achieved by growing GaAs ELO layers on SiO_2-coated GaAs substrates with increasing density of dislocations. Then, the ratio of vertical to lateral growth rates at the beginning of the growth was increased due to the higher density of surface steps, which in turn led to reduction of the mask-induced tilt of ELO wings [30.73, 74]. In the limiting case of heavily dislocated GaAs substrates, namely on GaAs buffers grown by MBE on Si substrates, the vertical growth of GaAs ELO was so fast that air-bridged structures without any interaction with the mask were obtained [30.55]. Basically the same approach has been used by *Fini* et al. [30.86]. They have shown that the crystallographic quality of coalescence of neighboring ELO GaN stripes can be improved if vertical development of ELO stripes is forced at the beginning of the growth, followed by a change of growth conditions and fast lateral overgrowth of the structure.

If thermally activated processes during the growth itself are considered as origins of wing tilt, reduction of the ELO growth temperature might be helpful. Indeed, *Silier* with coworkers have reported negligible adhesion to the SiO_2 mask of Si ELO layers grown by LPE at 450 °C, in contrast to their previous growth experiments performed at high temperature of ≈ 900 °C [30.17]. Also replacement of commonly used dielectric films by other mask materials seems to be a promising way of reduction of mask-induced wing tilt. In particular, we have found that bending of GaAs ELO layers grown by LPE is negligible when the SiO_2 mask is replaced by a thin graphite film [30.7]. This finding has been explained by delayed beginning of lateral growth due to a change of the shape of the liquid solution in the vicinity of the ELO edge when the SiO_2 was replaced by graphite film not wetted by the gallium melt [30.73]. Similarly, tungsten has been found to be a promising mask material for MOVPE growth of GaN ELO structures with very low tilt of the wings [30.4]. In principle, however, it is difficult to predict a priori the mask material that would be the most suitable for each particular ELO case.

The most natural way to eliminate the mask-induced strain would be to grow the ELO structures with their wings having no contact with the mask surface. An example of such a GaAs structure grown by LPE was shown in Fig. 30.27. Another example is shown

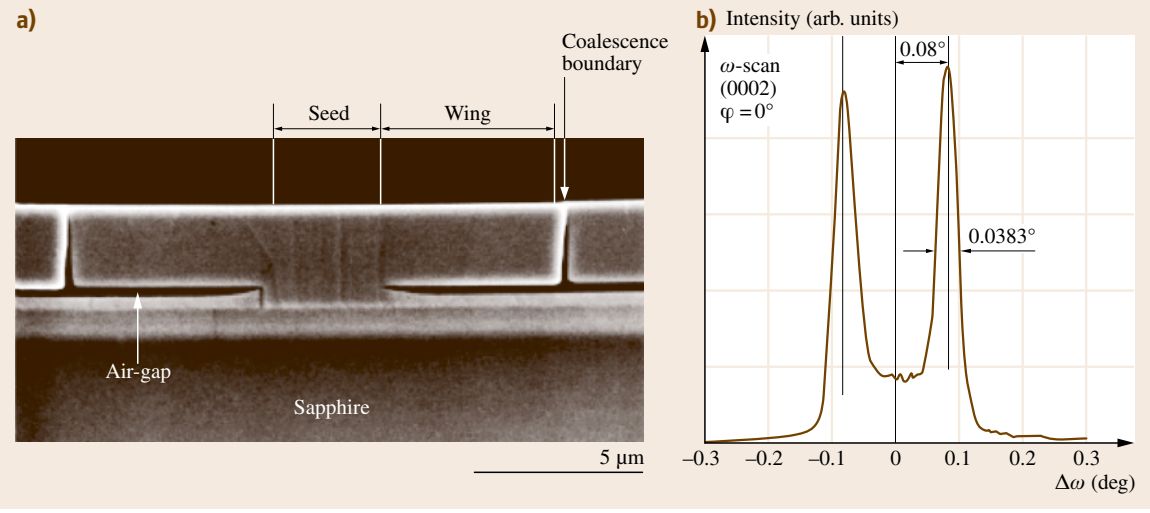


Fig. 30.32a,b Cross-sectional SEM image (**a**) and x-ray rocking curve measured for the scattering plane perpendicular to the seeding lines (**b**) of the air-bridged GaN on sapphire layers (after [30.85], © AIP 2000)

in Fig. 30.32a, which presents a cross section of air-bridged GaN layers grown by MOVPE on sapphire substrates coated by GaN buffer and silicon nitride mask [30.85]. Note that, due to profiling of the GaN buffer thickness, an air gap is formed during the growth above the mask and a free-standing laterally grown GaN is obtained. In this way, direct interaction of ELO wing with the mask is eliminated. Figure 30.32b shows the x-ray rocking curve of the layer measured with the x-ray scattering plane perpendicular to the seed direction. Two peaks originating from ELO wings are clearly visible. Each of them is much narrower than that obtained for standard GaN ELO layer (compare Fig. 30.31a), which means that the mask-induced bending of GaN wings is significantly suppressed. However, the different angular position of both peaks indicates that some residual tilt ($\approx 0.08°$) of the *c*-axis in the wing area relative to the underlying GaN is still present. Its origin is explained in the next section.

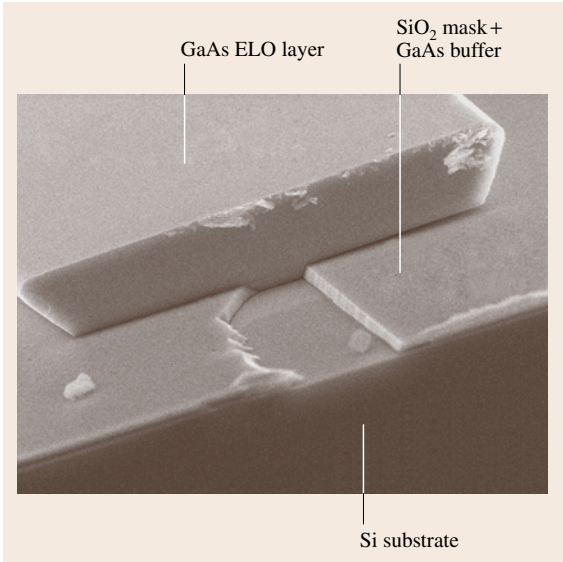


Fig. 30.33 Cross-sectional SEM image of the as-grown GaAs ELO layer on GaAs-coated Si substrate; the layer is 62 µm wide and 9.2 µm thick. Note that ELO wings overhang the mask

30.4.2 Thermal Strain in ELO Layers

Besides the mask-induced strain an additional deformation of ELO lattice planes may arise when the sample experiences a large stress upon cooling from the growth to room temperature due to the different thermal expansion coefficients of its components. This effect is commonly observed in planar heterostructures. In the following we will show that in the ELO structures thermal strain can be relaxed via additional tilting of the wings while still preserving their high quality.

Figure 30.33 shows a SEM image of a GaAs ELO layer grown by LPE on a silicon substrate coated by

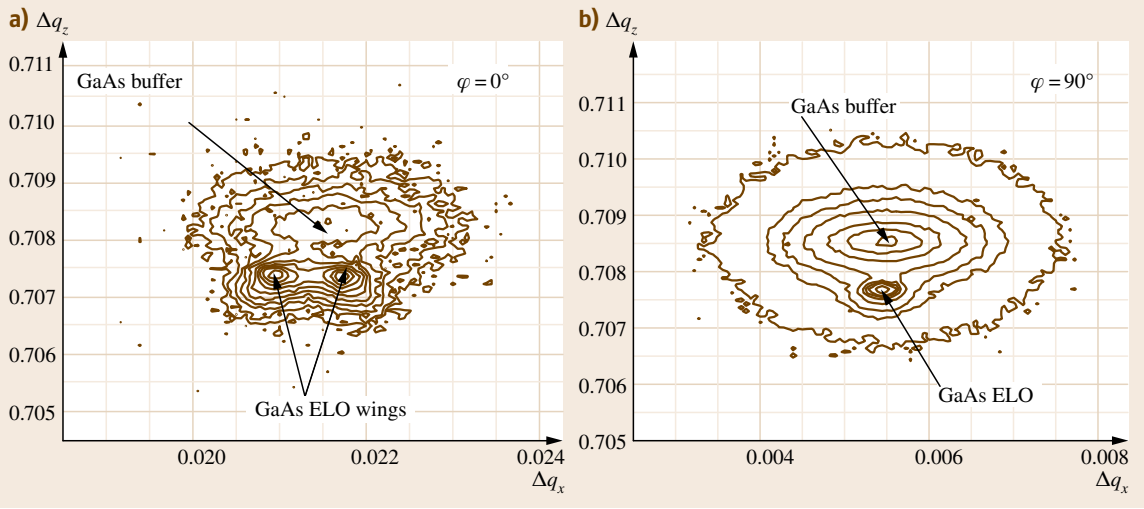


Fig. 30.34a,b Reciprocal-space map of (400) CuK_{α_1} reflection of the as-grown GaAs ELO layer on GaAs-coated Si substrate measured with the scattering plane perpendicular **(a)** and parallel **(b)** to the seeding lines. *Axes* are marked in units of $\lambda/2d$, where λ is the wavelength of the x-rays and d is the lattice spacing

a MBE-grown GaAs buffer layer and a SiO_2 mask. The layer was grown at 540 °C and was 62 µm wide and 9.2 µm thick. More details of the growth procedure can be found in the source publication [30.55]. Note that the layer does not stick to the substrate and its wings overhang the mask surface. This is the result of a fast vertical growth rate caused by buffer dislocations supplying surface steps as discussed in the previous section. Reciprocal-space maps measured for the same sample with the x-ray scattering plane perpendicular and parallel to the seeds are presented in Fig. 30.34a,b, respectively. The broad Bragg peak from the GaAs buffer as well as narrow, well-separated reflections originating from ELO wings are clearly visible. Note that, for $\varphi = 0°$, reflection from the layer consists of two peaks, whereas a single Bragg peak is visible only for $\varphi = 90°$. This is a quite similar behavior to that observed for our GaAs/GaAs ELO samples (compare Fig. 30.25) and indicates tilt of ELO wings in the plane perpendicular to the direction of the seeds.

Figure 30.35 shows x-ray rocking curves of the same sample before and after etching to remove selectively the SiO_2 mask. Note that the full-width at half-maximum of the reflection from the wing equals 94 arcsec, only. This is much less than that of the MBE-grown buffer ($\approx$ 435 arcsec) and confirms the very high structural quality of laterally overgrown parts of the layer. However, the whole curve is so wide that tilt of ELO wings caused by their interaction with underlying

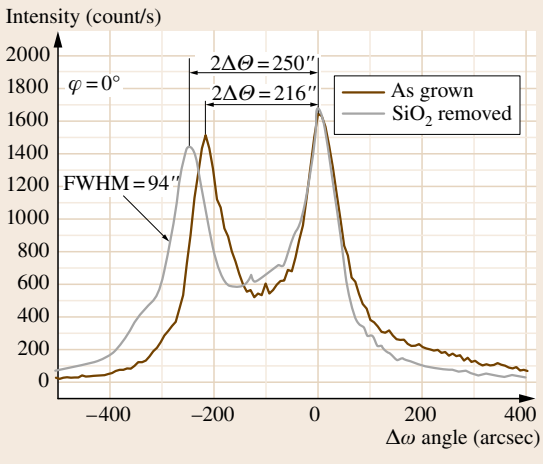


Fig. 30.35 X-ray rocking curves of (400) CuK_{α_1} reflection for the as-grown (*solid line*) and etched (*dashed line*) GaAs ELO layer on the SiO_2-coated Si substrate with GaAs buffer, with the x-ray scattering plane perpendicular to the direction of the seeding lines

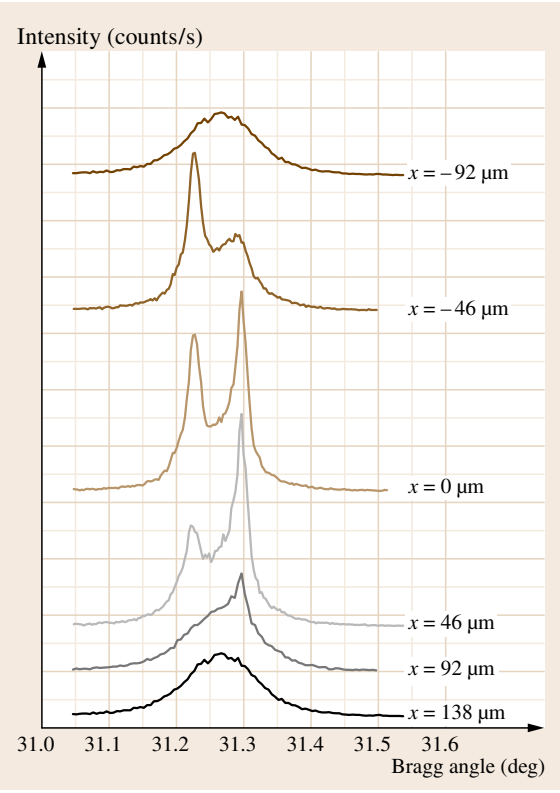


Fig. 30.36 (400) CuK_{α_1} rocking curves from as-grown GaAs on a Si ELO layer measured with a 90 µm-wide x-ray beam moved across the ELO stripe in 46 µm steps. The $x = -92$ and 138 µm positions correspond to reflections from the GaAs buffer outside the ELO stripe

SiO_2, similar to that presented for homoepitaxial GaAs ELO layers, might be suspected. It is worth noticing, however, that this time behavior of the system is quite different. The wing tilt angle increases when the mask is removed. Moreover, the wings hang over the mask, so the mask-induced strain, if any, should be negligible.

We have checked the direction of the tilt of the GaAs wings using the SRXRD technique described earlier (Fig. 30.29c). The set of x-ray rocking curves obtained from precisely defined areas of the GaAs stripe is shown in Fig. 30.36. The geometry of the experiment was exactly the same as before, i. e., the beam illuminated the left ELO wing first ($x = -46$ µm), then the central part of the ELO stripe ($x = 0$ µm), and finally left the sample on its right-hand side ($x = 92$ µm). Note the order in which ELO diffraction peaks appeared when the beam was illuminating subsequent parts of the stripe. The left ELO wing diffracted x-rays at smaller Bragg an-

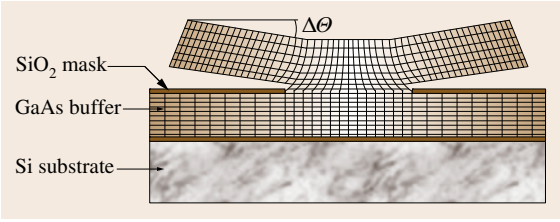


Fig. 30.37 Schematic drawing of the cross section of the as-grown GaAs ELO layer on a Si substrate. The tilt of the ELO wing is $\Delta\Theta$, where $2\Delta\Theta$ is the angular separation of the diffraction peaks, as marked in Fig. 30.35

gle than the right one. This is the opposite behavior to that observed for the GaAs/GaAs ELO system (compare Fig. 30.29) and indicates upward tilt of laterally overgrown parts of the layer [30.55]. All these findings indicate that processes other than wing adhesion to the mask must be involved.

We have explained the results presented above by taking into account the thermal strain that appears in the ELO structure during its postgrowth cooling to room temperature. This is illustrated in Fig. 30.37, which schematically shows deformation of lattice planes in the cross section plane of GaAs on Si ELO layer. A biaxial tensile strain caused by the different thermal contraction of epilayer and substrate is commonly observed in planar GaAs layers grown on Si substrates [30.87]. This strain disappears when the GaAs/Si structure is heated to $\approx 500\,^{\circ}\mathrm{C}$ [30.88]. Therefore, at the LPE growth temperature, the GaAs ELO layer grows essentially stress free. At room temperature, the basal plane of ELO layer should have the same tensile deformation in the direction perpendicular to the line seed as the upper surface of the buffer GaAs. We have shown that the ELO layer stands free and does not adhere to the mask. Therefore, unrestricted strain relaxation and free contraction of vertically grown volume of ELO should take place in its upper part, as shown in Fig. 30.37. This, in turn, must lead to the upward bending of ELO wings. Note, however, that this process does not affect the structural perfection of ELO wings, and the high crystallographic quality of laterally overgrown parts of the layer is preserved.

Finally, the shift of diffraction pattern seen in Fig. 30.35 for the etched sample should be explained. The thermal expansion coefficient of silicon oxide is much smaller than that for zincblende III–V semiconductors. Therefore, the SiO_2 film exerts a tensile stress on GaAs underneath and a compressive strain is present in oxide-free GaAs seeding windows [30.72]. In the GaAs/Si ELO system the thermal tensile strain in the GaAs seeding area exerted by the substrate is partially compensated by the mask-induced compression [30.55]. Therefore, deformation of ELO at its base, and consequently the ELO wings tilt angle, increase when the SiO_2 mask is etched away. This leads to increase of the angular separation of x-ray reflections, as shown in Fig. 30.35.

It is worth noting that our model of thermal strain relaxation in ELO layers predicts that the direction of wing tilt is correlated with the sign of the strain in the buffer. In particular, downward wings tilt should be observed if the buffer layer is under compressive thermal stress, which is usually the case for GaN ELO layers grown on SiC or sapphire substrates. This is exactly the behavior presented in Fig. 30.31 for the GaN on sapphire ELO system. At the growth temperature GaN ELO wings are tilted downward due to their interaction with the mask. Upon cooling to room temperature, the position of the wing peak increases from 1.19 to 1.36° (Fig. 30.31b), which means that an additional downward tilt induced by thermal stress has occurred. Qualitatively, the same phenomenon can be inferred from simulations via finite-element analysis of strain in ELO structures [30.89]. Most likely, the residual tilt found in GaN air-bridged structures illustrated in Fig. 30.32b might also be explained in a similar way by the thermal strain in the GaN buffer. All these examples prove that our model is correct. Moreover, they show that the phenomenon of thermal strain relaxation via tilting of ELO wings is not a specific attribute of GaAs/Si ELO structures grown by LPE but has much more general nature.

30.5 Recent Progress in Lateral Overgrowth of Semiconductor Structures

There is continuous progress in the application of the lateral overgrowth technology, and new versions of the ELO technique are being tested. In this section, we first present the recent progress made in liquid-phase ELO growth, including new mask materials and ELO growth by the liquid-phase electroepitaxy. Then, new concepts of ELO growth will be briefly reviewed and compared with the conventional ELO technique. Liquid-phase

epitaxy of InGaAs bridge layers and pendeo-epitaxy of GaN layers from vapor and liquid phase will be used as examples.

30.5.1 Developments in Liquid-Phase ELO Growth

Recent developments in liquid-phase ELO growth include the application of novel mask materials. In particular, electrically conductive masking films are interesting since they can be used in photovoltaic and thermophotovoltaic devices as back mirrors for photon recycling and/or as buried electrical contacts. However, the main problem is to find mask material that adheres strongly to the substrate and is thermally and chemically stable to remain undamaged during the aggressive conditions of epitaxial growth. Moreover, it is of prime importance that no nucleation takes place on the mask during the growth since in ELO technique the layer should start growing exclusively from the opening in the mask.

We have applied tungsten for substrate masking in lateral overgrowth of GaAs by LPE [30.5]. Tungsten masks were already successfully used to grow GaN ELO layers on sapphire by MOVPE and dramatic reduction of strain in the layers was found when tungsten film instead of common dielectrics was used to mask the substrate [30.4]. However, MOVPE growth conditions for GaN are quite different from those for solution growth of zincblende III–V compounds, so applicability of a W mask for selective growth of GaAs by LPE has to be verified experimentally.

Figure 30.38 shows a SEM image of GaAs layer grown at 600 °C on a tungsten-masked GaAs substrate. As seen, the epitaxial growth is perfectly selective, i. e., the layer starts growing from the line seed without any nucleation of GaAs on the mask. However, the tungsten mask peels off from the substrate, which often leads to cracking of the ELO layer. This effect is attributed to a large grain size and the strain built into the metallic film sputtered onto the bare GaAs. It has been reported that the application of a AuZn wetting layer between the substrate and the tungsten film helps to reduce the tungsten grain size, and consequently leads to lower strain inside the mask [30.90]. Indeed, we have found that use of a AuZn interlayer significantly improves adhesion of the W mask to the GaAs substrate during LPE growth of GaAs and cracking of ELO layers by the mask peeling off from the substrate is avoided [30.5]. Unfortunately, sometimes a melt removal from W-masked substrate after growth was incomplete since tungsten is well wetted

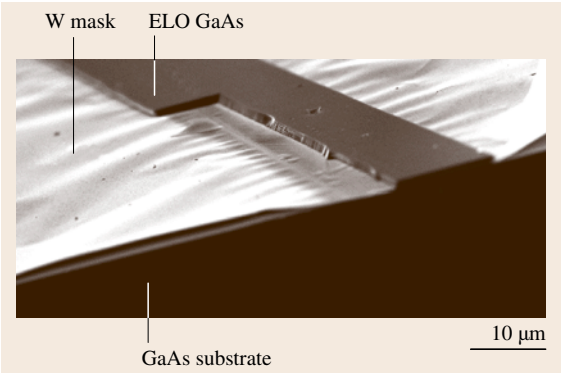


Fig. 30.38 SEM image of a GaAs ELO layer grown at 600 °C by LPE on a tungsten-masked GaAs substrate; mask deposited on bare GaAs substrate. Note that the mask peels off from the substrate, which leads to cracking of the layer

by gallium-rich solution. Consequently, local polycrystalline deposits from melt droplets left on the mask surface are often observed. This encouraged us to look for alternative mask materials.

We have found zirconium nitride films promising for electrically conductive masking in ELO technique [30.6]. This material has interesting properties resulting from its both metallic and covalent bonding characteristics. The covalent crystalline properties are: high melting point, extreme hardness and brittleness, and excellent thermal and chemical inertness. The metallic characteristics are high electrical conductivity and metallic reflectance [30.91]. Moreover, ZrN film when sputtered on a substrate usually has a fine-grained microstructure, thereby exhibiting a low strain level.

Figure 30.39 shows SEM image of a GaAs ELO layer grown at 750 °C by LPE on a ZrN-masked GaAs substrate. As seen from the figure, perfect growth selectivity is obtained. Moreover, the masking film is mechanically stable, adheres strongly to the substrate, and does not show any signs of mechanical degradation even at LPE growth temperature as high as 750 °C. A similar behavior can be seen in Fig. 30.40, which shows SEM image of a GaSb ELO layer grown at 480 °C on ZrN-masked GaSb substrate. This indicates that, due to the high stability of zirconium nitride, long time contact of the mask with Ga – As or Ga – Sb melts does not lead to degradation of its mechanical properties. Moreover, we have found that ZrN forms a low resistivity ohmic contact to n-type GaAs, which is important if the mask is to be further used as a buried electrical contact. Unfortunately, our experiments show

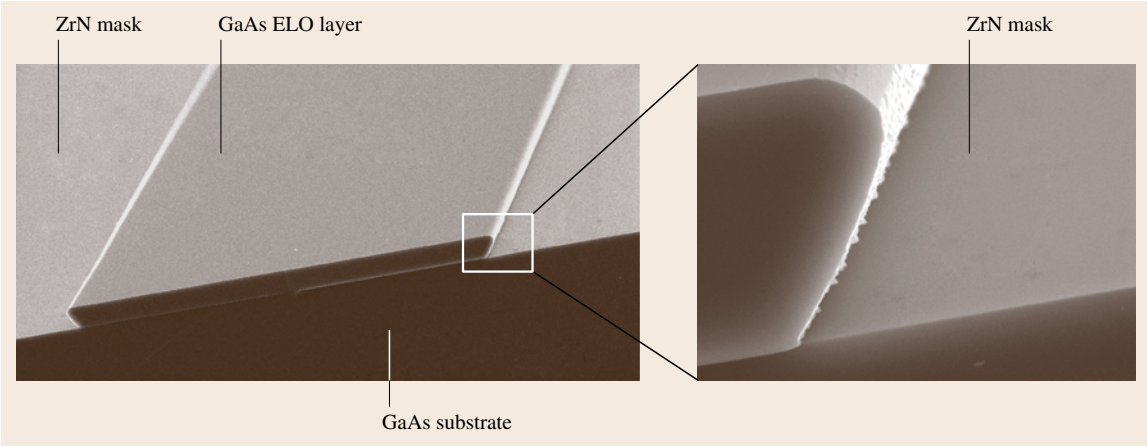


Fig. 30.39 SEM image of a GaAs ELO layer grown at 750 °C by LPE on a ZrN-masked GaAs substrate

that at temperatures higher than ≈ 580 °C the ZrN masks become highly resistive when they are heat-treated in hydrogen atmosphere during LPE growth. Most likely, ZrN is converted into insulating zirconium hydrides, though more work is needed to clarify this issue. This phenomenon sets an upper limit for LPE growth temperature if ZrN masks are to be used as electrical contacts. Thus, the ELO growth of GaAs becomes more difficult, though still possible. For ELO of GaSb layers however, typical LPE growth temperature is about 480 °C, which allowed us to grow high-quality GaSb ELO layers by LPE still preserving the high electrical conductivity of the ZrN mask.

As illustrated above, a proper choice of mask material is crucial for successful lateral overgrowth. However, masks that survive undamaged during a long growth process are difficult to find if extreme crystallization conditions are required. This is the case in solution growth of GaN. Since at typical conditions solubility of nitrogen in gallium is low, very high temperatures (≈ 1450 °C) and very high nitrogen pressure (≈ 1 GPa) are necessary to achieve a reasonable growth rate of GaN from Ga-N solution. *Bockowski* et al. have found that under such conditions silicon nitride, SiO_2, nickel, and molybdenum masks dissolve in the hot gallium while iridium reacts with Ga forming IrGa [30.92]. They have also found that tungsten platelets do not react with the Ga-N solution, so tungsten seemed to them a promising mask material for solution growth of GaN ELO layers. However this argument is not convincing enough since W films sputtered on a substrate have not been tested and other effects such as reduced chemical stability of thin films and cracking caused by built in strain or weak adhesion to the substrate (compare Fig. 30.38) might cause problems during ELO growth.

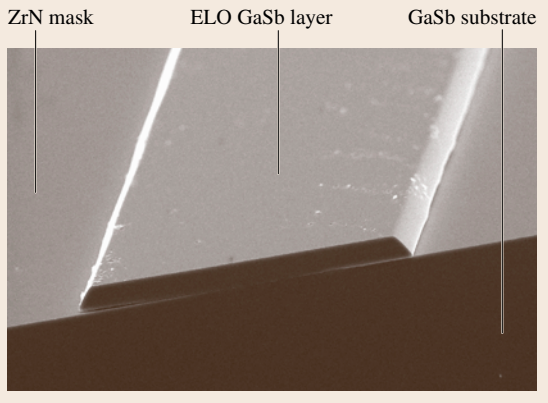


Fig. 30.40 SEM image of a GaSb ELO layer grown at 480 °C by LPE on a ZrN-masked GaSb substrate

Recently there has been increasing interest in lateral overgrowth of semiconductor layers by the liquid-phase electroepitaxy (LPEE) technique – a solution growth method in which layer crystallization is achieved by passing an electric current through the solid–liquid interface while the furnace temperature of the system is kept constant (see [30.93] for a review). The main mechanisms of solute transport toward the substrate during LPEE are diffusion due to the Peltier-effect-induced temperature gradient and electromigration of species (i. e., momentum exchange between free charge carriers and solute species). Technical realization of LPEE is slightly more complicated than that of LPE.

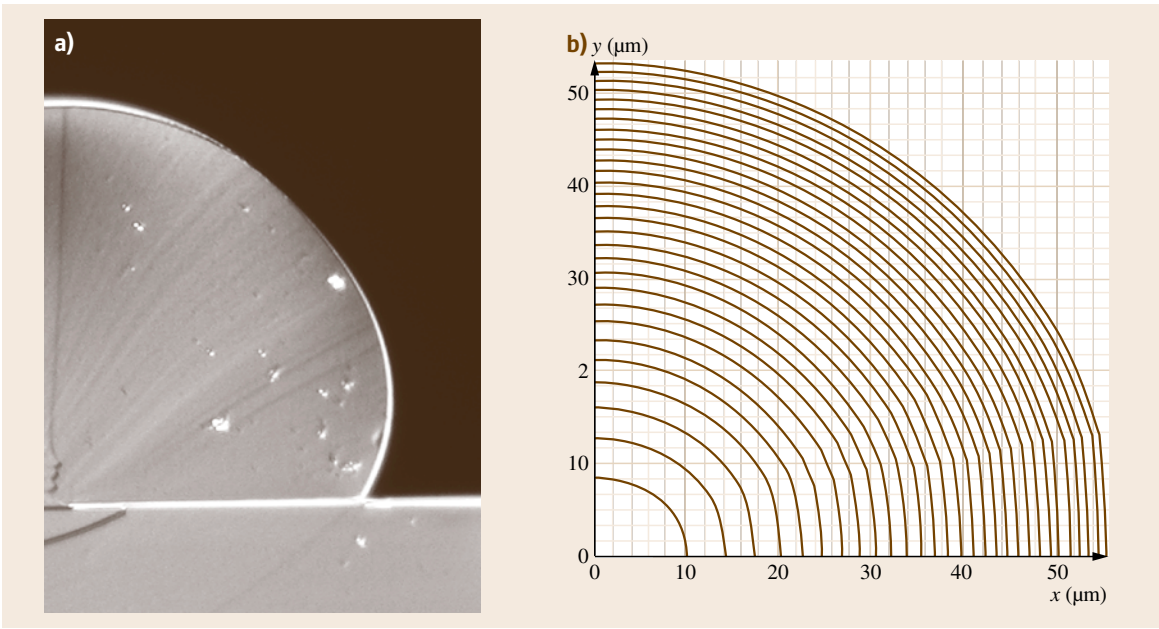


Fig. 30.41a,b SEM image of cross section (**a**) and calculated shape (**b**) of undoped GaAs ELO layer grown at 700 °C by LPEE on a SiO_2-masked GaAs substrate. The scale in both figures is the same for easier comparison. The time interval in (**b**) is 8 min

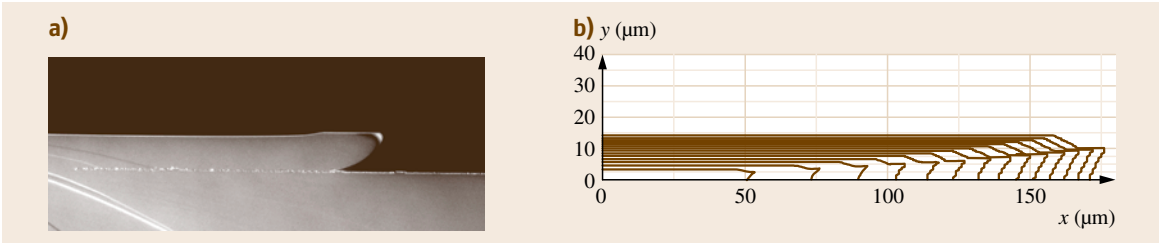


Fig. 30.42a,b SEM image of cross section (**a**) and calculated shape (**b**) of undoped GaAs ELO layer grown at 700 °C by LPEE on a GaAs substrate coated by electrically conductive (tungsten) mask. The scale in both figures is the same for easier comparison. The time interval in (**b**) is 16 min

However, electroepitaxy offers unique possibility to control the growth kinetics and properties of crystals produced. In particular, since the electric current flow affects the solute distribution in the liquid solution, by controlling the electrical current distribution areas with *local* supersaturation in the liquid zone can be produced, which is crucial for selective epitaxy techniques such as ELO. The progress achieved with the application of substrates coated by metallic masks has allowed a novel approach to ELO growth by LPEE.

Figures 30.41a and 30.42a show cross sections of undoped GaAs ELO layers grown at 700 °C by LPEE on (100) GaAs substrates. The difference between them is that they were grown on substrates masked by SiO_2 and tungsten films, respectively, while the growth parameters were kept the same in both cases (see [30.5] for more details). As seen, the ELO growth on the SiO_2-masked substrate is nearly isotropic while under the same conditions a wide and thin ELO layer is grown on the tungsten-masked substrate. This means that the shape and dimensions of the ELO layers grown by LPEE can be efficiently tailored by varying the electrical conductivity of the mask.

We have developed a simple mathematical model for LPEE growth of ELO layers, which predicts the electric current intensity and solute concentration dis-

tributions in the liquid zone under various LPEE growth conditions [30.94]. Moreover, the model allows for direct simulation of the evolution of the growth interface with time. Calculated shapes of the GaAs ELO layers grown by LPEE on GaAs substrates are plotted in Figs. 30.41b and 30.42b. They are shown on the same scale as the experimental images for easier comparison. As seen, the numerical results are in very good quantitative agreement with the experimental ones. The model allows us to postulate the following mechanism for ELO growth of semiconducting layers by LPEE: During growth on substrate masked by electrically insulating film electromigration of solute towards the seed dominates growth kinetics. Since the electric current distribution in the liquid zone around the seed is isotropic, this leads to a rounded shape of layers and to a low value of their aspect ratio despite surface kinetic processes on the upper ELO surface. On the other hand, if the substrate is coated by an electrically conductive mask and the resistivity of its contact to the substrate is low, after some initial ELO growth more electric current passes to the substrate directly through the melt–mask interface than through the ELO crystal and the seeding area. Here, the main role of electromigration is to supersaturate the solution just above the mask between the adjacent seeds. Then diffusion dominates ELO growth by supplying solute from these supersaturated areas directly to side-walls of growing layers. In agreement with our experiments, this mechanism significantly increases the lateral growth rate, thereby producing thin and wide ELO layers with flat upper facets. It is worth mentioning that a similar increase in the lateral growth rate in LPEE has been reported by *Mauk* and *Curran* for ELO of silicon from Si-Bi melt [30.95].

30.5.2 New Concepts of ELO Growth

The most characteristic feature of the ELO process presented so far is the change of the predominant growth direction – from vertical in the growth window to lateral in the regions over the masking film. In principle, however, it is possible to start lateral overgrowth from seeds oriented perpendicular to the substrate surface, so change of the growth direction is not necessary. If the substrate (bare substrate or substrate covered by a suitable buffer layer) is etched to create on its surface a pattern of ridges or elongated columns oriented similar to that of windows in ELO, conditions are provided such that the side-walls of the columns provide the crystallographic template for lateral growth.

As an example, Fig. 30.43 shows the geometry of a silicon substrate used for the LPE lateral overgrowth of Si from ridge seeds [30.17,96]. A thin SiO_2 mask has been deposited in the trenches to avoid direct epitaxy on the substrate. During LPE, the growth starts from the sidewalls of the ridges and proceeds laterally over the mask. The top surface of the ridge is the dislocation-free Si(111) facet, so it contains no step sources. Therefore, under proper growth conditions the vertical growth is completely eliminated and the layer grows laterally without any change of its thickness. As the thickness of the layer is determined by the ridge height, it can be easily controlled. However, very thin layers are difficult to obtain. When the thickness of the layer decreases the radius of curvature of its side-wall decreases as well (Fig. 30.43). This leads to an increase in the equilibrium concentration of the solute around this wall due to the Gibbs–Thomson effect (compare discussion in Sect. 30.2.2). Therefore, a higher melt supersaturation must be used to grow thinner layers. Then, however, the vertical growth may be initiated if the supersaturation is large enough for two-dimensional nucleation at the top surface of the ridge. Using this technique 18.8 μm wide and as thin as 0.23 μm, i. e., with L/h ratio up to 82, Si layers have been grown laterally upon SiO_2-coated Si substrates by LPE [30.96].

In contrast with the case of homoepitaxy, in lattice-mismatched heterostructures the top surface of the columns is not dislocation free and must be covered by an additional mask to avoid excessive vertical growth. Figure 30.44a shows schematically how this approach has been applied by *Iida* et al. to grow by LPE InGaAs bridge layers on GaAs substrates [30.21]. First, the substrate covered by the SiN mask has been processed by photolithography and etching to fabricate on its surface a pattern of deep circular trenches. Then, an $In_xGa_{1-x}As$ ($x \approx 0.06$) layer has been grown on such substrate by LPE. It is worth noting that no InGaAs buffer is used in this case. The layer grew inwards from the side-wall of the trench, forming a bridge over the

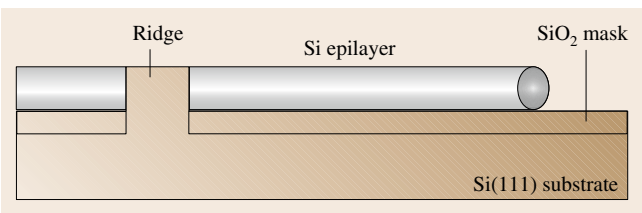


Fig. 30.43 Schematic illustration of the cross section of a Si epilayer grown by LPE on a SiO_2-masked Si substrate from a ridge seed

trench. Some growth over the mask outside the trench area was observed as well. Analysis of the layer surface revealed presence of dislocated zone with etch pits density $\approx 10^6\,\mathrm{cm}^{-2}$ only above the small area where the layer was in direct contact with the GaAs substrate (Fig. 30.44b). The rest of the layer, having diameter above 1 mm, was of very high quality with dislocation density below $10^4\,\mathrm{cm}^{-2}$ (Fig. 30.44c), so it could be used as substrate for further deposition of device layers. The same authors have tested the bridge growth from maskless trenches. Then the lateral growth from the trench wall shown in Fig. 30.44a is accompanied by additional growth of InGaAs from the trench bottom. As a result, new dislocations are found in the InGaAs bridge layer at the places where the heavily dislocated material growing from the trench bottom comes in contact with the bridge. If GaAs substrates are used the presented technique allows for growth of $In_xGa_{1-x}As$ layers with very low In concentrations ($x \approx 0.06$). For a higher indium content ($x \approx 0.20$), InAs wafer converted into InGaAs instead of GaAs has been used as the substrate [30.97]. The technology of LPE growth of bridge layers over the trenches seems to be very promising due to its simplicity, very high efficiency, and high quality of layers produced.

The method of lateral overgrowth that has received much attention recently is pendeo- (from the Latin: to *hang* or to be *suspended from*) epitaxy (PE). To grow PE GaN layers, a thin AlN wetting layer and GaN buffer are grown first on an available substrate as shown in Fig. 30.45a. Next, trenches deep enough to reach the substrate surface (or even slightly deeper) are etched in the structure. The bottoms of trenches and top surfaces of GaN columns are then covered by a suitable masking film, so the GaN column side-walls are the only surfaces exposed for subsequent GaN regrowth. During pendeo-epitaxy, the GaN layers overflow the trenches growing laterally from the column side-walls (Fig. 30.45b). Also some vertical growth takes place followed by lateral growth over the mask covering the top of GaN mesas. After sufficiently long growth time the PE layers starting from adjacent seeds merge at the fronts marked A and B in Fig. 30.45c, so a continuous PE GaN layer is formed. As mentioned earlier, dislocations thread in the GaN buffer mainly along the *c*-axis (i. e., perpendicular to the surface). Thus, only a very small part of them intersect the side-wall of GaN columns and propagate into the PE GaN layer. Therefore, the regrown material contains four to five orders of magnitude lower density of dislocations than that in the buffer [30.98]. Moreover, in contrast to the standard ELO case, also the dislocations originally present in the buffer (marked as TDs in Fig. 30.45) are prevented from reaching the PE layer if the GaN columns are capped by the mask. Thus, in principle, the density of dislocations should be reduced over the whole area of the substrate and the need to locate devices only in the laterally overgrown parts of the structures (as was in the case of ELO; see Sect. 30.3) could be eliminated. The PE technique in its version shown in Fig. 30.45 has been used by *Chen* et al. to grow by MOVPE GaN layers on SiO_2-masked sapphire substrates [30.99]. Studies of the layers with the use of transmission electron microscopy have proved that laterally overgrown parts of the layers are nearly dislo-

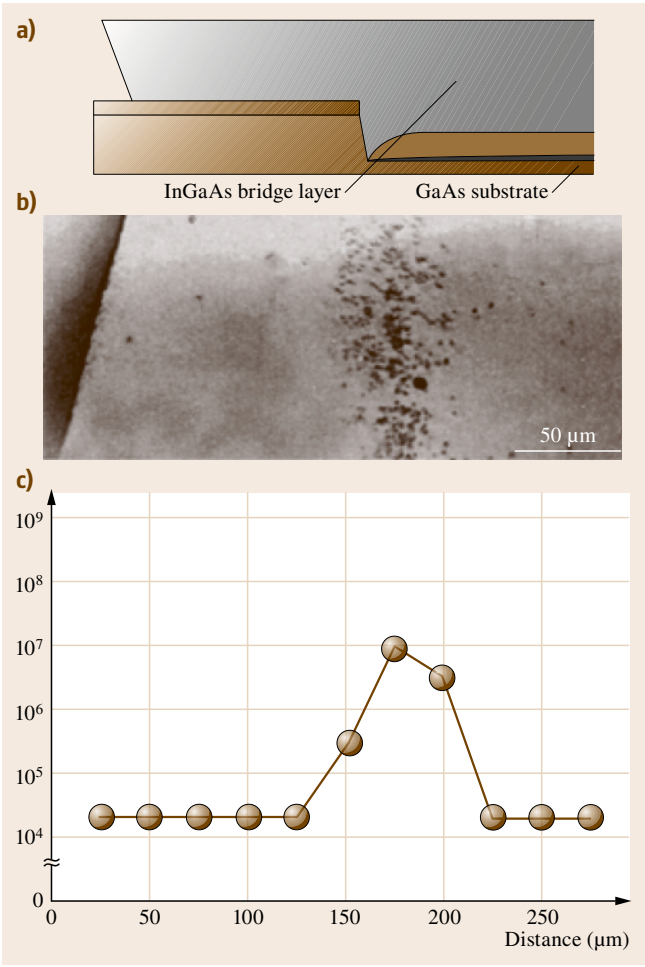


Fig. 30.44a–c Schematic illustrations of the layer cross section **(a)**, photographs of the etched surface **(b)**, and the profiles of etch pits density **(c)** in an InGaAs bridge layer grown by LPE on a SiO_2-masked (111) GaAs substrate (courtesy of Hayakawa)

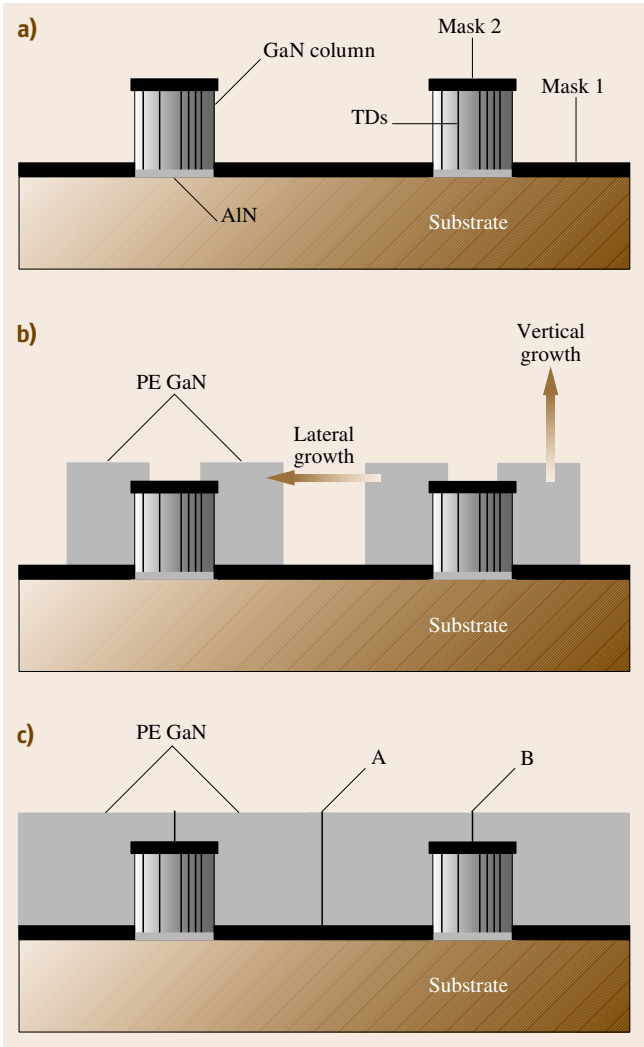


Fig. 30.45a–c Schematic of the process steps for pendeo-epitaxy of GaN: **(a)** etching of elongated columns in GaN buffer layer followed by masking of the substrate (mask 1) and the top column surfaces (mask 2); **(b)** epitaxial growth of GaN starts from exposed side-walls of GaN columns and proceeds laterally to fill the trenches; also some vertical growth occurs; **(c)** PE layers merge over the trenches (seam A) and over the capped GaN columns (seam B), so continuous PE GaN film is formed. TDs are the dislocations threading to the surface of the GaN column

cation free. However, new dislocations have been found at the fronts where the layers merge (marked A and B in Fig. 30.45c) as well as over the edges of masked GaN columns. In the light of what has been written in Sect. 30.4 above, this finding is not surprising. Most probably, interaction of the laterally overgrown parts of the layers with the SiO_2 mask has occurred, leading to their relative misorientation and defect creation above the mask edges and at the coalescence fronts. Thus, this is the same phenomenon as we have discussed already.

These results, as well as the whole discussion in Sect. 30.4, clearly show that properties of laterally overgrown layers are strongly influenced by the presence of a mask. In particular, the strain introduced by the mask and poisoning by impurities diffusing out from the mask at high temperature have negative impact on quality of the layers. Therefore, processes are needed in which layers can be grown laterally without any masking film. It is important to note that each mask requires additional technological steps. Thus, a maskless version of selective epitaxy would result in significant simplification of the whole procedure, which is especially important if the growth process under development is to be applied on industrial scale. *Zheleva* et al. have found that under proper MOVPE conditions neither GaN nor free Ga accumulate on 6H-SiC (0001) surface [30.98]. Thus, pendeo-epitaxy of GaN without a mask between the GaN columns (mask 1 in Fig. 30.45) became possible by replacement of sapphire by SiC. As before, PE growth started from the side-walls of the GaN columns and proceeded laterally at some distance from the SiC surface. Transmission electron microscopy showed that neither tilting nor low-angle tilt boundaries were present at the coalescence boundary over the trench (seam A in Fig. 30.45c). However, tilting of 0.2° and defects at seam B were still found in the portion of the coalesced GaN layer that interacted with the mask covering the GaN column [30.100]. This tilt can be further eliminated if a completely mask-free PE process (i. e., without masks 1 and 2) is employed, as has been reported for GaN [30.98] and AlGaN [30.101] on SiC systems. Then, defect-free coalescence can be obtained at both coalescence fronts [30.98]. The price paid for that, however, is enhanced vertical growth of the PE layer and free propagation into regrown GaN of threading dislocations originally present in the GaN buffer.

The next step towards commercialization of pendeo-epitaxy of GaN would be to replace SiC substrates, which are expensive and available with low diameters, by large-area silicon wafers. However, direct epitaxy of GaN on Si is difficult and usually results in polycrystalline films, most likely due to the prior formation of SiN_x film on the Si surface. Therefore, *Davis* et al. have employed a procedure in which Si(111) substrates

are first covered by an epitaxial film of SiC, followed by their transfer to the MOVPE system for deposition of an AlN wetting layer and a GaN buffer [30.100]. Next, they could make use of the process route elaborated earlier for PE of GaN on bulk SiC, namely, substrate etching to define GaN columns and PE regrowth. The quality of the GaN layers they obtained was quite similar to that on bulk SiC substrates. The problem, however, was the thermal stress arising due to the large difference of thermal expansion coefficients between subsequent layers, which resulted in cracking of the structures on cooling if the SiC transition layer was too thin.

As the final example of new developments in lateral overgrowth, successful liquid-phase growth of GaN layers will be briefly presented. Solution growth is preferred for ELO of GaN since supersaturation at the crystal surface can be kept lower than that during vapor-phase epitaxy. Moreover, the growth velocity along the *c*-axis of GaN is usually much smaller than in perpendicular directions [30.102], so fast lateral growth is expected if (0001)-oriented substrates are used. Unfortunately, as discussed in the Sect. 30.5.1, mask materials that survive undamaged extreme conditions of GaN growth from liquid solution are difficult to find. Therefore, for ELO of GaN *Bockowski* with coworkers [30.103] have used unmasked sapphire substrate with GaN ridges cut from 3 μm-thick GaN buffer grown by MOVPE (Fig. 30.46a). Basically, this is the maskless version of the substrate drawn schematically in Fig. 30.45a. Figure 30.46b shows a plane view of the epilayer grown on such substrate from Ga-N solution at 1430 °C and nitrogen pressure of 1 GPa. The GaN stripes increased their width from 20 to 60 μm and grew ≈ 20 μm high during 5 h-long crystallization. Unfortunately, simultaneous direct crystallization of GaN on unprotected sapphire took place too, so small GaN grains are visible between the stripes. These grains disturbed lateral growth by consuming solute arriving from the bulk solution. Moreover, dislocations were generated in the layer when grains came into contact with laterally grown GaN, in the similar way to that observed in InGaAs bridge layers (compare Sect. 30.5.1). As the result, quite high dislocation density of $\approx 10^6\,\mathrm{cm}^{-2}$ was found in the GaN wing area, though it was much smaller than the $\approx 10^8\,\mathrm{cm}^{-2}$ dislocation density in the areas above the GaN ridges. It has been expected that parasitic GaN nucleation on bare sapphire might be reduced by decreasing the distance between neighboring GaN ridges [30.92]. It is not clear if this has really happened since decrease of ridges spacing from 300 to 60 μm did not lead to a significant reduction of dislocation density in laterally grown GaN.

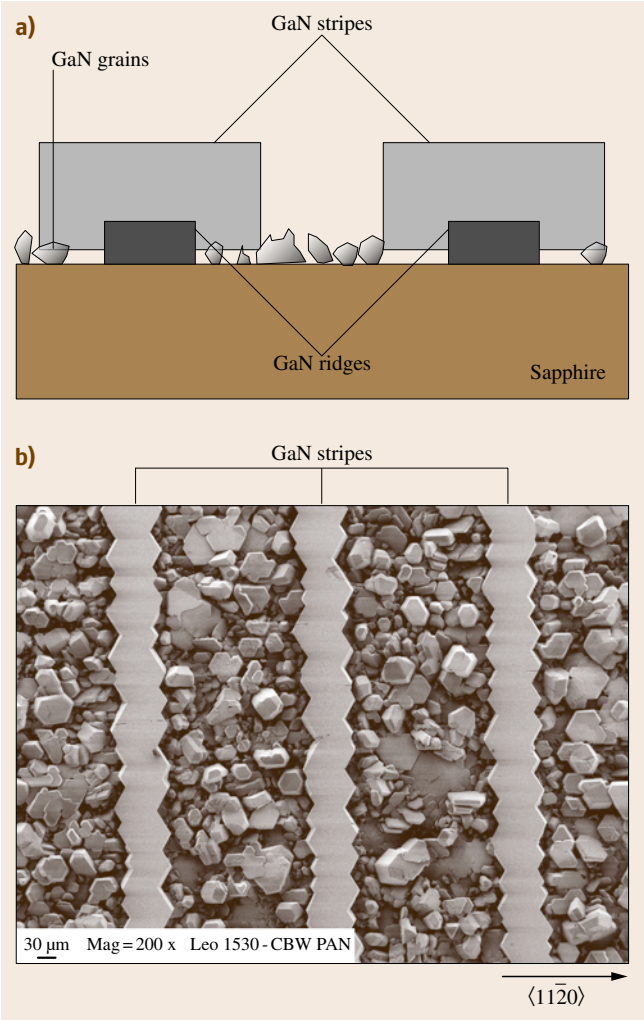


Fig. 30.46a,b Schematic cross section (not to scale) **(a)** and plane-view SEM image **(b)** of GaN layers grown from the Ga-N solution at high nitrogen pressure on sapphire substrate with 20 μm-wide 3 μm-high GaN ridges cut from 3 μm-thick MOVPE-grown GaN buffer. Spacing between the ridges was 300 μm (courtesy of Boćkowski)

Another mode of lateral growth of GaN from Ga-N solution at high nitrogen pressure was found when ≈ 800 μm-thick free-standing GaN substrate grown by HVPE was patterned to create on its surface 5–10 μm-high GaN ridges [30.104]. During GaN growth at 1425 °C and nitrogen pressure of 1 GPa some growth took place directly from the unprotected HVPE sub-

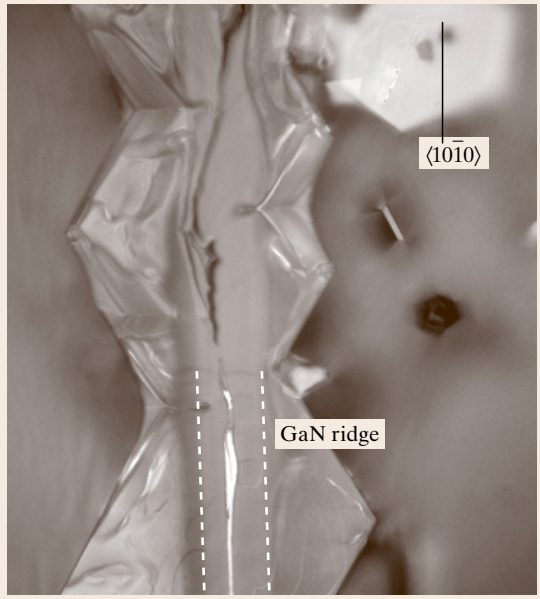


Fig. 30.47 SEM image of GaN layers grown from Ga-N solution at high nitrogen pressure on free-standing HVPE GaN substrate with 20 μm-wide 10 μm-high GaN ridges (*dashed lines*). Spacing between the ridges was 300 μm (courtesy of Boćkowski) ▶

strate as in the previous case. However, due to the specific distribution of solution supersaturation at the substrate surface, preferential lateral overgrowth from ridge edges was also observed. A plan view of such a sample is shown in Fig. 30.47. An important finding by Bockowski et al. was that laterally overgrown parts hang over the substrate if the ridge height was increased to 10 μm. This resulted in their better crystallographic quality. Indeed, analysis of the sample by defect-selective etching has shown that ELO wing areas were dislocation free.

These initial results obtained in lateral overgrowth of GaN from high-temperature Ga-N solution at high nitrogen pressure are very promising. Note that the widths of GaN wings are much larger than those commonly obtained by vapor growth techniques. However, it seems that for further progress application of a suitable mask will be needed to avoid direct crystallization on the substrate. If such a mask is found, much wider wings with lower dislocation density could be grown. Also, improvement of upper surface morphology is required for free-standing layers, similar to that shown in Fig. 30.47.

30.6 Concluding Remarks

A review of epitaxial lateral overgrowth is provided. The main idea behind lateral overgrowth techniques is selective epitaxy of a homoepitaxial layer on a relaxed buffer having a required lattice parameter. The seeds can be defined by patterning of the mask covering the buffer layer (ELO) or the growth starts from the exposed side-walls of ridges etched in the buffer (PE). Selective nucleation of the layer in the narrow seeds allows information on buffer lattice spacing to pass through, while blocking all (or the main part of) defects created due to lattice misfit between the buffer and its substrate. We have shown numerous examples in which this procedure works efficiently. Usually, practical realization of the idea above depends on the system considered. Quite often specific properties of the materials and growth technique used determine the way in which laterally overgrown structures are produced (*vide* maskless pendeo-epitaxy of GaN on SiC, solution growth of GaN, etc.). There are, however, some basic phenomena typical for all laterally overgrown structures. These include some general rules that should be followed to obtain wide and thin laterally overgrown layers (i. e., choice of optimal direction of the seeds, high quality of the buffer, low growth rate, use of doping to enhance lateral growth, etc.). Efficient filtration of dislocations originally present in the buffer is well evidenced in all laterally overgrown structures. Tilt of wings caused by their interaction with the mask underneath is commonly observed in both ELO and PE layers, independently of the growth technique used. If the tilt angle is small the accompanied strain can be accommodated elastically (e.g., in GaAs/GaAs ELOs). Otherwise, arrays of dislocations creating low-angle boundaries appear above the mask edges (e.g., in GaN ELO structures). Tilting of laterally grown parts of the layer leads to the generation of new defects at the coalescence front where two layers grown from neighboring seeds and having opposite tilt direction merge. Growing the layers standing free above the mask seems to be the best way to improve the quality of the coalescence front, although there are no gen-

eral recipes for how to grow such structures. Thermal strain due to different thermal expansion coefficients of subsequent layers is also commonly found in laterally overgrown heterostructures. However, if the layer is attached to the buffer through the narrow seeds only, this strain is usually much smaller than in reference planar structures.

It is difficult to predict future developments in the ELO technology. From a scientific point of view there are still many questions to be answered. In particular, understanding of the microscopic interaction of ELO wings with the mask, identification of mechanisms leading to bending of dislocations threading in a seed area, and many other issues require further work. Basic crystal growth research including process modeling is needed to increase the width of laterally overgrown layers, improve the quality of coalescence fronts, develop technology for lateral overgrowth of ternary layers on binary substrates, find new mask materials, and simplify the whole growth procedure to meet industrial standards. Application of laterally overgrown layers will depend on progress in these areas. Market demands for novel semiconductor devices will decide if and which techniques of lateral epitaxy will find their permanent position in industrial practice. Nowadays there is great demand for group III nitride substrates, so lateral overgrowth of these compounds is most studied in many laboratories worldwide. It is believed that the importance of lateral overgrowth of *traditional* zincblende III–V ternary substrates will increase in the near future.

References

30.1 T. Kawai, H. Yonezu, Y. Ogasawara, D. Saito, K. Pak: Segregation and interdiffusion of In atoms in GaAs/InAs/GaAs heterostructures, J. Appl. Phys. **74**, 1770–1775 (1993)

30.2 N. Chand, R. People, F.A. Baiocchi, W.K. Wecht, A.Y. Cho: Significant improvement in crystalline quality of molecular beam epitaxially grown GaAs on Si (100) by rapid thermal annealing, Appl. Phys. Lett. **49**, 815–817 (1986)

30.3 N. Gopalakrishnan, K. Baskar, H. Kawanami, I. Sakata: Effects of the low temperature grown buffer layer thickness on the growth of GaAs on Si by MBE, J. Cryst. Growth **250**, 29–33 (2003)

30.4 Y. Honda, Y. Iyechika, T. Maeda, H. Miyake, K. Hiramatsu, H. Sone, N. Sawaki: Crystal orientation fluctuation of epitaxial-lateral-overgrown GaN with W mask and SiO_2 mask observed by transmission electron diffraction and x-ray rocking curves, Jpn. J. Appl. Phys. **38**, L1299–L1302 (1999)

30.5 D. Dobosz, Z.R. Zytkiewicz, T.T. Piotrowski, E. Kaminska, E. Papis, A. Piotrowska: Application of tungsten films for substrate masking in liquid phase epitaxial lateral overgrowth of GaAs, Cryst. Res. Technol. **38**, 297–301 (2003)

30.6 D. Dobosz, K. Golaszewska, Z.R. Zytkiewicz, E. Kaminska, A. Piotrowska, T.T. Piotrowski, A. Barcz, R. Jakiela: Properties of ZrN films as substrate masks in liquid phase epitaxial lateral overgrowth of compound semiconductors, Cryst. Res. Technol. **40**, 492–497 (2005)

30.7 Z.R. Zytkiewicz, J. Domagla, D. Dobosz, J. Bak-Misiuk: Microscopic bending of GaAs layers grown by epitaxial lateral overgrowth, J. Appl. Phys. **84**, 6937–6939 (1998)

30.8 S. Sakawa, T. Nishinaga: Effect of Si doping on epitaxial lateral overgrowth of GaAs on GaAs-coated Si substrate, Jpn. J. Appl. Phys. **31**, L359–L361 (1992)

30.9 L. Jastrzebski, J.F. Corboy, R. Soydan: Issues and problems involved in selective epitaxial growth of silicon for SOI fabrication, J. Electrochem. Soc. **136**, 3506–3513 (1989)

30.10 R. Bergmann, E. Czech, I. Silier, N. Nagel, E. Bauser, H.J. Queisser, R.P. Zingg, B. Hofflinger: MOS transistors with epitaxial Si, laterally grown over SiO_2 by liquid phase epitaxy, Appl. Phys. A **54**, 103–105 (1992)

30.11 H. Asai, S. Adachi, S. Ando, K. Oe: Lateral GaAs growth over tungsten gratings on (001) GaAs substrates by metalorganic chemical vapor deposition and applications to vertical field-effect transistors, J. Appl. Phys. **55**, 3868–3870 (1984)

30.12 J.M. Olchowik, A. Fave, B. Semmache, A. Laugier, A. Kaminski, W. Sadowski: Crystallisation of Si epitaxial lateral overgrowth layers for photovoltaic structures, Proc. 16th Eur. Photovolt. Sol. Energy Conf., Vol. 2 (2000) pp. 1286–1288

30.13 J.J. Pak, G.W. Neudeck, A.E. Kabir, D.W. DeRoo, S.E. Staller, J.H. Logsdon: A new method of forming a thin single-crystal silicon diaphragm using merged epitaxial lateral overgrowth for sensor applications, IEEE Electron. Dev. Lett. **12**, 614–616 (1991)

30.14 G.W. Neudeck, P.J. Schubert, J.L. Glenn Jr., J.A. Friedrich, W.A. Klaasen, R.P. Zingg, J.P. Denton: Three dimensional devices fabricated by silicon epitaxial lateral overgrowth, J. Electron. Mater. **19**, 1111–1117 (1990)

30.15 R.W. McClelland, C.O. Bozler, J.C.C. Fan: A technique for producing epitaxial films on reuseable substrates, Appl. Phys. Lett. **37**, 560–562 (1980)

30.16 B.-Y. Tsaur, R.W. McClelland, J.C.C. Fan, R.P. Gale, J.P. Salerno, B.A. Vojak, C.O. Bozler: Low-dislocation-density GaAs epilayers grown on Ge-coated Si substrates by means of lateral epi-

taxial overgrowth, Appl. Phys. Lett. **41**, 347–349 (1982)
30.17 I. Silier, A. Gutjahr, N. Nagel, P.O. Hansson, E. Czech, M. Konuma, E. Bauser, F. Banhart, R. Kohler, H. Raidt, B. Jenichen: Solution growth of epitaxial semiconductor-on-insulator layers, J. Cryst. Growth **166**, 727–730 (1996)
30.18 T. Nishinaga: Epitaxial lateral overgrowth of III–V compounds for obtaining dislocation free layers, Cryst. Prop. Prep. **31**, 92–99 (1991)
30.19 P.O. Hansson, A. Gustafsson, M. Albrecht, R. Bergmann, H.P. Strunk, E. Bauser: High quality Ge_xSi_{1-x} by heteroepitaxial lateral overgrowth, J. Cryst. Growth **121**, 790–794 (1992)
30.20 Z.R. Zytkiewicz: Laterally overgrown structures as substrates for lattice mismatched epitaxy, Thin Solid Films **412**, 64–75 (2002)
30.21 Y. Hayakawa, S. Iida, T. Sakurai, H. Yanagida, M. Kikuzawa, T. Koyama, M. Kumagawa: Epitaxial lateral overgrowth of InGaAs on patterned GaAs substrates by liquid phase epitaxy, J. Cryst. Growth **169**, 613–620 (1996)
30.22 S. Nakayama, M. Kaneko, S. Aizawa, K. Kashiwa, N.S. Takahashi: InGaP lattice-mismatched LPE growth on GaAs substrates by epitaxial lateral overgrowth technique, J. Cryst. Growth **236**, 132–136 (2002)
30.23 S. Nakamura, M. Senoh, S. Nagahama, N. Iwasa, T. Yamada, T. Matsushita, H. Kiyoku, Y. Sugimoto, T. Kozaki, H. Umemoto, M. Sano, K. Chocho: InGaN/GaN/AlGaN-based laser diodes with modulation-doped strained-layer superlattices, Jpn. J. Appl. Phys. **36**, L1568–L1571 (1997)
30.24 Y.C. Liu, Z.R. Zytkiewicz, S. Dost: Computational analysis of lateral overgrowth of GaAs by liquid-phase epitaxy, J. Cryst. Growth **275**, e953–e957 (2005)
30.25 Z. Yan, S. Naritsuka, T. Nishinaga: Interface supersaturation in microchannel epitaxy of InP, J. Cryst. Growth **203**, 25–30 (1999)
30.26 Z. Yan, S. Naritsuka, T. Nishinaga: Two-dimensional numerical calculation of solute diffusion in microchannel epitaxy of InP, J. Cryst. Growth **209**, 1–7 (2000)
30.27 B. Beaumont, P. Vennegues, P. Gibart: Epitaxial lateral overgrowth of GaN, Phys. Status Solidi (b) **227**, 1–43 (2001)
30.28 R.J. Matyi, H. Shichijo, T.M. Moore, H.-L. Tsai: Microstructural characterization of patterned gallium arsenide grown on ⟨001⟩ silicon substrates, Appl. Phys. Lett. **51**, 18–20 (1987)
30.29 J.T. Torvik, J.I. Pankove, E. Iliopoulos, H.M. Ng, T.D. Moustakas: Optical properties of GaN grown over SiO_2 on SiC substrates by molecular beam epitaxy, Appl. Phys. Lett. **72**, 244–245 (1998)
30.30 G. Bacchin, T. Nishinaga: A new way to achieve both selective and lateral growth by molecular beam epitaxy: Low angle incidence microchannel epitaxy, J. Cryst. Growth **208**, 1–10 (2000)
30.31 S.C. Lee, K.J. Malloy, L.R. Dawson, S.R.J. Brueck: Selective growth and associated faceting and lateral overgrowth of GaAs on a nanoscale limited area bounded by a SiO_2 mask in molecular beam epitaxy, J. Appl. Phys. **92**, 6567–6571 (2002)
30.32 H. Tang, J.A. Bardwell, J.B. Webb, S. Moisa, J. Fraser, S. Rolfe: Selective growth of GaN on a SiC substrate patterned with an AlN seed layer by ammonia molecular-beam epitaxy, Appl. Phys. Lett. **79**, 2764–2766 (2001)
30.33 Z.R. Zytkiewicz: Epitaxial lateral overgrowth of GaAs: Principle and growth mechanism, Cryst. Res. Technol. **34**, 573–582 (1999)
30.34 S. Zhang, T. Nishinaga: Epitaxial lateral overgrowths of GaAs on (001) GaAs substrates by LPE: Growth behavior and mechanism, J. Cryst. Growth **99**, 292–296 (1990)
30.35 S. Naritsuka, T. Nishinaga: Epitaxial lateral overgrowth of InP by liquid phase epitaxy, J. Cryst. Growth **146**, 314–318 (1995)
30.36 R.P. Gale, R.W. McClelland, J.C.C. Fan, C.O. Bozler: Lateral epitaxial overgrowth of GaAs by organometallic chemical vapor deposition, Appl. Phys. Lett. **41**, 545–547 (1982)
30.37 J. Park, P.A. Barnes, C.C. Tin, A. Allerman: Lateral overgrowth and epitaxial lift-off of InP by halide vapor-phase epitaxy, J. Cryst. Growth **187**, 185–193 (1998)
30.38 T. Nishinaga, T. Nakano, S. Zhang: Epitaxial lateral overgrowth of GaAs by LPE, Jpn. J. Appl. Phys. **27**, L964–L967 (1988)
30.39 S. Zhang, T. Nishinaga: LPE lateral overgrowth of GaP, Jpn. J. Appl. Phys. **29**, 545–550 (1990)
30.40 R. Bergmann, E. Bauser, J.H. Werner: Defect-free epitaxial lateral overgrowth of oxidized (111) Si by liquid phase epitaxy, Appl. Phys. Lett. **57**, 351–353 (1990)
30.41 R. Bergmann: Model for defect-free epitaxial lateral overgrowth of Si over SiO_2 by liquid phase epitaxy, J. Cryst. Growth **110**, 823–834 (1991)
30.42 S. Sakawa, T. Nishinaga: Faceting of LPE GaAs grown on a misoriented Si(100) substrate, J. Cryst. Growth **115**, 145–149 (1991)
30.43 D. Dobosz, Z.R. Zytkiewicz: Epitaxial lateral overgrowth of semiconductor structures by liquid phase epitaxy, Int. J. Mater. Prod. Technol. **22**, 50–63 (2005)
30.44 F. Banhart, R. Bergmann, F. Phillipp, E. Bauser: Dislocation generation in silicon grown laterally over SiO_2 by liquid phase epitaxy, Appl. Phys. A **53**, 317–323 (1991)
30.45 Y.S. Chang, S. Naritsuka, T. Nishinaga: Optimization of growth condition for wide dislocation-free GaAs on Si substrate by microchannel epitaxy, J. Cryst. Growth **192**, 18–22 (1998)

30.46 D. Dobosz, Z.R. Zytkiewicz, E. Papis, E. Kaminska, A. Piotrowska: Epitaxial lateral overgrowth of GaSb layers by liquid phase epitaxy, J. Cryst. Growth **253**, 102–106 (2003)

30.47 H. Raidt, R. Kohler, F. Banhart, B. Jenichen, A. Gutjahr, M. Konuma, I. Silier, E. Bauser: Adhesion in growth of defect-free silicon over silicon oxide, J. Appl. Phys. **80**, 4101–4107 (1996)

30.48 Z.R. Zytkiewicz, D. Dobosz, Y.C. Liu, S. Dost: Recent progress in lateral overgrowth of semiconductor structures from the liquid phase, Cryst. Res. Technol. **40**, 321–328 (2005)

30.49 Z.R. Zytkiewicz, D. Dobosz, M. Pawlowska: Epitaxial lateral overgrowth of GaAs: Effect of doping on LPE growth behaviour, Semicond. Sci. Technol. **14**, 465–469 (1999)

30.50 J.C. Brice: *The Growth of Crystals from Liquids* (North-Holland, Amsterdam 1973)

30.51 M. Ohara, R.C. Reid: *Modelling Crystal Growth Rates from Solution* (Prentice-Hall, Englewood Cliffs 1973)

30.52 W.J.P. van Enckevort, A.C.J.F. van den Berg: Impurity blocking of crystal growth: a Monte Carlo study, J. Cryst. Growth **183**, 441–455 (1998)

30.53 D.-K. Kim, B.-T. Lee: Heteroepitaxial growth of GaAs on (100) GaAs and InP by selective liquid phase epitaxy, Jpn. J. Appl. Phys. **33**, 5870–5874 (1994)

30.54 B. Beaumont, S. Haffouz, P. Gibart: Magnesium induced changes in the selective growth of GaN by metalorganic vapor phase epitaxy, Appl. Phys. Lett. **72**, 921–923 (1998)

30.55 Z.R. Zytkiewicz, J. Domagala: Thermal strain in GaAs layers grown by epitaxial lateral overgrowth on Si substrates, Appl. Phys. Lett. **75**, 2749–2751 (1999)

30.56 T. Nishinaga: Microchannel epitaxy: an overview, J. Cryst. Growth **237–239**, 1410–1417 (2002)

30.57 A. Sakai, H. Sunakawa, A. Usui: Transmission electron microscopy of defects in GaN films formed by epitaxial lateral overgrowth, Appl. Phys. Lett. **73**, 481–483 (1998)

30.58 A. Sakai, H. Sunakawa, A. Usui: Defect structure in selectively grown GaN films with low threading dislocation density, Appl. Phys. Lett. **71**, 2259–2261 (1997)

30.59 Z. Yu, M.A.L. Johnson, T. McNulty, J.D. Brown, J.W. Cook, J.F. Schetzina: Study of the epitaxial lateral overgrowth (ELO) process for GaN on sapphire using scanning electron microscopy and monochromatic cathodoluminescence, MRS Internet J. Nitride Semicond. Res. **3**, 6 (1998)

30.60 A. Gustafsson, P.O. Hansson, E. Bauser: Cathodoluminescence from relaxed Ge_xSi_{1-x} grown by heteroepitaxial lateral overgrowth, J. Cryst. Growth **141**, 363–370 (1994)

30.61 Z.I. Kazi, P. Thilakan, T. Egawa, M. Umeno, T. Jimbo: Realization of GaAs/AlGaAs lasers on Si substrates using epitaxial lateral overgrowth by metalorganic chemical vapor deposition, Jpn. J. Appl. Phys., Part 1 **40**, 4903–4906 (2001)

30.62 P. Kozodoy, J.P. Ibbetson, H. Marchand, P.T. Fini, S. Keller, J.S. Speck, S.P. DenBaars, U.K. Mishra: Electrical characterization of GaN p-n junctions with and without threading dislocations, Appl. Phys. Lett. **73**, 975–977 (1998)

30.63 T. Mukai, K. Takekawa, S. Nakamura: InGaN-based blue light-emitting diodes grown on epitaxially laterally overgrown GaN substrates, Jpn. J. Appl. Phys., Part 2 **37**, L839–L841 (1998)

30.64 C. Sasaoka, H. Sunakawa, A. Kimura, M. Nido, A. Usui, A. Sakai: High-quality InGaN MQW on low-dislocation-density GaN substrate grown by hydride vapor-phase epitaxy, J. Cryst. Growth **189/190**, 61–66 (1998)

30.65 G. Parish, S. Keller, P. Kozodoy, J.P. Ibbetson, H. Marchand, P.T. Fini, S.B. Fleischer, S.P. DenBaars, U.K. Mishra, E.J. Tarsa: High-performance (Al,Ga)N-based solar-blind ultraviolet p-i-n detectors on laterally epitaxially overgrown GaN, Appl. Phys. Lett. **75**, 247–249 (1999)

30.66 Z.R. Zytkiewicz, J.Z. Domagala, D. Dobosz, L. Dobaczewski, A. Rocher, C. Clement, J. Crestou: Tilt and dislocations in epitaxial laterally overgrown GaAs layers, J. Appl. Phys. **101**, 013508 (2007)

30.67 N. Nagel, F. Banhart, E. Czech, I. Silier, F. Phillipp, E. Bauser: The coalescence of silicon layers grown over SiO_2 by liquid-phase epitaxy, Appl. Phys. A **57**, 249–254 (1993)

30.68 F. Banhart, N. Nagel, F. Phillipp, E. Czech, I. Silier, E. Bauser: The coalescence of silicon layers grown over SiO_2 by liquid-phase epitaxy, Appl. Phys. A **57**, 441–448 (1993)

30.69 W. Huang, T. Nishinaga, S. Naritsuka: Microchannel epitaxy of GaAs from parallel and nonparallel seeds, Jpn. J. Appl. Phys. **40**, 5373–5376 (2001)

30.70 Z. Yan, Y. Hamaoka, S. Naritsuka, T. Nishinaga: Coalescence in microchannel epitaxy of InP, J. Cryst. Growth **212**, 1–10 (2000)

30.71 Y.-C. Shih, J.-C. Lou, W.O. Oldham: Seam line defects in silicon-on-insulator by merged epitaxial lateral overgrowth, Appl. Phys. Lett. **65**, 1638–1640 (1994)

30.72 Z.R. Zytkiewicz, J. Domagala, D. Dobosz, J. Bak-Misiuk: Strain in GaAs layers grown by liquid phase epitaxial lateral overgrowth, J. Appl. Phys. **86**, 1965–1969 (1999)

30.73 Z.R. Zytkiewicz, J. Domagala, D. Dobosz: Control of adhesion to the mask of epitaxial laterally overgrown GaAs layers, J. Appl. Phys. **90**, 6140–6144 (2001)

30.74 Z.R. Zytkiewicz, J. Domagala, D. Dobosz: Mask-induced strain in GaAs layers grown by liquid phase epitaxial lateral overgrowth, Mater. Res. Soc. Symp. Proc. **570**, 273–278 (1999)

30.75 J.Z. Domagala, A. Czyzak, Z.R. Zytkiewicz: Imaging of strain in laterally overgrown GaAs layers by spatially resolved x-ray diffraction, Appl. Phys. Lett. **90**, 241904 (2007)

30.76 I.-H. Kim, C. Sone, O.-H. Nam, Y.-J. Park, T. Kim: Crystal tilting in GaN grown by pendoepitaxy method on sapphire substrate, Appl. Phys. Lett. **75**, 4109–4111 (1999)
30.77 D. Lubbert, T. Baumbach, P. Mikulik, P. Pernot, L. Helfen, R. Kohler, T.M. Katona, S. Keller, S.P. DenBaars: Local wing tilt analysis of laterally overgrown GaN by x-ray rocking curve imaging, J. Phys. D Appl. Phys. **38**, A50–A54 (2005)
30.78 R. Rantamaki, T. Tuomi, Z.R. Zytkiewicz, J. Domagala, P.J. McNally, A.N. Danilewsky: Synchrotron x-ray topographic and high-resolution diffraction analysis of mask-induced strain in epitaxial laterally overgrown GaAs layers, J. Appl. Phys. **86**, 4298–4303 (1999)
30.79 R. Rantamaki, T. Tuomi, Z.R. Zytkiewicz, J. Domagala, P.J. McNally, A.N. Danilewsky: Comparative analysis of synchrotron x-ray transmission and reflection topography techniques applied to epitaxial laterally overgrown GaAs layers, J. X-ray Sci. Technol. **8**, 277–288 (2000)
30.80 P. Fini, H. Marchand, J.P. Ibbetson, S.P. DenBaars, U.K. Mishra, J.S. Speck: Determination of tilt in the lateral epitaxial overgrowth of GaN using x-ray diffraction, J. Cryst. Growth **209**, 581–590 (2000)
30.81 M.H. Kim, Y. Choi, J. Yi, M. Yang, J. Jeon, S. Khym, S.J. Leem: Reduction in crystallographic tilting of lateral epitaxial overgrown GaN by removal of oxide mask, Appl. Phys. Lett. **79**, 1619–1621 (2001)
30.82 K. Hiramatsu, H. Matsushima, T. Shibata, Y. Kawagachi, N. Sawaki: Selective area growth and epitaxial lateral overgrowth of GaN by metalorganic vapor phase epitaxy and hydride vapor phase epitaxy, Mater. Sci. Eng. B **59**, 104–111 (1999)
30.83 P. Fini, A. Munkholm, C. Thompson, G.B. Stephenson, J.A. Eastman, M.V. Ramana Murthy, O. Auciello, L. Zhao, S.P. DenBaars, J.S. Speck: In situ, real-time measurement of wing tilt during lateral epitaxial overgrowth of GaN, Appl. Phys. Lett. **76**, 3893–3895 (2000)
30.84 R. Kohler, B. Jenichen, H. Raidt, E. Bauser, N. Nagel: Vertical stress in liquid-phase epitaxy Si layers on SiO_2/Si evaluated by x-ray double-crystal topography, J. Phys. D Appl. Phys. **28**, A50–A55 (1995)
30.85 I. Kidoguch, A. Ishibashi, G. Sugahara, Y. Ban: Air-bridged lateral epitaxial overgrowth of GaN thin films, Appl. Phys. Lett. **76**, 3768–3770 (2000)
30.86 P. Fini, L. Zhao, B. Morgan, M. Hansen, H. Marchand, J.P. Ibbetson, S.P. DenBaars, U.K. Mishra, J.S. Speck: High-quality coalescence of laterally overgrown GaN stripes on GaN/sapphire seed layers, Appl. Phys. Lett. **75**, 1706–1708 (1999)
30.87 S.F. Fang, K. Adomi, S. Iyer, H. Morkoc, H. Zabel, C. Choi, N. Otsuka: Gallium arsenide and other compound semiconductors on silicon, J. Appl. Phys. **68**, R31–R58 (1990), and references therein
30.88 N. Lucas, H. Zabel, H. Morkoc, H. Unlu: Anisotropy of thermal expansion of GaAs on Si(001), Appl. Phys. Lett. **52**, 2117–2119 (1988)
30.89 T.S. Zheleva, W.M. Asmawi, K.A. Jones: Pendeoepitaxy versus lateral epitaxial overgrowth of GaN: A comparative study via finite element analysis, Phys. Status Solidi (a) **176**, 545–551 (1999)
30.90 E. Kaminska, A. Piotrowska, E. Mizera, M. Guziewicz, A. Barcz, E. Dynowska, S. Kwiatkowski: Rapid thermal nitridation of tungsten-capped shallow ohmic contacts to GaAs, Mater. Res. Soc. Symp. Proc. **337**, 349–354 (1994)
30.91 L.E. Toth: *Transition Metals Carbides and Nitrides* (Academic, New York 1971)
30.92 M. Boćkowski, I. Grzegory, G. Nowak, G. Kamler, B. Łucznik, M. Wróblewski, P. Kwiatkowski, K. Jasik, S. Krukowski, S. Porowski: Growth of GaN on patterned GaN/sapphire substrates with various metallic masks by high pressure solution method, Proc. SPIE **6121**, 612103–1–612103–9 (2006)
30.93 T. Bryskiewicz: Liquid phase electroepitaxy of semiconductor compounds, Prog. Cryst. Growth Charact. **12**, 29–43 (1986), and references therein
30.94 Y.C. Liu, Z.R. Zytkiewicz, S. Dost: A model for epitaxial lateral overgrowth of GaAs by liquid-phase electroepitaxy, J. Cryst. Growth **265**, 341–350 (2004)
30.95 M.G. Mauk, J.P. Curran: Electro-epitaxial lateral overgrowth of silicon from liquid-metal solutions, J. Cryst. Growth **225**, 348–353 (2001)
30.96 Y. Suzuki, T. Nishinaga: Si LPE lateral overgrowth from a ridge seed, Jpn. J. Appl. Phys. **29**, 2685–2689 (1990)
30.97 K. Balakrishnan, S. Iida, M. Kumagawa, Y. Hayakawa: A novel method to grow high quality $In_{1-x}Ga_xAs$ ELO and bridge layers with high indium compositions, J. Cryst. Growth **237–239**, 1525–1530 (2002)
30.98 T.S. Zheleva, S.A. Smith, D.B. Thomson, K.J. Linthicum, P. Rajagopal, R.F. Davis: Pendeoepitaxy: A new approach for lateral growth of gallium nitride films, J. Electron. Mater. **28**, L5–L8 (1999)
30.99 Y. Chen, R. Schneider, S.Y. Wang, R.S. Kern, C.H. Chen, C.P. Kou: Dislocation reduction in GaN thin films via lateral overgrowth from trenches, Appl. Phys. Lett. **75**, 2062–2063 (1999)
30.100 R.F. Davis, T. Gehrke, K.J. Linthicum, E. Preble, P. Rajagopal, C. Ronning, C. Zorman, M. Mehregany: Conventional and pendeo-epitaxial growth of GaN(0001) thin films on Si(111) substrates, J. Cryst. Growth **231**, 335–341 (2001)
30.101 R.F. Davis, T. Gehrke, K.J. Linthicum, T.S. Zheleva, E.A. Preble, P. Rajagopal, C.A. Zorman, M. Mehregany: Pendeo-epitaxial growth of thin films of gallium nitride and related materials and their

characterization, J. Cryst. Growth **225**, 134–140 (2001)

30.102 I. Grzegory: High pressure growth of bulk GaN from solutions in gallium, J. Phys. Condens. Matter **13**, 6875–6892 (2001)

30.103 M. Boćkowski, I. Grzegory, J. Borysiuk, G. Kamler, B. Łucznik, M. Wróblewski, P. Kwiatkowski, K. Jasik, S. Krukowski, S. Porowski: Growth of GaN on patterned GaN/sapphire substrates by high pressure solution method, J. Cryst. Growth **281**, 11–16 (2005)

30.104 M. Boćkowski, I. Grzegory, G. Nowak, B. Łucznik, B. Pastuszka, G. Klamer, M. Wróblewski, P. Kwiatkowski, K. Jasik, W. Wawer, S. Krukowski, S. Porowski: Growth of GaN on patterned thick HVPE free standing GaN substrates by high pressure solution method, Phys. Status Solidi (c) **3**, 1487–1490 (2006)

31. Liquid-Phase Epitaxy of Advanced Materials

Christine F. Klemenz Rivenbark

The performance of many electronic and optoelectronic devices critically depends on the structural quality and homogeneity of the base material, which is often an epitaxial film grown by either vapor-phase epitaxy (VPE) or liquid-phase epitaxy (LPE).

This chapter presents the state of the art in LPE growth of selected advanced materials:

1. High-temperature superconductors
2. Calcium gallium germanates (langasite-type materials)
3. III–V wide-bandgap nitrides.

It is not the aim, herein, to present LPE growth of more *traditional* III–V semiconductors (Si, Ge, GaAs, GaP, InP, GaP) and garnets, which have already been described extensively in the literature since about 1960. Instead, some of the most relevant literature references are given in the historical overview, which also provides a very good insight into the potential of LPE growth of newer materials. Despite the fact that LPE growth has gained less attention over the past decades, mainly due to the development of VPE growth techniques, there is a silver lining which clearly indicates that the highest-quality epitaxial films, for most efficient electronic and optoelectronic devices, will ultimately be achieved from liquid-assisted or LPE-grown films.

31.1 **Historical Development of LPE** 1042
31.2 **Fundamentals of LPE and Solution Growth** 1042
31.3 **Requirements for Liquid-Phase Epitaxy** .. 1044
31.4 **Developing New Materials: On the Choice of the Epitaxial Deposition Method** 1044
31.5 **LPE of High-Temperature Superconductors** ... 1046
31.5.1 Phase Relations, Solvent System, and Solubility Curves 1046
31.5.2 Heat of Solution 1048
31.5.3 Supersaturation and Driving Force for LPE 1048
31.5.4 Substrates and Epitaxial Relationship 1049
31.5.5 LPE Growth System and Film Growth Procedure 1051
31.5.6 Growth Mechanisms and Growth Parameters: Theory Versus Experiment 1052
31.6 **LPE of Calcium Gallium Germanates** 1055
31.6.1 Solvent System 1055
31.6.2 Substrates for Homoepitaxial LGT LPE Film Growth 1056
31.6.3 LPE growth of LGS, LGT, and LGN 1057
31.6.4 Structural and Chemical Characterization of Doped LGT LPE Films 1057
31.7 **Liquid-Phase Epitaxy of Nitrides** 1059
31.7.1 Developments and Trends in LPE of GaN and AlN 1060
31.7.2 Substrates for Epitaxy of Nitrides 1061
31.7.3 Growth System and Optimization ... 1062
31.7.4 Morphological Evolution of LPE-Grown Nitride Films 1062
31.8 **Conclusions** 1063
References 1064

Liquid-phase epitaxy (LPE) is a growth process where a film of crystalline material is deposited from a supersaturated solution onto a single-crystal substrate. This method was first described by *Nelson* in 1963 [31.1], who used a tipping LPE apparatus for the fabrication of GaAs laser diodes. Since then, the LPE tech-

nique has been widely investigated and applied to a variety of materials, among others II–V and III–V semiconductors, carbides, nitrides, and a variety of oxides.

31.1 Historical Development of LPE

Liquid-phase epitaxy has traditionally been the optimal method of choice for the synthesis of highest-quality material. However, this method present also challenges. Today, more *flexible* vapor growth techniques are generally preferred, especially when film defects are not a limiting factor for a given device application.

The potential of LPE to yield low-defect material has been documented in many instances, following the path of the development of technologically important materials. In LPE, the driving force (or supersaturation) for epitaxy is very low. Hence, growth occurs near to thermodynamic equilibrium. The atoms can migrate efficiently on the growing surface until they find the most energetically favorable position where they will incorporate. If supersaturation and other requirements are sufficiently met, the LPE growth process may result in single-crystalline films with very high homogeneity and purity. This was best demonstrated with the development of light-emitting diodes (LEDs), with compound semiconductors such as GaAs [31.2] and GaP [31.3]. Films grown by LPE contain inherently a very low concentration of point defects, which is often orders of magnitude lower than material grown by any other growth technique. Most relevant evidence is found in the research and development of green GaP LEDs. Due to an extremely low concentration of nonstoichiometric point defects in GaP LPE films, even at room temperature, the dominant emission transition mechanism is free exciton recombination. This enables a highly efficient nitrogen-free pure-green light emission (550 nm) [31.4] that cannot be obtained from GaP produced by other growth techniques.

As early as in 1974, studies on selective LPE growth of GaAlAs led to the discovery of the epitaxial lateral overgrowth (ELO) technique [31.5]. The combination of ELO and LPE [31.6] enabled dislocation-free films of many semiconductors, such as Si, GaAs, GaP, GaAs/Si, and InP [31.7–10]. In view of this, one may wonder why LPE growth is only marginally studied today. For example, research on wide-bandgap nitride semiconductors could highly profit from dislocation-free LPE films to be used as substrates. Often, the potential of a growth technique is evaluated from its commercialization perspective. Without contest, the broadest commercialization to report for films grown by LPE resides in the field of optoelectronic devices. Even today, the highest-efficiency red GaAs LEDs are commercially produced by LPE. Films grown by LPE are also used in other applications, such as γ-ray [31.11] and high-end infrared (IR) detectors based on CdHgTe/CdTe [31.12], for example, where low-defect films are required.

The LPE growth technique has been applied to oxides as early as 1968 by *Linares* [31.13], who was able to demonstrate high-quality yttrium iron garnet (YIG) films for the first time. Different magnetic garnet compositions were studied for their use in bubble domain devices, and this led to extensive studies on many fundamental and experimental aspects of LPE of oxides, including the development of different LPE growth techniques, which are summarized in excellent review papers [31.14–16]. The LPE growth method has been successfully applied for the growth of many oxides, among which one can also cite high-temperature superconductors (HTSCs), optical, nonlinear optical, and electrooptical materials, and ferroelectric and dielectric materials.

31.2 Fundamentals of LPE and Solution Growth

The fundamental mechanisms for LPE growth are similar to the growth of free-standing crystals from a solution (or flux), or to the growth of larger crystals by top-seeded solution growth (TSSG). Usually, the same solution can be used for the three methods. In growth from solutions, a supersaturation is created either by cooling or by solvent evaporation of the saturated solution. Another method that can be used is the transport of the growth species through a temperature gradient in the solution. In each case, the growth is controlled by the

diffusion of growth species towards the growing interface and the counterdiffusion of solvent species towards the bulk of the solution.

In the initial stages of crystallization from solutions, some degree of supersaturation, or driving force, has to be established for nucleation to take place. A crystal can only develop once a nuclei has reached a critical size and thus is stable. In the phase diagram, the region between the liquidus temperature and the temperature at which nucleation proceeds spontaneously is called the metastable or Ostwald–Miers region. Within this region, heteronucleation is possible, for example, on the crucible, on a substrate or seed, or on an impurity. The width of the metastable region can vary greatly for different solvent–solute systems. It depends on many factors, such as the nature and complexity (number of components) of the crystallizing material, the viscosity of the solution, additives/dopants, and hydrodynamic flows. For oxides, undercoolings up to 150 °C are possible [31.17], whereas metallic semiconductor solutions allow only a few degrees or less of undercooling. When a seed or a substrate is used, the initial supersaturation required for crystallization is less than that required for spontaneous and free nucleation in the solution. This means that, in LPE, nucleation starts within a range between the liquidus temperature and the temperature for spontaneous homogeneous nucleation.

The relative supersaturation is expressed by

$$\sigma = \frac{x - x_e}{x_e} , \tag{31.1}$$

where x and x_e represent the actual and equilibrium concentration of the solution.

The undercooling $\Delta T = T_e - T$ is related to the respective concentrations through the specific phase diagram of the system considered.

During the LPE film growth, a boundary layer of thickness δ is created at the growth front, in which a temperature and a concentration gradient exist. At the growth interface, two processes are determining: hydrodynamic flow and diffusion. For mass transport, flow is far more effective than diffusion. However, the growth is governed by the slowest process. In addition, the hydrodynamic flow decreases from the bulk of the solution to become negligible at the growth interface. Thus, at the growing interface, within the boundary layer, the process is solely controlled by diffusion, and the thickness of the boundary layer will be determined by the hydrodynamic flow. As a consequence, for a given system at a given supersaturation, the thickness of the boundary layer will ultimately determine the growth rate, which in turn can be adjusted through the substrate rotation rate.

In the case of LPE growth, for systems with relatively large Schmidt number ($\mathrm{Sc} = \nu/D$), the thickness of the diffusion boundary layer on the rotating substrate interface can be estimated from

$$\delta = 1.61 \left(\frac{\nu}{\varpi}\right)^{1/2} \left(\frac{D}{\nu}\right)^{-1/3} , \tag{31.2}$$

where ν is the kinematic viscosity, ω is the rotation rate of the substrate, and D is the diffusion coefficient of the solute [31.18].

In LPE, the growth rate will not only be determined by the mass transport process of the growth species to the growing interface, but also by the incorporation of the solute atoms on the growing surface. Thus, supersaturation and surface kinetics should both be considered, and *Ghez* and *Giess* [31.19] developed a model for the growth of garnets that includes surface kinetic terms. However, in most growth rate studies, it is generally assumed that surface diffusion can be neglected [31.20]. For supersaturations σ below a critical value σ_c, the dependence between growth rate and supersaturation is quadratic, with $r \sim \sigma^2$ for $\sigma < \sigma_c$, and linear above, with $r \sim \sigma$ for $\sigma > \sigma_c$. High growth rates can be achieved when a continuous source of steps is present on the growing surface, for example, when growth spirals are present [31.20, 21].

In solution growth or LPE of complex crystals, the solute usually has a different composition than the solution, and the tendency for the incorporation of specific growth species in the growing crystals is expressed by the distribution coefficient k. The equilibrium distribution coefficient k_0 can be determined from the phase diagram. The effective distribution coefficient describes the concentration ratio between an element in the solution and in the crystal, as $k_{\mathrm{eff}} = c_{\mathrm{crystal}}/c_{\mathrm{solution}}$. The relation describing the dependence of the effective segregation coefficient on the diffusion boundary layer δ and growth rate r has been derived by *Burton* et al. [31.22] as

$$k_{\mathrm{eff}} = \frac{k_0}{k_0 + (1 - k_0) \exp - \left(\frac{cr\delta}{D}\right)} . \tag{31.3}$$

In this equation, c is a constant for the ratio of the mass density between the whole system and the element in the solution. The effective distribution coefficient is an extremely important parameter. If k_{eff} for a certain element is known, it is possible to adjust its concentration in a thin film or crystal, thereby opening the possibil-

ity to *engineer* the properties of a given material, which is important, for example, for the development of laser materials. For $k_{\text{eff}} < 1$, the concentration of the element considered will be smaller in the crystal than in the solution. Hence, this can be used in purification processes by crystallization.

31.3 Requirements for Liquid-Phase Epitaxy

For LPE growth of a new material, the greatest challenge is usually to find a suitable solvent system in which the material crystallizes first and as a single phase. This requires sufficient knowledge of the phase relations in the system considered. Unfortunately, complex phase diagrams are often not available, therefore the first choice is usually made empirically, based on experience with similar material systems.

Besides the solvent, the substrate has to fulfill several requirement, and not only for the epitaxial deposition itself, but there are also requirements depending on foreseen applications. If heteroepitaxy is considered, the growth of layers by LPE is very demanding. Since LPE is a near-equilibrium process, it is very sensitive to all parameters. Basic substrate requirements are low misorientation, low misfit at growth temperature, similar thermal expansion coefficients, and excellent chemical/thermal stability. An important factor for control of the growth mode is the misfit between substrate and film. In the case of LPE of garnets [31.13, 23] it has been demonstrated that the misfit Δa ($= a_{\text{substrate}} - a_{\text{film}}$), with a being the lattice constant, had to be less than 0.01Å in order to avoid cracking. In LPE, there are three major parameters that govern the film growth mode: the supersaturation, the substrate misfit, and the substrate misorientation. Extremely flat surfaces can only be obtained if all three values are extremely low; for example, in the case of c-oriented $YBa_2Cu_3O_{7-x}$ (YBCO) films grown by LPE on (110)$NdGaO_3$, it has been demonstrated that surfaces with interstep distances of 10 µm between monosteps of 12 Å can only be achieved if the following conditions are jointly met: substrate misfit of $\leq 0.08\%$, substrate misorientation of $\leq 0.02\%$, and relative supersaturation of $\leq 0.18\%$ [31.24]. Even if values will vary for other material systems, these values can be taken as order-of-magnitude requirements.

There exist three major methods commonly used for LPE growth: the tipping, dipping, and sliding boat techniques. They have been extensively described in the literature [31.1, 14–16, 23, 25], and will not be reviewed herein, except for the specific materials discussed further below.

31.4 Developing New Materials: On the Choice of the Epitaxial Deposition Method

Deposition methods can be divided into two categories: growth from the vapor phase, and growth from the liquid phase. In vapor growth, or vapor-phase epitaxy (VPE), deposition occurs under vacuum conditions. VPE techniques include radiofrequency (RF) sputtering, laser ablation, thermal and electron-beam evaporation, ion-beam deposition, molecular-beam epitaxy (MBE), atomic layer epitaxy (ALE), chemical vapor deposition (CVD), metalorganic chemical vapor deposition (MOCVD), and plasma-assisted processes. These methods either create a vapor or molecular nutrient by thermal or electron beam evaporation, or the effusion of gas atoms from a Knudsen cell. The growth species are transported to the heated substrate, where nucleation and growth occur. Vapor-phase epitaxy is a very powerful method for the growth of semiconductors. In particular, VPE has the flexibility needed to grow and explore a variety of p/n-multilayer and quantum well structures, and allows the growth and and control of the deposition process down to atomic scale. Thus, over the past decades, VPE growth techniques have constantly improved and become widely applied, whereas interest in LPE growth declined.

The successes of VPE growth techniques, however, resides mainly in the field of conventional semiconductors, such as InGaAsSb and similar systems. Surprisingly, VPE growth of oxides and nitrides exhibits much lower performance, yielding rather defect-rich films that often contain secondary phases and numerous grain boundaries.

In oxides, difficulties typically arise due to the complexity of the material [31.25], the high melting point of constituent oxides, complex phase diagrams, and also often a limited temperature–pressure stability range which restricts the growth conditions that can be applied. Especially important is also the low mobility of growth species under vacuum conditions, which limits surface diffusion of adatoms. Thus, oxide films grown by VPE typically show two-dimensional nucleation of small islands that spread over the growing surface and coalesce. This *birth-and-spread* growth mechanism results in films with numerous grain boundaries, and surfaces with high density of steps with small interstep distances (typically $< 50\,\text{nm}$) [31.26]. Surface nucleation can be suppressed by large misorientation angles of the substrates, thereby providing kinks and steps. However, films grown on misoriented substrates show a very high step density with very small interstep distances, and hence are rough. With increasing thickness of oxide VPE films, a columnar structure often develops, which is not desirable for most applications [31.27]. Similar behavior has been observed for nitride VPE films. In VPE, near-equilibrium conditions may exist very close to the growth steps, and kinetics is the dominant factor that governs the deposition. In contrast, in LPE, the low driving force is maintained over a much larger extensions, and when all substrate and growth parameters are sufficiently met, the layer-by-layer growth mechanism can proceed over macroscopic dimensions. In LPE, extremely flat surfaces may develop, and interstep distances of several microns are not uncommon [31.28, 29].

For the achievement of high structural film perfection, it is essential that growth and decomposition reactions coexist at the growing interface. It is postulated that the high stability of nitrides, combined with a low mobility of growth species under vacuum (VPE) conditions, limits the surface mobility of atoms and thereby also reversible reactions associated with surface diffusion. This could also tentatively explain the experimental limit to the minimum dislocation density observed in vapor-grown nitride films, where dislocations can only be reduced to about $10^5\,\text{cm}^{-2}$ despite the use of ELO techniques [31.30–33]. The nucleation and growth kinetic processes are extremely complex, and the reasons for this experimentally observed limit are not well understood.

The achievable surface flatness, density of defects (point defects, dislocations), and film homogeneity/perfection are three aspects where the LPE technique may surpass VPE, and the choice of the deposition method should also be based on such consideration.

The growth system required for LPE growth (in any configuration) is very inexpensive compared with the equipment required for VPE growth. However, the achievement of high-quality LPE layers of complex oxides, nitrides or carbides can be very challenging, because this requires a multidisciplinary approach and high degree of expertise, especially in solution chemistry and thermodynamics, phase diagrams, and an excellent knowledge of the fundamentals of crystal growth and epitaxy. Like any other growth method, LPE presents advantages and disadvantages. Among the advantages, we can mention:

1. Films with highest structural perfection (low dislocation density, low concentration of point defects)
2. Automatic stoichiometry control
3. Dopant incorporation can be extremely homogeneous
4. Extremely flat surfaces (facets) can develop
5. Upscaling and mass production are possible.

Among the disadvantages, we can mention:

1. Usually thick films: minimum achievable is about 500 nm (atomic layer epitaxy not possible)
2. Not practical for the growth of heterostructures/quantum/multilayers
3. Rough interfaces frequent, due to back-dissolution
4. Very sensitive to substrate surface defects and parameters (misfit, misorientation)
5. Three major conditions have to be satisfied for the highest-quality LPE films:
   - Very small misfit
   - Misorientation
   - Supersaturation.

Only a few aspects were discussed in this introductory section. In the following sections, this paper discusses these challenges more precisely, with examples from recent studies on LPE growth of three different advanced materials, developed for different applications:

1. High-temperature superconductors, for flat (Josephson-junction) tunnel-device technology
2. Nitrides, for optoelectronic and high-power electronic devices
3. Novel frequency-agile piezoelectrics, for high-precision resonators and oscillators.

31.5 LPE of High-Temperature Superconductors

In 1986, the discovery of superconductivity at 35 K in a La-Ba cuprate [31.35], above the 23 K of Nb_3Ge, initiated the search for other cuprates with a higher critical temperature T_c. One year later, *Wu* et al. [31.36] found $YBa_2Cu_3O_7$ (YBCO) with a T_c of 90 K. This superconductivity above the boiling point of liquid nitrogen (77 K) stimulated worldwide research on this new group of superconducting materials (*ceramics*), and numerous applications based on tunneling were envisaged assuming that a timely solution of the material and crystal growth problems could be found. A new technological era seemed to be born, and this discovery initiated a race for the identification of new ceramic compounds with even higher critical temperatures. Among all high-temperature superconductors (HTSC), the so-called 123-cuprates became the most investigated. In particular, $YBa_2Cu_3O_{7-x}$ (YBCO) and $NdBa_2Cu_3O_{7-x}$ (NdBCO) attracted significant attention.

Due to their complexity and limited thermal and chemical stability, the growth process parameters that can be used for the synthesis of HTSCs are limited; for example, YBCO and NdBCO compounds melt incongruently below 1100 °C, and their oxygen content will depend on the growth conditions. In most cases, YBCO is grown as tetragonal $YBa_2Cu_3O_{6.2}$, which has to be oxidized at high temperature to be transformed to the superconducting orthorhombic $YBa_2Cu_3O_{6.93}$ phase with the highest T_c. This structural phase transition results in twinning, which in epitaxial layers depends on the substrate misfit and on the film thickness. Most thin-film growth efforts were concentrated on vapor-phase epitaxy (VPE), with more than 500 groups worldwide exploring chemical and physical vapor deposition of HTSCs. These VPE films typically show two-dimensional (2-D) nucleation and localized step flow, or spiral-island formation [31.37, 38]. Thus, step densities were very high, with interstep distances typically between 10 and 30 nm. For planar tunnel device technology, due to the very short coherence length in HTSCs, extremely flat film surfaces and interfaces were required. Thus, despite the challenges, researchers started to investigate liquid-phase epitaxy (LPE) growth of HTSCs. Early attempts in LPE growth of YBCO, NdBCO, $Bi_2Sr_2CaCu_2O_y$(2212), and Tl(1223)/(1324) can be found in [31.39–52].

31.5.1 Phase Relations, Solvent System, and Solubility Curves

Phase Relations

Knowledge of phase relations and solubilities is of fundamental importance for crystal growth and liquid-phase epitaxy of a given compound. In order to avoid the crystallization of competing secondary phases, the growth has to occur in the primary crystallization field (PCF) of the material to be grown, where it crystallizes

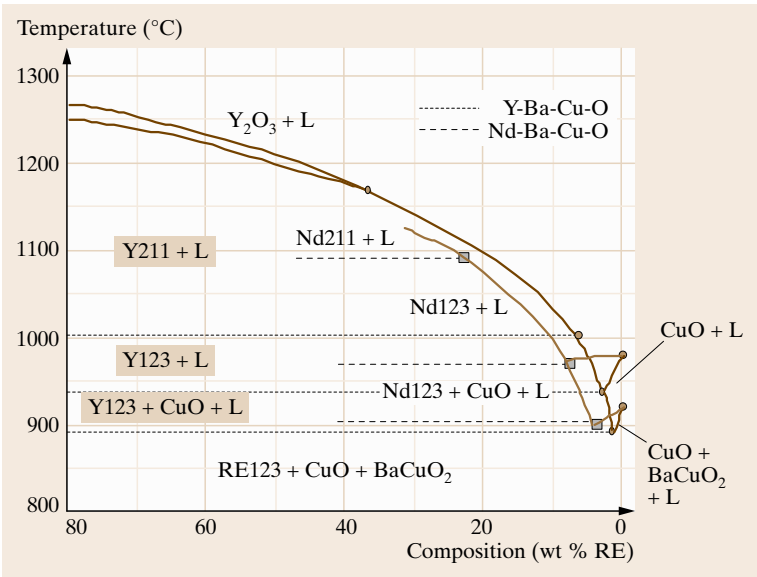


Fig. 31.1 Pseudobinary cut in the phase diagram of YBCO compared with NdBCO (after [31.34])

first, as a single phase, upon cooling. Due to the chemical and structural complexity of YBCO and NdBCO, this turned out to be very challenging [31.53, 54].

YBCO and NdBCO are incongruently melting in air, i. e., they decompose upon melting. Therefore, it is not possible to grow YBCO and NdBCO substrate crystals directly from their melt, for example, by Czochralski or Bridgman techniques. Therefore, they have to be grown from high-temperature solutions.

Solvent System

Finding a suitable solvent for a given crystal may sometimes represent a real challenge. The old concept *similia similibus solvuntur*, (Latin: similar is dissolved in similar) is the basic criterion to be followed. The optimum choice is a solvent which is chemically similar (in the type of bonding) to the solute, but which has sufficient crystal-chemical differences between the solvent and solute constituents in order to prevent the incorporation of solvent species into the solute structure. For example, for the growth of metals, metallic solutions can be used. For high-melting oxide compounds, oxides or fluorides (or a mixture of both) are generally used.

In the case of YBCO and NdBCO, attempts to use solvents traditionally applied to other oxides, for example, lead oxide and lead fluoride, were not successful [31.55]. Hence, a self-flux was chosen. Since the lowest possible growth temperature is desired, a solution composition near the binary eutectic between $BaCuO_2$ and CuO, which is at about 29 ± 1.5 mol % BaO at a temperature of 910 ± 10 °C in air [31.54], is used as solvent.

The two systems, NdBCO and YBCO, are not fundamentally different. Compared with YBCO, NdBCO has two advantages: a higher thermal stability reaching nearly 1100 °C in air, and a much wider PCF allowing concentrations up to approximately 20 wt % for single-phase growth. Figure 31.1 shows a pseudobinary cut in the phase diagrams of YBCO compared with NdBCO [31.34].

Thus, YBCO LPE films (or crystals) can be grown from about 920 to 1000 °C from the BaCuO/CuO eutectic solvent; for NdBCO, temperatures up to 1100 °C can be applied. Crystals and films grown from this solvent are not superconducting, because they are depleted in oxygen. They have to be annealed in oxygen after the growth to become superconducting. This will be discussed in detail below.

Solubility Curves

In order to adjust the supersaturation, knowledge of the solubility of the solute in the solution is essential. In the case of YBCO and NdBCO, solubility studies were performed by observation of the formation of crystals on the top of the solution, and by crystallization experiments on nonpolished seeds dipped into the saturated solution [31.34]. Figure 31.2 shows the solubility data obtained (our work) compared with the data published by other authors.

There is a strong scatter in the published solubility data shown in Fig. 31.2, which can be tentatively explained by the difficulties of HTSCs and the quite different methods, solution compositions, chemicals, and crucibles used by the different authors [31.56, 57].

Observation of the formation and dissolution of crystals on top of a solution is usually not sufficiently accurate for the determination of the solubility curve. Indeed, due to the metastable (Ostwald–Miers) region,

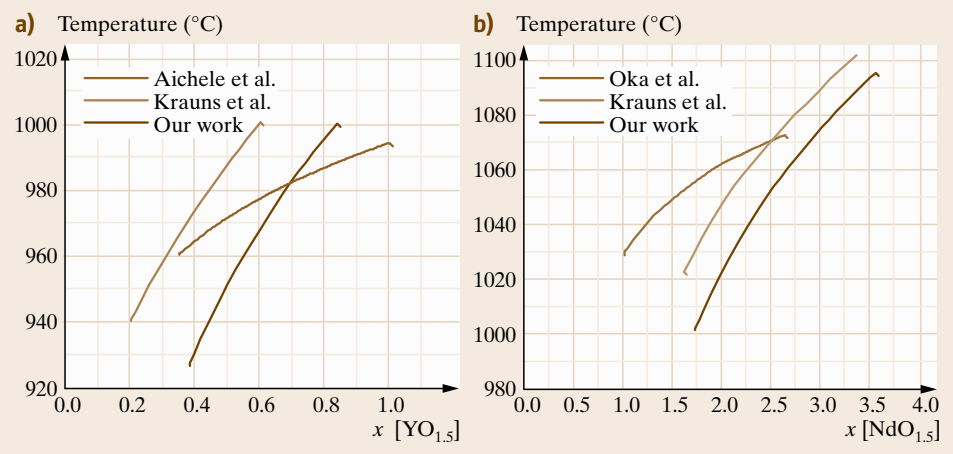


Fig. 31.2a,b Solubility of YBCO (**a**) and NdBCO (**b**) in the BaO/CuO flux at 31 mol %, compared with data published by other authors (after [31.34])

spontaneous crystallization does not necessarily occur at the liquidus temperature, because a stable nucleus of critical size needs to be formed first. Thus, the solution can be supersaturated by several degrees before a visible crystallite is observed. The use of a seed may allow the supersaturation required for nucleation and growth to be reduced, and may thus give a better estimate of the liquidus temperature. This became clear during our investigations, and our values shown in Fig. 31.2 were mainly obtained using seeds, and were further confirmed by the LPE experiments.

31.5.2 Heat of Solution

In an ideal solution there are no attraction forces between the chemical species in solution, and there is no change in internal energy upon mixing. Thus, all the chemical potentials can be expressed as

$$\mu_j = \mu_{0j} + \mathrm{RT} \ln x_j \,, \tag{31.4}$$

where μ_{0j} is a function of temperature and pressure only, x_j is the mole fraction of the component j, and each component's chemical potential is a linear function of $\ln x_j$.

In the case of nonideal solutions, the mole fractions in (31.1) have to be replaced by the activities a_j, which are related to the concentrations by the activity coefficients γ_j according to

$$a_j = \gamma_j x_j \,. \tag{31.5}$$

For YBCO and NdBCO, there exist different possibilities to express the heat of solution. One may use a one-particle model by considering the reaction

$$\mathrm{REBa_2Cu_3O_{7-}}_x(\mathrm{s}) \rightleftharpoons \mathrm{REBa_2Cu_3O_{7-}}_y(\mathrm{l}) \,,$$

with the solid (s), liquid (l), and with RE = Y, Nd.

For an ideal solution, the relationship between the concentration, expressed in mole fraction, x_1 and x_2 at the temperatures T_1 and T_2, respectively, is then given by

$$\ln\left(\frac{x_2}{x_1}\right) = -\frac{\Delta H_{\mathrm{fus}}}{R}\left(\frac{1}{T_2} - \frac{1}{T_1}\right) \,, \tag{31.6}$$

where ΔH_{fus} represents the enthalpy of fusion. For an ideal solution, a linear plot of $\ln(x)$ versus $1/T$ is obtained. This Arrhenius-type relation allows the whole solubility curve to be extrapolated based on the melting point of the solute and a few solubility points. A deviation from linearity would indicate nonideal solution behavior. Since the volume change on melting is small, ΔH_{fus} may be approximated by the heat of solution L_{sol} [31.58, 59]. Then the heat of solution can be related to the saturation concentration of the solute in the solvent by the relationship

$$L_{\mathrm{sol}} = \mathrm{RT}^2 \frac{\mathrm{d} \ln x_{\mathrm{sat}}}{\mathrm{d}T} \tag{31.7}$$

or

$$L_{\mathrm{sol}} = 4.574 T_1 T_2 \frac{\log x_{\mathrm{sat1}} - \log x_{\mathrm{sat2}}}{T_1 - T_2} \,, \tag{31.8}$$

where x_{sat1} and x_{sat2} are the saturation concentrations at temperatures T_1 and T_2.

By using the above expressions and according to our refined experimental solubility curve, for $\mathrm{YBa_2Cu_3O_{7-}}_x$ crystals with $x \approx 0.8$, grown in air ($P_{\mathrm{O_2}} \approx 0.2$ atm) in a flux with Ba-to-Cu ratio of 31 : 69, a value of $L_{\mathrm{sol}} = 34.7$ kcal/mol at 1273 K is obtained [31.34], which is in good agreement with the calculations of *Tsagareisvili* et al. [31.60, 61], who obtained an enthalpy of melting of ΔH_{melt} of 36.44 and 41.03 kcal/mol for $\mathrm{YBa_2Cu_3O_6}$ at 1446 K and $\mathrm{YBa_2Cu_3O_7}$ at 1503 K, respectively.

For $\mathrm{Nd_{1.1}Ba_{1.9}Cu_3O_{7+}}_x$ crystals grown in air, at 1060 °C, from the same flux composition, a heat of solution of $L_{\mathrm{sol}} = 28.1$ kcal/mol at 1333 K [31.34] is obtained.

The heat of solution can also be expressed by using a multiparticle model, similar to the eight-particle model used for garnets [31.62–64]. In this case, it is assumed that the REBCO molecule splits into several particles or molecules upon dissolution according to

$$\mathrm{REBa_2Cu_3O_{7-}}_x(\mathrm{s}) \rightleftharpoons \mathrm{RE^{3+}} + 2\mathrm{Ba^{2+}} + 3\mathrm{Cu^{2+}} + (7-x)\mathrm{O^{2-}} \,,$$

or

$$\mathrm{REBa_2Cu_3O_{7-}}_x(\mathrm{s}) \rightleftharpoons \mathrm{REO_{1.5}(l)} + 2\mathrm{BaO(l)} + 3\mathrm{CuO(l)} \,.$$

31.5.3 Supersaturation and Driving Force for LPE

An important condition to achieve high-quality films and flat surfaces by LPE is the control of supersaturation. Knowledge of the solubility curve of a system allows one to estimate the supersaturation and the driving force for epitaxy for given growth conditions.

Crystals grow when the total free energy of the system can be decreased. Thus, the driving force for

crystallization is the free energy difference

$$\Delta G = \Delta H - T\Delta S \tag{31.9}$$

between the solid and the supersaturated solution. Here, ΔH and ΔS are the differences in enthalpy and entropy of the crystalline and fluid phase, respectively.

For near-equilibrium conditions, we can write

$$\Delta G = \Delta H \left(1 - \frac{T}{T_e}\right) \approx \Delta H \left(\frac{\Delta T}{T_e}\right) , \tag{31.10}$$

with $\Delta T = T_e - T$ being the difference between the equilibrium and actual temperature.

When we cool a solution that is at equilibrium, the concentration of solute x becomes greater than the equilibrium concentration x_e. Then, the free energy difference at the solid–liquid interface can be written as

$$\Delta G = \mathrm{RT} \ln \left(\frac{x}{x_e}\right) = \mathrm{RT} \ln \left(1 + \frac{x - x_e}{x_e}\right) . \tag{31.11}$$

For small values of supersaturation σ, we have

$$\ln \left(1 + \frac{x - x_e}{x_e}\right) \approx \frac{x - x_e}{x_e} = \sigma . \tag{31.12}$$

Substitution into (31.11) yields

$$\Delta G = \mathrm{RT}\sigma . \tag{31.13}$$

and by using (31.10), the following relation is obtained

$$\sigma = \frac{x - x_e}{x_e} = \Delta H \frac{\Delta T}{RT\, T_e} , \tag{31.14}$$

where σ is the relative supersaturation, x and x_e are the actual and equilibrium concentrations, and ΔH is the enthalpy of solution.

Besides σ, other expressions are often used

$$\alpha = \frac{x}{x_e} \quad \text{the supersaturation ratio ,}$$
$$\Delta T = T_e - T \quad \text{the undercooling .}$$

From the solubility curve of YBCO, if we assume an undercooling of $\Delta T = 0.5\,\mathrm{K}$, the Gibbs free energy difference (driving force) would then be $\Delta G(\Delta T = 0.5) = 13.6\,\mathrm{cal/mol}$, and for $\Delta T = 2.5\,\mathrm{K}$, $\Delta G(\Delta T = 2.5) = 68.1\,\mathrm{cal/mol}$.

31.5.4 Substrates and Epitaxial Relationship

For the growth of high-quality LPE films, the substrate has to fulfill several requirements: low misfit at the growth temperature, similar thermal expansion coefficients to film material, high chemical stability in the solvent, and excellent thermal stability, from the film growth to the application temperature. In the case of YBCO and NdBCO, the considered temperature range is from about 1100 °C down to 77 K (liquid nitrogen). Furthermore, in LPE, a very low supersaturation has to be applied during film growth. Since all these requirements have to be satisfied for best results, it is evident that homoepitaxy would represent the best choice. In heteroepitaxy, there is generally no substrate material fulfilling all requirements, and compromises are necessary. Often, buffer layers (also called templates) of the same or a similar film material are used for better lattice matching. In the case of LPE of YBCO and NdBCO, the gallates $LaGaO_3$ (LGO) and $NdGaO_3$ (NGO) were extensively explored as substrate materials, as they present a better lattice match compared with other substrates such as $SrTiO_3$ or MgO. Both gallates and their solid solutions can be grown by Czochralski technique.

Due to the severe misfit requirements in LPE [31.65], precise knowledge of the lattice constants and thermal expansion coefficients of the substrate and HTSC film material is essential [31.66]. High-temperature lattice constant data available in the literature can vary significantly. This is not necessarily due to different determination methods or experimental inaccuracies, but can also be due to the crystal material itself. Indeed, the lattice constants of the same oxide material grown by Czochralski in different laboratories can vary, depending on the starting oxides and growth conditions applied. In addition, radial and longitudinal compositional variations are frequently observed in oxides grown by Czochralski technique. In the case of HTSCs, it was found more accurate to perform high-temperature x-ray measurements on film and specific substrate material used for LPE growth, using the same diffractometer/method [31.66]. For the measurements, single-crystalline film and substrate samples were ground into powder. A Philips Xpert diffractometer with CuK_α-radiation was used for the θ–2θ scans between 20 and 75° in the temperature range of about 35–1000 °C. The Bühler heating chamber HDK was equipped with a tightening tunable electrode for compensation of heat expansion of the Pt-strip heater, keeping the sample in focusing position. The samples were fixed with a Pt/Rh net with mesh size 200 µm × 200 µm, length 20 mm, and thickness 100 µm, attached to the strip heater, reducing existing lateral temperature gradients over the sample. Lattice parameters

were refined by Rietveld analysis. The most challenging aspect during such measurements (and also the greatest source of error) is usually lack of knowledge of the exact temperature of the sample. The use of an internal standard for calibration allows more accurate determinations. The thermal expansion behavior of alumina (Al_2O_3) and its chemical stability is well established. Therefore, Al_2O_3 reference powder was mixed with the gallates for calibration of the temperature [31.66].

The lattice parameters and lattice angles of the orthorhombic, tetragonal, and rhombohedral rare-earth gallates and cuprates (YBCO and NdBCO) can be transformed into a pseudocubic system. The pseudocubic cell of orthorhombic $REGaO_3$ (with RE = Y, Nd) consists of two [110]/2 axes, of which the distance is $0.5(a^2+b^2)^{1/2}$, and the orthogonal axis $c/2$. The pseudocubic angle $\gamma = 2\arctan(a/b) \neq 90°$ exists between the [110] directions. The substrate gallate planes (001) and (110) fit the film lattice parameters a, b, and $c/3$, depending on film orientation [31.65]. The epitaxial relations, in terms of pseudocubic lattice parameters, are shown schematically in Fig. 31.3, for c-oriented YBCO films on (001) and (110)NGO, on the left, compared with a-oriented films on (110)NGO, on the right. For epitaxy on the (001) gallate plane, both $d/2$ axes include the pseudocubic angle γ which slightly deviates from the cubic angle by $\Delta\gamma = 90° - \delta$. Thus, in-plane shear strains in the orthogonal film cell are introduced. For epitaxy on (110) planes the [110] direction pointing out of the (110) plane is inclined by $\Delta\gamma$, leading to *out-of-plane* shear strains in the film. The values of $\Delta\gamma$ at 1000 and 25 °C are 0.24 and 0.32° ($LaGaO_3$), and 0.34 and 0.74° ($NdGaO_3$).

Both compounds, YBCO and NdBCO, undergo a structural phase transition when they are cooled to room temperature after the growth. They are tetragonal ($a = b, c$) at LPE growth temperature of about

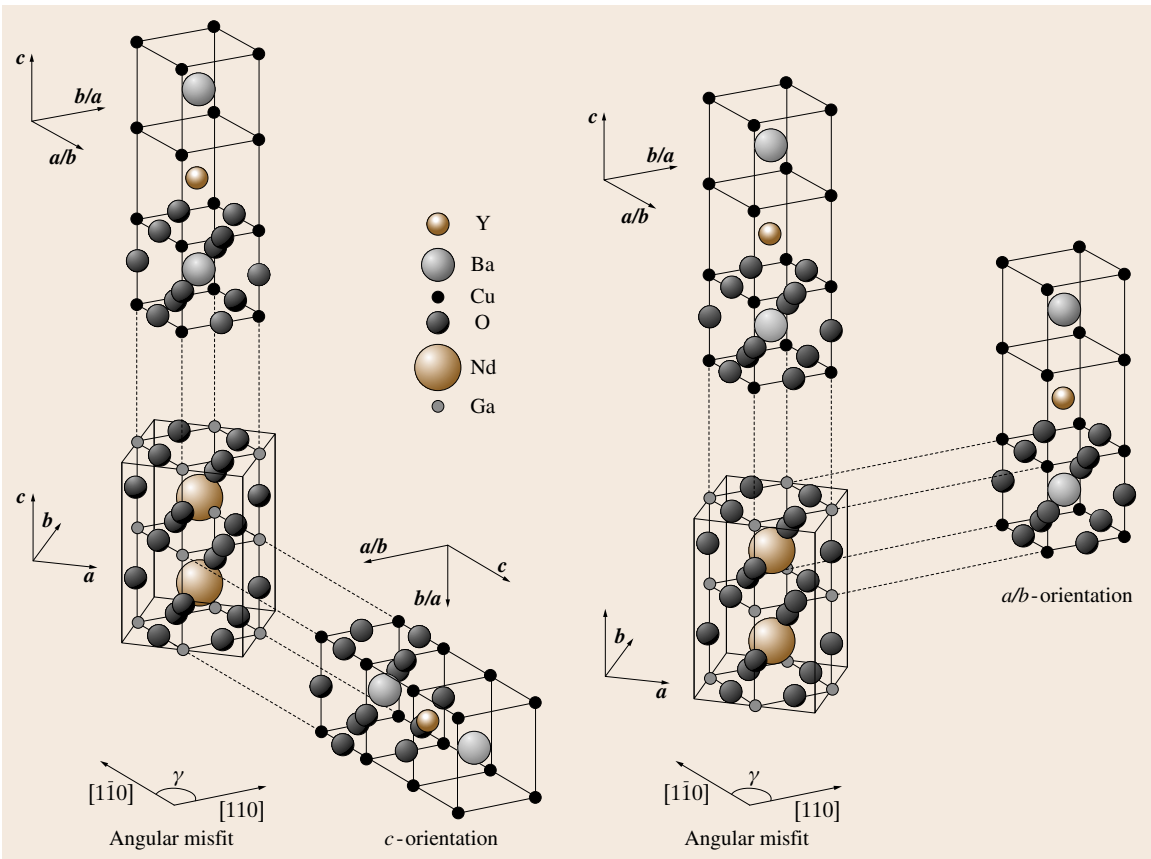


Fig. 31.3 Substrate–film orientations observed in epitaxial growth of YBCO on NGO

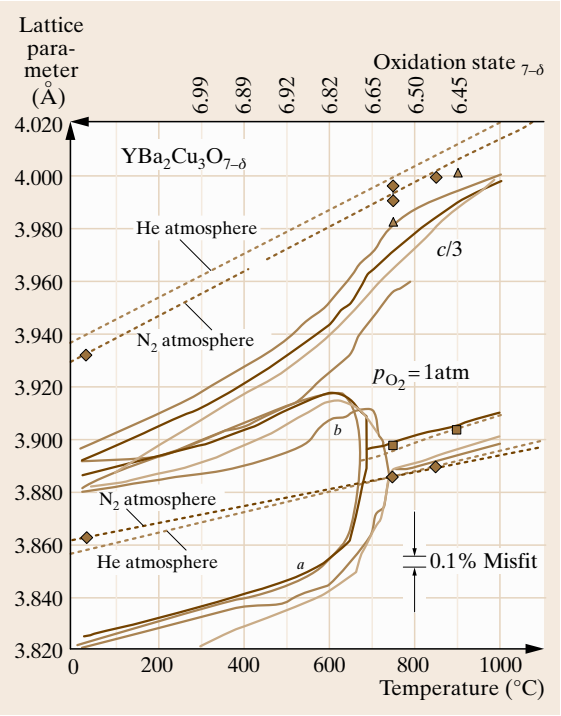


Fig. 31.4 Thermal expansion and oxidation behavior for YBCO according to different authors at $p_{O_2} = 1$ atm (after [31.66])

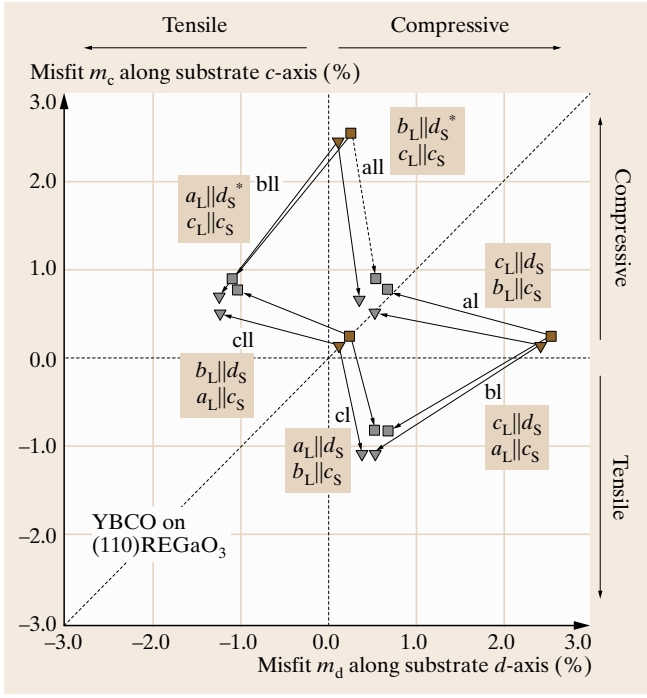


Fig. 31.5 Misfit relations of YBCO on (001)$NdGaO_3$ (□) and $PrGaO_3$ (▽) at 1000 °C (*full symbols*) and room temperature (*open symbols*). *Arrows* mark the path of cooling (and oxidation) after growth. Films are *a*-, *b*- or *c*-oriented (after [31.65])

1000 °C, and orthorhombic ($a \neq b, c$) once at room temperature. For YBCO, the splitting of the tetragonal *a*-axis into two *a*- and *b*-axes during this phase transition occurs at a temperature of about 600 °C, depending on the oxidation state of the material, as shown in Fig. 31.4. Since YBCO and NdBCO cannot directly be grown as orthorhombic phase, they have to be annealed in oxygen after the growth to become superconducting. This poses a formidable challenge for LPE growth of high-quality films, and will be discussed in detail in the film morphology section later.

The misfit relations between YBCO and (110) $NdGaO_3$ and $PrGaO_3$ at room temperature and at 1000 °C (approximate film growth temperature) [31.65, 66] are shown in Fig. 31.5. In this figure, assuming a rigid substrate, one can conclude that for *a*- and *b*-films at growth temperature the in-plane *c*-axis of the film, $c_L \parallel d_S$ (where "L" indicates film and "S" indicates substrate), is under high compressive strain of 2.5%, whereas the tetragonal *a*-axis (= *b*-axis) of the film shows only a small misfit of 0.15 and 0.2% for $PrGaO_3$ and $NdGaO_3$, respectively.

31.5.5 LPE Growth System and Film Growth Procedure

An important requirement for the achievement of high-quality LPE layers is precise control of the supersaturation σ. The supersaturation has to be small enough to prevent spontaneous three-dimensional nucleation, to prevent step bunching (described by the kinematic wave theory), and for layer-by-layer growth. The achievable growth mode depends also on substrate parameters, as discussed previously. Technically, precise control of supersaturation requires a corresponding precision of temperature control and programming, in combination with homogenization of the solution by forced convection. For preliminary LPE growth experiments, a chamber furnace can be sufficient. For highest-quality oxide films, low-gradient platinum-wound three-zone vertical LPE furnace systems allow

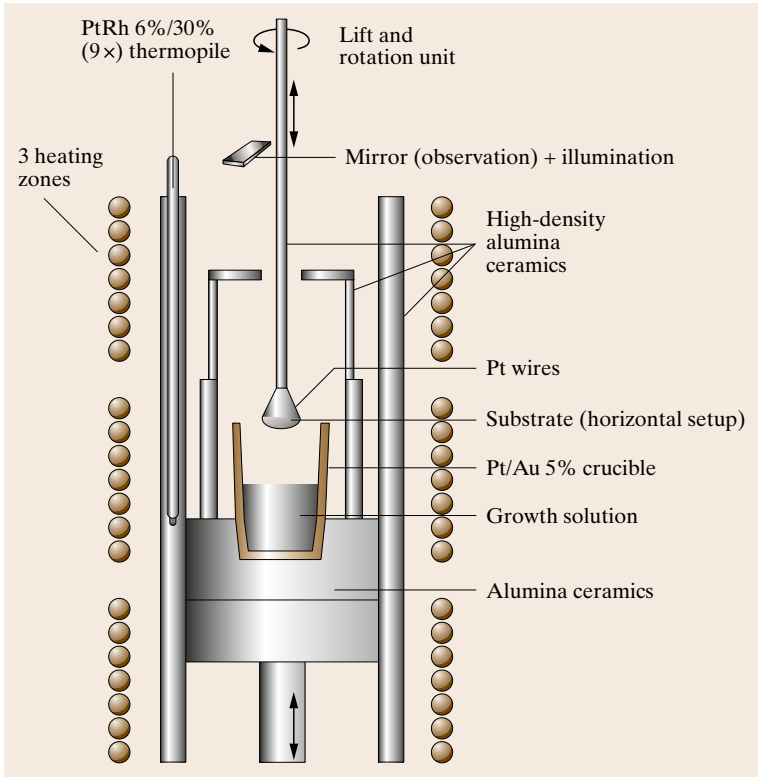


Fig. 31.6 Vertical three-zone LPE growth furnace system

the required control of temperature gradients [31.15, 16]. A typical three-heating-zone LPE growth furnace system is shown schematically in Fig. 31.6.

Such platinum LPE growth system can be used for LPE growth of a variety of oxides, up to 1200 °C. It is equipped with lift and rotation mechanisms. Optical-fiber illumination and mirror are used to observe the surface of the solution and for precise substrate dipping.

In a typical YBCO LPE growth experiment, high-purity starting chemicals (BaO_2, CuO, and Y_2O_3) are well mixed and introduced into the ceramic crucible. The crucible is placed in the furnace and heated to about 1010 °C during 9 h. After soaking for 7 h, the temperature is increased to 1040 °C and reduced to 1000 °C within 90 min. The equilibration to the liquidus temperature is done by the dipping of nonpolished test substrates into the solution. Then, several YBCO films are successively grown on (110)NGO by slow cooling of the solution. The substrates can be mounted in vertical or horizontal position, and the substrate rotation rate is typically between 15–30 and 100–120 rpm, for each configuration, respectively.

31.5.6 Growth Mechanisms and Growth Parameters: Theory Versus Experiment

The primary motivation for using the LPE growth technique for film growth has traditionally been the need for extremely homogeneous crystal material and/or flat surfaces for a given application. In the case of HTSCs, due to their short coherence length (≈ 4 and 12 Å, for *a*-oriented and *c*-oriented films, respectively), extremely flat film surfaces are required for planar tunnel-device applications. For *c*-oriented YBCO films, this represents one *c*-axis monostep height of 12 Å, separated from another monostep by 10 μm lateral distance. This formidable requirement can be better understood if we consider, for example, a step with a height of 1 m that would need to be separated from another step by a distance of 10 km. From this, it is obvious that only a growth mechanism providing a continuous source of steps, such as the layer-by-layer or spiral growth mode, would allow such surfaces to be obtained over macroscopic dimensions. This requirement leads to unprecedented challenges in film synthesis, and therefore

precise evaluation of the tolerances on all substrate and growth parameters is required.

Besides substrate misfit and misorientation, knowledge of the supersaturation is especially critical. Supersaturation can be calculated from the phase diagram, for a given undercooling applied during the film growth. However, the values that are obtained are, at best, indicative only. Growth features, such as spirals, reflect the real conditions existing at the growing interface, and can thereby provide important fundamental data on growth parameters that cannot be measured. The approach discussed below can be applied to other materials as well, and the reader is encouraged to look into the original literature for full details.

In LPE of YBCO on (110)$NdGaO_3$ the film orientation depends on the supersaturation. It was found experimentally that the transition from a-oriented to mixed a/c to c-oriented films occurs with decreasing undercooling (supersaturation), with a threshold at about $\Delta T \approx 3\,\mathrm{K}$ ($\sigma \approx 0.03$) [31.28].

Screw dislocations provide a continuous source of steps, which can propagate over macroscopic distances. It has been shown [31.20], that the distance between steps originating from a single dislocation is proportional to the size of the 2-D nucleus at a given supersaturation. For an Archimedean spiral, the relationship between supersaturation σ and interstep distance y_0 can be expressed by [31.67]

$$y_0 = 19\rho_c = \frac{19\gamma_m a}{k_B T \vartheta}\,, \tag{31.15}$$

with the relative supersaturation $\sigma = (n - n_e)/n_e$, where n and n_e represent the actual and equilibrium concentration of the solute, and a is the length of a growth unit. In the case of solution growth, the energy per growth unit (or molecule) γ_m on the edge of the critical nucleus is of the order of 1/6 of the heat of solution per molecule. The anisotropy of γ_m, which will be higher for low-energy planes, is neglected in the following development.

In the case of YBCO, with $\gamma_m = 4.01 \times 10^{-20}\,\mathrm{J/molecule}$, $a = 3.9\,\text{Å}$ ($a = b$ for tetragonal YBCO at the growth temperature of 996 °C), and $T = 1269\,\mathrm{K}$, we can estimate the achievable interstep distances as a function of undercooling. This is shown in Fig. 31.7, for an Archimedean, monostep, single-source spiral. From this, the size of the critical nucleus was estimated to be about 23 nm for this surface. This is one order of magnitude larger than observed for YBCO films grown by VPE [31.68].

In this figure, we recognize that the undercooling has to be smaller than 0.17 °C in order to obtain interstep distances larger than 10 µm, as required for the specific application discussed herein. Such small under-

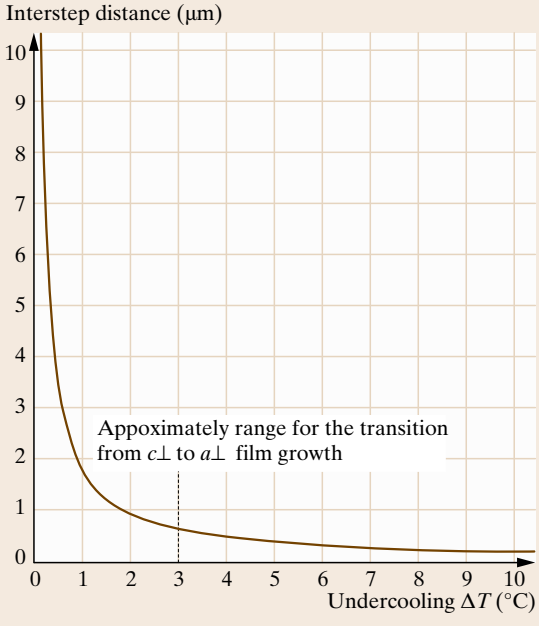


Fig. 31.7 Interstep distances as a function of undercooling (after [31.28])

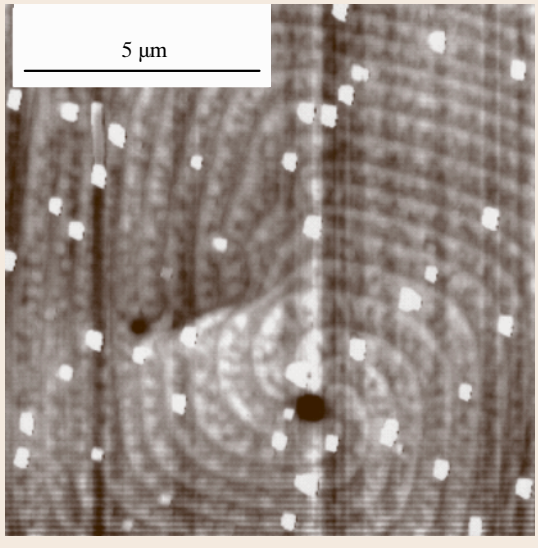


Fig. 31.8 AFM image of a complex YBCO spiral ($m = 2$ and $S = 6$) observed on a LPE film surface (after [31.29])

cooling sets stringent requirements on the temperature stability of the growth process.

Growth spirals can also emerge from groups of dislocations. In this case, the interstep distance is given by $y_0 = 4\pi r_c/\varepsilon$ [31.20], where r_c is the radius of the critical nucleus. For a group of dislocations of the same sign, ε can be as large as the number of dislocations contained in it. A group of S dislocations of the same sign, arranged on an array of length L, has an activity of ε times that of a single dislocation. Accordingly [31.20]

$$\begin{aligned} &\text{If} \quad 2\pi r_c \gg L\,, \quad \text{then} && \varepsilon = S\,, \\ &\qquad 2\pi r_c = L\,, && \varepsilon = \frac{S}{2}\,, \\ &\qquad 2\pi r_c \ll L\,, && \varepsilon = 2\pi r_c \frac{S}{L}\,. \end{aligned}$$

A comparison was made with complex growth spirals observed on *c*-oriented YBCO films grown on (110)NGO [31.29]. Figure 31.8 shows a typical YBCO spiral where six double steps (2×12 Å) emerge from the core region.

Interstep distances and step height of several complex growth spirals observed on single film surface of 1 mm×1 mm, were measured by atomic force microscopy (AFM), and the following relation was found

$$y_0 = \frac{19\gamma_m am}{\varepsilon k_B T \vartheta}\,, \tag{31.16}$$

with $2\pi r_c S/L < \varepsilon < S/2$, and where m represents the number of monosteps of the step source.

The high nucleus density in vapor-grown films results in high screw dislocation densities of typically $10^9\,\mathrm{cm}^{-2}$ [31.68–70] and high grain boundary densities due to the coalescence of misaligned islands [31.71]. High critical current densities are generally observed for such defect-rich films. Island formation can be prevented and a pure step flow mode achieved when the substrate misorientation, which is 3–5° for physical vapor deposition (PVD) and chemical vapor deposition (CVD) [31.71], is adapted to the size of the critical nucleus. In contrast, in LPE, the nucleus density is lower.

With increasing radius of critical nuclei at low supersaturations, the substrate parameters misfit and misorientation become crucial for the initial growth stage and the growth mode of the films. Hence, in LPE, to understand the *a*/*c*-transition of YBCO on (110)$NdGaO_3$, the substrate strain energy contribution must be taken into account. Furthermore, the surface energy between a liquid and its solid is lower than the surface energy between a vapor and its solid. Therefore, in LPE, the strain energy is not negligible with respect to surface energy terms. In order to estimate the influence of substrate misfit and misorientation on the thermodynamically determined equilibrium size of the nucleus, characteristic substrate parameters were compared [31.24]. A relationship between misfit, relative supersaturation, radius of critical nucleus, and film orientation could be established, as shown in Fig. 31.9.

For the achievement of large interstep distances of $y_0 \approx 10\,\mu\mathrm{m}$ between monosteps of 12 Å height for *c*-oriented YBCO layers on (110)NGO, the estimated corresponding undercooling is $\Delta T \approx 0.17\,\mathrm{K}$. From Fig. 31.8, one can see that the maximum tolerable misfit with the substrate would then be 0.08%. For thin, strained *c*-YBCO films on (110)NGO with average misfit of 0.28%, it follows that a maximum interstep distance of $y_0 \approx 2.5\,\mu\mathrm{m}$ between monosteps of 12 Å height can be obtained. It is also found that the substrate misorientation has to be smaller than 0.02° to avoid step bunching [31.24].

Only selected aspects were discussed in this section on HTSCs, to demonstrate the challenges and also

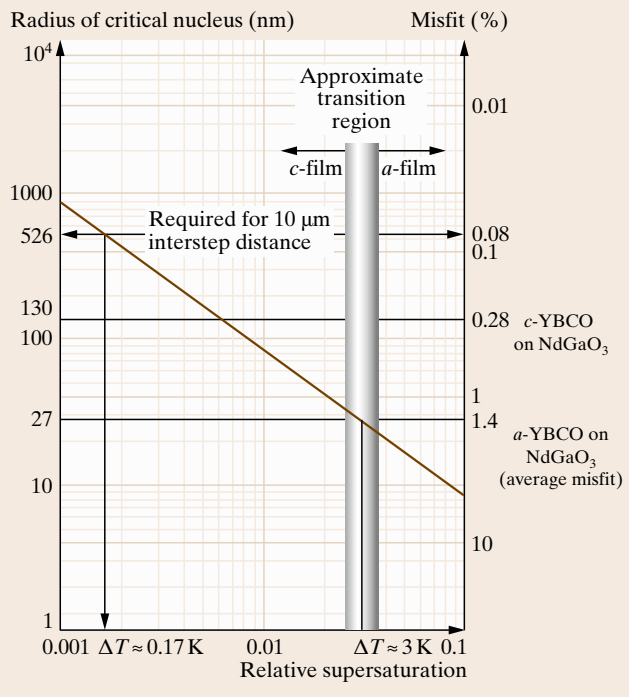


Fig. 31.9 Radius of critical nucleus as a function of misfit and supersaturation in LPE of YBCO (after [31.24])

the potential of this growth technique for the development of high-quality complex oxide films. Cracking and twinning of high-quality LPE REBCO films are further challenges to their application in devices. Whereas the T_c of LPE HTSC films was found comparable, after oxidation, to that of VPE-grown material, the critical current density was usually lower due the high quality of the LPE films (lack of pinning centers). The interested reader should refer to the specific literature for additional details.

31.6 LPE of Calcium Gallium Germanates

Langasite-type materials represent a family of trigonal acentric crystals whose structure was found to be similar to that of calcium gallium germanate, $Ca_3Ga_2Ge_4O_{14}$, (CGG, space group D_3^2-$P321$). They were first investigated as laser hosts [31.72], and rediscovered later for their outstanding piezoelectric properties [31.73]. So far, the most investigated ternary compositions have been $La_3Ga_5SiO_{14}$ (LGS), $La_3Ga_{5.5}Ta_{0.5}O_{14}$ (LGT), and $La_3Ga_{5.5}Nb_{0.5}O_{14}$ (LGN). Langasite-type materials are the first real competitors to quartz in high-precision resonators and high-temperature high-pressure sensors and displace a quartz technology that has been dominant over the past 50 years [31.74–76]. The advantages of langasites over quartz include lower acceleration sensitivity; higher piezoelectric coupling, which enables devices to be made smaller; and higher Q (quality factor), which reduces phase noise and enables higher-frequency operation. Unlike quartz, these materials have no phase transition below the melting point, which enables devices capable of high-temperature operation.

Ternary langasites LGS, LGT, and LGN, can be grown by Czochralski technique. However, due to inherent materials properties (phase relations/stability) and particularities of the growth process, these crystals and wafers show a defect structure and inhomogeneities [31.77] that lead to nonreproducibility in surface acoustic wave (SAW) device parameters [31.78] and discrepancies in fundamental measurements. Recently, high-quality 2 inch diameter LGT crystals have enabled the achievement of high Q-values [31.79] and overall better reproducibility in resonator parameters. It is also acknowledged that better values would probably by possible if the crystal quality could be further improved. In particular, better ordering of atoms in the LGT lattice and lower density of point defects could be beneficial.

Ternary langasites can be described as a mixed framework consisting of an octahedra (1a), and two types of tetrahedra, a small (2d) and a larger (3f). The holes of the Thomson cubes are occupied by large cations (3e) [31.80,81]. Due to the presence of four sites for cations, the CGG-type crystal structure accommodates a wide spectrum of ions. Structure-forming ions, dopant ions, and their radii range (tolerance factor) for corresponding coordination number have been determined [31.81], and a stability diagram derived [31.82]. In ternary CGG-type materials, the four sites can be populated by more than one kind of cation, and this leads to a disordered structure of these materials. In LGT, the Ga and Ta ions share the same octahedral site, which leads to compositional inhomogeneities. Best ordering of atoms, and hence, highest Q-value, is expected from crystals grown near thermodynamic equilibrium. This was one of the reasons to investigate LPE growth of langasite materials. In addition, as we know from semiconductor studies (Sect. 31.1), the concentration of native intrinsic (point) defects is very low in crystal material grown by LPE or from solutions. From fundamental thermodynamics, for the same material, the point defect concentration will be lower if a growth process is used that allows a lower growth temperature. In Czochralski growth of ternary langasites, the growth temperature is about 1450 °C, whereas the same materials can be grown from solutions at about 950 °C [31.83–85]. This growth temperature difference results in a remarkable reduction in the concentration of intrinsic point defects. This may result in higher Q-values, and hence this was a further motivation to investigate LPE growth of langasites.

31.6.1 Solvent System

Finding a suitable solvent for the growth of such complex materials represents one of the major challenges. A first selection of possible fluxes is always done according to chemical and structural aspects, and through systematic studies of available respective or related phase diagrams. Compared with the solute, the solvent should be chemically similar in the type of bonding, but is should have components that present sufficient

crystal-chemical differences to avoid incorporation of solvent species into the crystal.

In langasites, with decreasing temperature, the following crystallization sequence of possible phases is expected from respective melting points and considerations of the complexity of the corresponding oxide. Simple oxides (R_2O_3; R = A, B) have the highest melting temperature, followed by the perovskite (ABO_3) phase at a lower melting temperature, then the garnet ($A_3B_5O_{12}$) phase could be expected, and finally the langasite phase at the lowest temperature. However, when $A = La^{3+}$, the garnet phase does not exist, because La^{3+} is too large for the garnet crystal structure. Hence, in ternary langasites, $LaGaO_3$ is the phase likely to crystallize in a temperature region above the langasite liquidus. Hence, an ideal solvent should present sufficient solubility for the langasite phase below this temperature.

Alkali vandates, molybdates, and tungstates were widely used as fluxes for crystal growth of silicates and germanates. Li_2MoO_4:MoO_3 readily dissolves many oxides, and was used for the growth of BeO, GeO_2, SiO_2, TiO_2, Be_2SiO_4, Y_2SiO_5, as well as for emerald ($Be_3Al_2Si_6O_{18}$). Early attempts to crystallize the langasite phase from Li_2MoO_4:MoO_3 flux were not successful [31.85].

Another typical solvent for oxides is PbO. PbO-B_2O_3 fluxes as well as BaO-BaF_2-B_2O_3 ternary solvent systems were successfully applied for the flux and LPE growth of perovskites and garnets [31.16]. Bi_2O_3-based fluxes have similarities to PbO-based fluxes, and were explored as lead-free alternates for the growth of magnetic garnets. However, two major problems were reported: a flux difficult to remove after the growth, leaving usually a rough film surface after cleaning, and a growth-induced anisotropy of about one order of magnitude lower than with a PbO-B_2O_3 flux [31.16]. Due to the valence state of Bi^{3+}, bismuth-based solvents cannot be used for the growth of many oxides containing large rare-earth ions such as La^{3+}, because of possible substitution, therefore they were not investigated in those studies.

For the growth of LGS, LGT, and LGN, PbO-based solvent systems with various solute ion ratios and concentrations were tested by flux-growth experiments. In preliminary experiments performed between 950 and 1000 °C, lanthanum gallate was often obtained as secondary phase, and it became evident that the growth temperature had to be lower. However, PbO has a melting point of 886 °C and becomes very viscous around 900 °C. Thus, a eutectic-forming additive was sought to lower the growth temperature towards the expected stability field of the ternary langasites. The addition of MoO_3 was found to be beneficial, and LPE films could then be successfully grown in a lower temperature range from a solvent with PbO-to-MoO_3 ratio of 12 : 1 [31.85].

31.6.2 Substrates for Homoepitaxial LGT LPE Film Growth

In LPE, the substrate misfit is a determining factor in the achievable growth mode and film quality. The availability of Czochralski-grown LGS, LGN, and LGT substrates to be used as substrates for homoepitaxial LPE growth therefore represents an important advantage.

In case of homoepitaxy (LPE), the supersaturation required to initiate nucleation is extremely small. Hence, substrate surface defects, scratches, as well as residual strain due to the polishing process, will strongly affect the initial stages of film growth. Chemical etching is an efficient method that can be used to remove the damaged surface substrate layer prior to LPE. For LGT, similarly to perovskites and garnets, hot orthophosphoric acid was found to be particularly suitable, also to reveal substrate defects [31.77]. In langasites, striations can be revealed by etching the substrate in H_3PO_4 at 130 °C during 2–3 h, as shown in Fig. 31.10.

In ternary langasites, striations are likely due to the solid solubility between Si-Ga in LGS (tetrahedral site, CN IV), between Nb-Ga in LGN, and between Ta-Ga in LGT (octahedral site CN VI), associated with temperature fluctuations at the growth interface during Czochralski growth. These striations cause periodic (three-dimensional, 3-D) variations of lattice parameters and inhomogeneous dopant/impurity incorporation. In Czochralski growth of LGT, striations can be strongly reduced, below the detection limit of x-ray synchrotron topography, when all growth parameters are properly optimized [31.79].

When a film is grown by LPE on a striated substrate of same composition (homoepitaxy), the first monolayers of the film will try to adapt their composition to minimize the misfit strain. This phenomenon is well known, especially in semiconductor heteroepitaxy where it has been documented as the lattice pulling effect. A film grown on a striated substrate will show a variation of composition as a function of its thickness, which in turn may affect the film properties. It is therefore important to investigate such defects in the substrates used for epitaxial deposition.

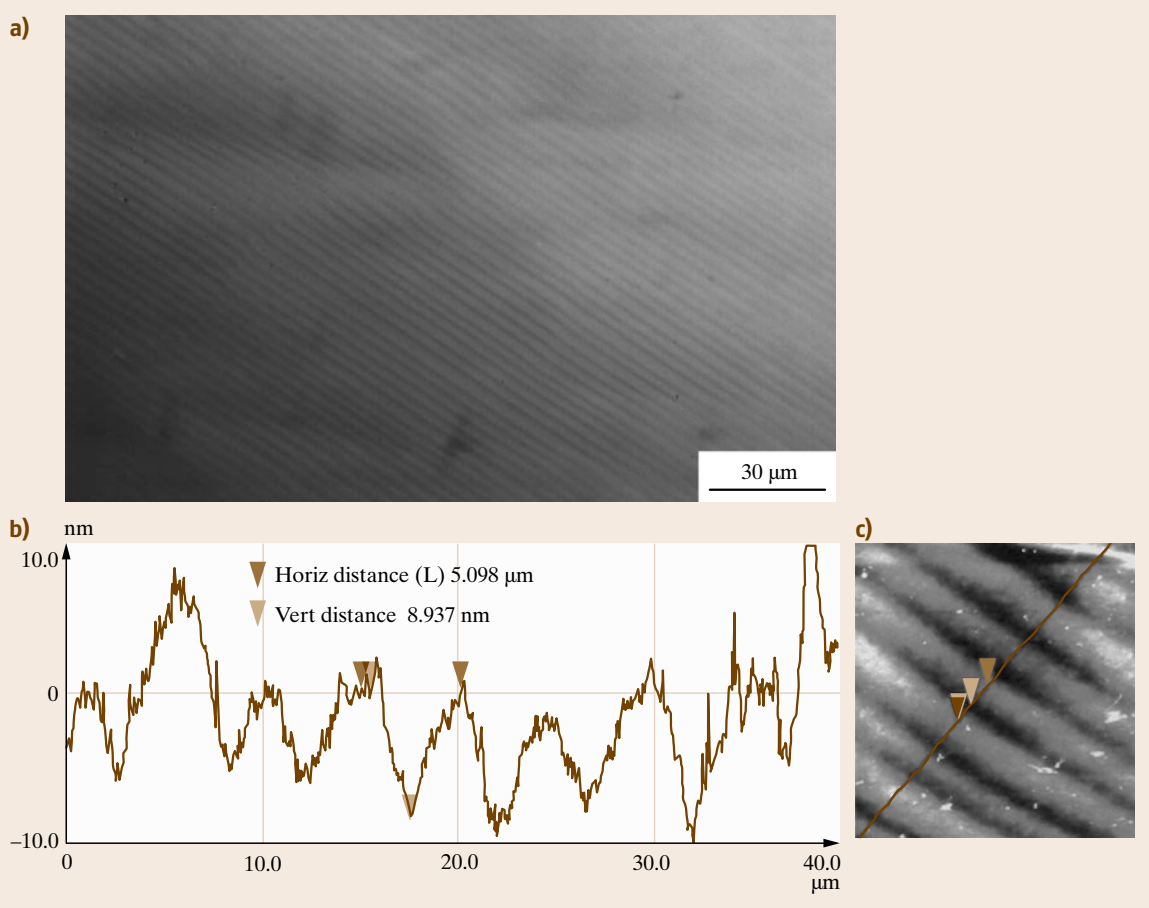


Fig. 31.10a–c Striations on x-cut LGT substrates after chemical etching in hot H_3PO_4. **(a)** Nomarski microphotography, **(b,c)** AFM profile

31.6.3 LPE growth of LGS, LGT, and LGN

Langasite LPE films are grown in a similar vertical furnace system to that previously described. In a typical experiment [31.84, 85], the starting oxides and dopants, consisting of 10–15 wt % stoichiometric LGT (or LGN) in the PbO solvent, are hand-mixed, introduced in a Pt crucible of about 45 cm^2, and covered with a lid. The crucible is placed in the furnace, and the furnace is heated to 980 °C in 5 h, followed by 24–48 h soaking at this temperature. Once during this time, the flux is gently mixed by using a fork made of Pt stripes mounted on an alumina rod. After soaking, the lid is removed and the temperature progressively is lowered to 900 °C and kept there for 5–10 h. The search for the liquidus is then carried out by slowly cooling the furnace while dipping unpolished LGT seeds and analyzing the phases that crystallize onto them by x-ray diffraction (XRD). Several LPE films are then grown in a temperature range of typically 900–850 °C. After the growth, the residual flux can easily be removed by cleaning the film in diluted nitric acid at room temperature.

31.6.4 Structural and Chemical Characterization of Doped LGT LPE Films

Various techniques have been used for the structural and chemical characterization of high-quality LPE films, and the best way to evaluate the films is usually to compare their quality with the substrates. In the following, only selected methods are discussed.

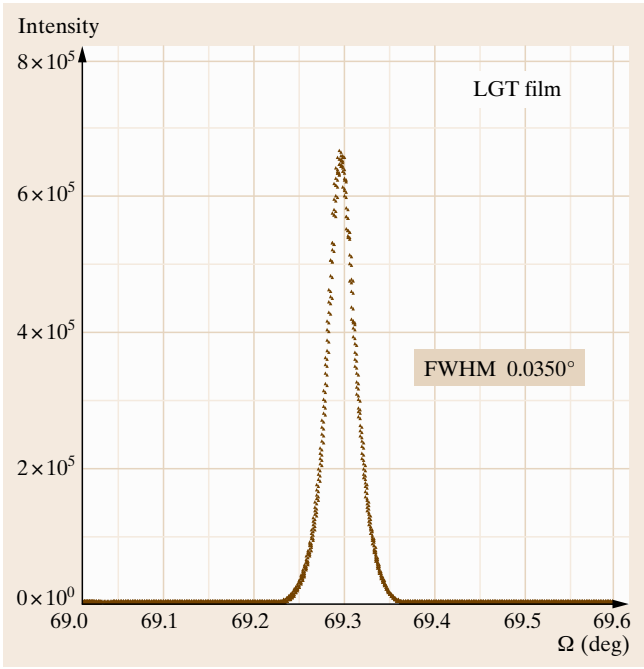


Fig. 31.11 Rocking curve of *x*-LGT film, (004) reflection (after [31.86])

Rocking Curve Measurements

The structural characterization of high-quality crystals and films can be done be rocking curve measurements. During such measurements, the crystal is rotated (rocked) through the Bragg angle, and the reflected beam is measured by a fixed counter. Each potentially slightly misoriented subgrain comes into reflection as the crystal is rotated. Hence, the width of a rocking curve is a direct measure of the range of orientation present in a crystal. For a perfect crystal, the theoretical full-width at half-maximum (FWHM) is on the order of 0.003°. Very few natural crystals reach this value, and most single-crystals have FWHM 10–100 times larger.

In the case of LGT, films and substrates with different orientation were analyzed [31.86]. For *x*-oriented LGT substrates, typical value of about FWHM of 0.044° was obtained for the (004) rocking curve. The rocking curve of an *x*-oriented Al:Ti:LGT film is shown in Fig. 31.11. A FWHM of 0.035° was obtained, which indicates its very high structural perfection, better than that of the substrate.

The structural perfection of *x*-oriented LGT films was consistently better than that of LGT substrates. Also, the surface of *x*-LGT films is usually extremely

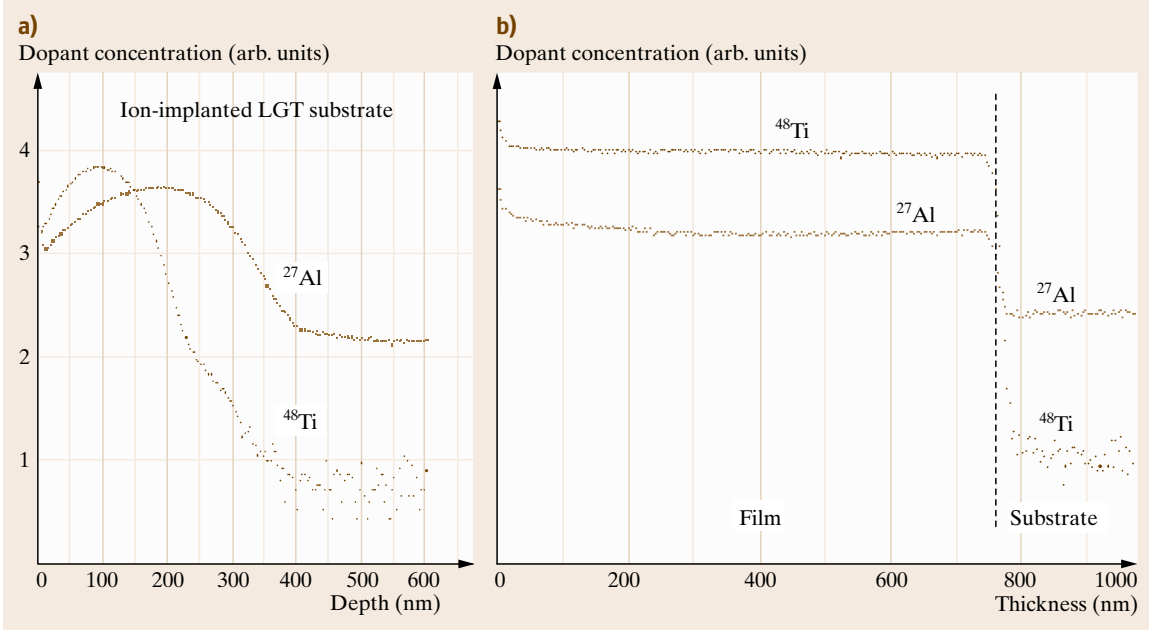


Fig. 31.12a,b Dopant concentration SIMS profile of: Al:Ti ion-co-implanted LGT substrate **(a)**, and in situ codoped Al:Ti:LGT LPE film **(b)**

smooth, whereas *y*-oriented LGT films traditionally show a stepped surface [31.83]. Consequently, we obtained FWHM values slightly larger for *y*-oriented LGT LPE films, typically around 0.0465°. These values are quite remarkable for oxide films.

Quantitative Estimation of Dopant Concentration by SIMS

Secondary-ion mass spectroscopy (SIMS) is an analytical technique used for near-surface and small-area analysis. Depth profiling from about 1 nm to a few microns is possible. All elements and their isotopes can be detected by SIMS. Dynamic SIMS is used for bulk compositional analysis and depth profiles of trace elements. Several dopants were explored. Herein we focus on Al:Ti:LGT LPE films [31.86]. For quantitative measurement of the dopants concentration films by SIMS, standards are needed. Hence, Al and Ti ion-(co)implanted LGT substrates were used as standards. Figure 31.12a shows the SIMS profile of an Al:Ti ion-(co)implanted *x*-LGT substrate. The curved shape of the depth distribution is typical for ion-implanted samples, and demonstrate the inhomogeneous dopant concentration obtained by ion implantation. This should be compared with Fig. 31.12b, which is the SIMS profile obtained for an *x*-oriented Al:Ti:LGT LPE film.

One can see that the Al and Ti dopants are extremely homogeneously distributed through the whole thickness of the LGT LPE film, with almost constant value, and a very abrupt decrease at the film–substrate interface. The SIMS profile also allows a precise measurement of the thickness of the films, which is challenging to measure by other techniques in the case of almost perfect lattice match with the substrate (homoepitaxy).

The thickness of this film is about 770 nm. For a growth time of 30 min, these LGT films had typically thickness of 0.3–4 μm, depending on applied growth parameters. This also demonstrates that, in principle, relatively thin films can be grown by LPE technique. The mean dopant concentration values obtained for this film from the SIMS profile are: ^{48}Ti(100–700 nm) $\approx 2.72\times10^{19}$ atoms/cm^3, and ^{27}Al(200–700 nm) $\approx 5.67\times10^{18}$ atoms/cm^3 within the film depth given in brackets. If one assumes that the isotopic abundance is preserved, the corrected value for Ti is 3.69×10^{19} atoms/cm^3. These are equilibrium concentrations for the applied film growth temperature and conditions.

31.7 Liquid-Phase Epitaxy of Nitrides

Group III nitrides such as AlN, GaN, InN, and their alloys are the focus of intense research for their high potential in optoelectronic devices and high-power high-temperature electronics. These nitrides form ternary and quaternary solid solutions which enable the energy bandgap of the devices to be tuned from about 0.7 eV (InN) to 3.45 eV (GaN), to 6.2 eV (AlN). Hence, they are excellent candidates for engineering of materials with a specific bandgap. The III–V nitrides and their solid solutions are suitable for a wide rang of applications, for example, for high-temperature piezoelectric and pyroelectric applications, surface acoustic wave (SAW) devices, light-emitting diodes (LEDs), laser diodes (LDs), and ultraviolet (UV) detectors and sensors.

Wide-bandgap nitride semiconductors have been successfully commercialized as LEDs for solid-state lighting/illumination, compact laser sources for digital versatile disk (DVD) heads, and even high-power electronics in the form of high-electron-mobility transistors (HEMTs). However, doping issues, the lack of lattice-matched substrates, and high film defect density are currently major obstacles for the reliable development of high-performance devices.

The field of nitrides would highly profit from the availability of bulk GaN and AlN single crystals. However, their synthesis is faced with several difficulties. They cannot be grown from the melt by Czochralski or Bridgman technique due to their extremely high decomposition temperature of about 45 000 atm for GaN at a (theoretical) melting temperature of about 2500 °C. Various methods have been applied to synthesis of bulk nitrides, but none of them has yet provided substrates of sufficient quality and/or sufficient yield to be successfully implemented in device technology.

Interest in nitrides as semiconducting materials in the blue/UV range dates back to early 1970, with the first studies on epitaxial deposition of GaN with promising electrical and optical properties. However, the material showed strong n-type character, and p-type doping could not be achieved. Thus, interest in GaN faded until 1989, when p-type conduction could be

demonstrated for the first time by low-energy electron-beam irradiation (LEEBI) of Mg-doped films [31.87]. The majority of GaN layers are grown from the vapor phase, by MOVPE, MBE, or HVPE, on mismatched substrates. On sapphire and SiC, the layers present a columnar structure consisting of many small hexagonal grains [31.88], which are tilted and rotated within the GaN film and which give rise to very high dislocation densities of about 10^9–$10^{10}\,\mathrm{cm}^{-2}$. In spite of this, very bright LEDs and LDs have been demonstrated [31.89]. The highest brightness and lifetimes are reported for layers grown by the epitaxial lateral overgrowth (ELO) technique on sapphire [31.90] and SiC [31.91], which is due to the reduction of the dislocation density to about $10^5\,\mathrm{cm}^{-2}$ that can be obtained using ELO. This clearly demonstrates the impact of the crystalline perfection of the GaN layer on device performance.

Aluminum nitride has great potential in UV sensor devices. It also presents the required attributes for high-temperature high-power applications, where a high bandgap energy, high thermal and chemical stability, low leakage currents, high breakdown currents, high dielectric constant, and high resistance are required. In particular, AlN has emerged as a potential candidate for high-energy-density capacitors. Film defects such as low-angle grain boundaries (LAGBs), threading dislocations (TDs), and the development of a columnar structure are responsible for leakage currents. Hence, despite its potential, this material has not yet found broad application in such devices. Improving the quality of AlN films (and single crystals) is a critical challenge towards applications of AlN.

The case of InN and InGaN solid solutions is particularly interesting. This material received less attention than AlN and GaN until recently, when the energy gap of wurtzite InN was found to be about 0.7 eV [31.92–94] instead of the previously reported value of 1.9 eV [31.95]. This has extended the range of the energy gaps of group III nitride alloys from the deep-ultraviolet to the near-infrared spectral region. It has been shown that the bandgap of InGaN solid solutions can be varied continuously from 0.7 to 3.4 eV [31.96], providing a full-solar-spectrum material system for multijunction solar cells. In addition, a much greater radiation resistance has been reported for InGaN alloys compared with materials such as GaAs and GaInP. This makes them particularly suitable for radiation-hard high-efficiency solar cells for space exploration [31.97]. Here also, the highest photovoltaic cell efficiencies are expected from low-defect films and lattice-matched high-quality substrates, and this is a further area where LPE growth is particularly important to explore.

Today, the majority of AlInGaN devices rely on vapor-phase growth, often on highly mismatched substrates such as sapphire and SiC, leading to epilayers with large density of defects. As can be recognized from above, almost all applications of nitrides would greatly benefit from high-quality crystal/film nitride material. It is therefore surprising that LPE growth of nitrides is only marginally studied by a few groups.

31.7.1 Developments and Trends in LPE of GaN and AlN

LPE growth of GaN has been explored as early as 1972, by *Logan* and *Thurmond* [31.98], with the successful growth of GaN films on sapphire from Ga and Ga + Bi solvents, and ammonia gas as nitrogen source. A low solubility of $x_{\mathrm{GaN}} \approx 3\times10^{-5}$ at 1150 °C at 1 atm pressure was reported. LPE growth of GaN on highly mismatched sapphire substrates (misfit ≈ 1.6%) turned out to be very challenging, and n-type material was obtained, similarly to vapor-grown films. GaN and InGaN films were successfully grown by LPE on a variety of substrates: sapphire, $LiGaO_2$ (LG), $LiAlO_2$ (LAO), and on MOVPE/HVPE buffer layers [31.99]. The kinetics of GaN formation from Ga solvent in ammonia atmosphere was investigated [31.100]. In 1964, *Glemser* [31.101] showed that Li_3N dissolves in melts based on lithium halides. Li_3N is one of the few nitrides that can be grown by Czochralski technique, and the solubility of nitrogen in molten Li_3N has been investigated [31.102]. Later, Li and Li_3N [31.103] fluxes were applied for the growth of GaN. Other solvent systems were investigated for the synthesis of nitride single crystals. Solutions based on Ga with Na have been successfully applied to the growth of GaN [31.104, 105], and dislocation densities as low as $2.3\times10^5\,\mathrm{cm}^{-2}$ were reported [31.106]. Considering the fact that the LPE growth process that can be applied is very inexpensive compared with the multistep VPE growth technology necessary for similar film quality, there is significant potential in this approach. A remarkable aspect of LPE of GaN is that films with dislocation densities as low as $10^4\,\mathrm{cm}^{-2}$ can be grown on MOCVD GaN templates that have a much higher dislocation density (10^7–$10^8\,\mathrm{cm}^{-2}$) [31.107]. Most GaN LPE and flux-growth experiments were performed at pressure of about 40–50 atm in sealed crucibles [31.106, 107]. In such systems, not only the

temperature but also the pressure becomes a variable. Besides challenges in nucleation control, growth stability is difficult to achieve in such systems. Hence, the use of solvents and LPE methods that allow the growth of GaN at atmospheric pressure may represent a better choice for the growth of thin high-quality LPE GaN films. From the low defect densities already obtained despite the use of nonideal templates [31.107] and nonoptimal LPE growth conditions, we can certainly predict that lowest dislocation (defects) densities will some day be reported for LPE-grown nitride films.

In the case of AlN, *Dugger* was the first to demonstrate the growth of AlN crystals of about 1 mm length from AlN-Ca_3N_2 solutions [31.108]. Recent studies with similar fluxes were only marginally successful, and the growth of AlN single crystals and LPE films from such solutions represents still a challenge.

31.7.2 Substrates for Epitaxy of Nitrides

Under near-equilibrium growth conditions, the substrate parameters misfit, misorientation, as well as the substrate surface structure after polishing/cleaning play a fundamental role. Only at practically zero misfit and low supersaturation can the layer-by-layer or Frank–van der Merwe growth mode, yielding films with highest structural perfection, be expected. Besides sapphire, MgO, SiC, $MgAl_2O_4$, and Si have been investigated as substrate material for nitrides, but all of them have too high a misfit for coherent overgrowth and layer-by-layer growth by LPE. Promising substrate materials such as $LiGaO_2$ (LGO) and $LiAlO_2$ (LAO) have emerged. However, their low chemical and thermal stability (about 900 °C) limits the growth parameters to be applied for any growth technique. Zinc oxide is actually the best lattice-matched substrate for nitrides, available at high yield [31.109]. The combination of ZnO and GaN in multilayer device structures would also be of particular interest to optoelectronic devices. However, ZnO has limited chemical stability with respect to halogenides at high temperature, and therefore cannot be used in chloride-HVPE nitride growth processes. This does not represent a major issue in LPE growth, where the substrate is immersed in the (metallic) solvent and the process is performed in such a way that the substrate never comes into direct contact with the reactive atmosphere.

Among all choices, homoepitaxial deposition provides the best possible conditions for highest-quality GaN LPE-grown films [31.99]. Hence, thick GaN films or GaN substrate crystals would be of greatest interest, also as base structure for the fabrication of devices by VPE. Among all methods, the ammonothermal is actually the most promising method for obtaining large-size ($\geq$ 2 inch) GaN at high yield. In 1997/1998 *Dwilinski* et al. demonstrated ammonothermal growth of GaN and AlN from supercritical ammonia and lithium and potassium amides [31.110]. The excellent properties reported led to intensive research in this area by different groups. It was found that GaN show a retrograde solubility when basic mineralizers are used, and quite high pressures (100–300 MPa) have to be applied [31.111, 112]. In acid mineralizers, the solubility behavior of GaN is conventional, and the process can be carried out at lower pressures ($\leq$ 150 MPa) [31.113], which makes growth stability easier to achieve. The increasing availability of near-lattice-matched substrates will be very significant for major advances in the field.

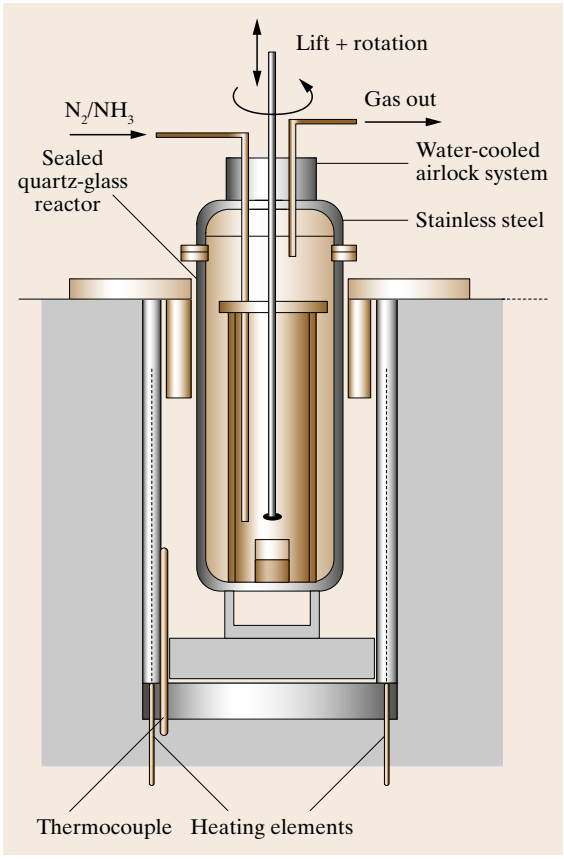


Fig. 31.13 Schematic view of the nitride LPE growth system (after [31.99])

31.7.3 Growth System and Optimization

In the following, a relatively simple growth system and method is presented, which was successfully used for the growth of GaN and InGaN films from Ga or Ga + Bi solutions, using NH_3 nitrogen source [31.99]. The film growth was performed in a sealed quartz glass reactor placed in a chamber furnace schematically shown in Fig. 31.13.

The crucible is placed in an appropriate temperature gradient in the middle of the reactor. A quartz liner is placed between the crucible and the reactor wall, with quartz glass plates placed on its top. This arrangement allows to reduce the convection of the gases above the crucible and is beneficial to prevent contamination of the solution by impurities/contaminants from the quartz reactor and/or airlock system.

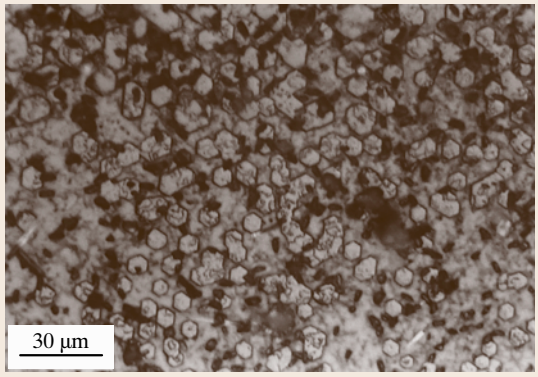


Fig. 31.14 Typical surface morphology of GaN LPE film on sapphire (basal plane). A large lattice mismatch results in the Volmer–Weber growth mode with development of hexagonal GaN islands

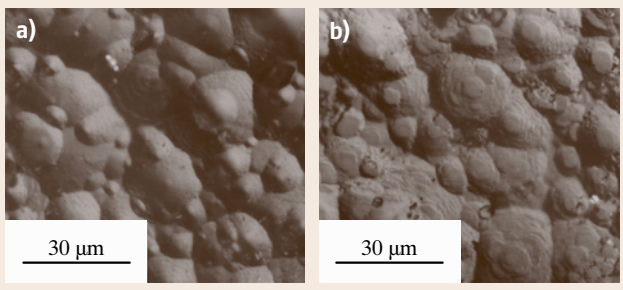


Fig. 31.15 **(a)** Nomarski microphotography of the surface of a HVPE GaN template before LPE growth. **(b)** GaN film morphology at the early stage of LPE growth on the GaN HVPE template

In a typical experiment, the reactor with empty crucible is evacuated and cleaned several times with nitrogen gas and heated to about 850 °C. The liquid gallium or other (liquid/preheated) solvent can be introduced into the crucible from the top of the reactor, through the airlock system. The suitability of this method depend on the solvent used. If it is applied, this procedure has to be done under nitrogen overpressure, to prevent oxygen contamination. The temperature is then raised to the growth temperature. The substrate is fixed on an alumina or graphite substrate holder in horizontal, vertical or inclined position, and introduced into the reactor through the airlock system. The substrate is dipped into the solution, rotated, and the reactive gas mixture is then introduced into the reactor. This procedure allows the use of substrates that would not withstand the reactive gas atmosphere without degradation, for example, ZnO. In principle, various reactive gases can be applied. In the specific experiments described herein [31.99], a gas mixture of NH_3 and N_2 with ratio 1 : 4 was used. After the growth, before removing the substrates from the solution, pure nitrogen is introduced and the temperature reduced to about 800 °C while removing the film from the solution. Growth times of 10–72 h were applied, and several films were grown successively.

31.7.4 Morphological Evolution of LPE-Grown Nitride Films

Sapphire substrates (basal plane) have a lattice mismatch of −13.7% with GaN along the a direction, which leads to the Volmer–Weber (or islands) film growth mechanism. GaN LPE films grown directly on sapphire typically present hexagonal islands, which may spread over the entire substrate surface by coalescence, resulting in grain-boundary-rich films. Figure 31.14 shows a Nomarski microphotography of a GaN LPE on sapphire during the coalescence phase.

When GaN LPE films are grown on GaN templates, growth starts preferentially on existing defects. Etch back and regrowth mechanisms can easily take place, as growth proceeds near thermodynamic equilibrium, with very low supersaturation. Hence, surface mobility of atoms is high, which leads to the smoothing effect on growth hillocks present on GaN buffer layer. This is shown in the Nomarski microscopy images of Fig. 31.15. On the left, we can see the surface morphology of a GaN HVPE seed layer grown on sapphire before LPE. Conical islands and overall roughness along the macrosteps are clearly visible. On the right,

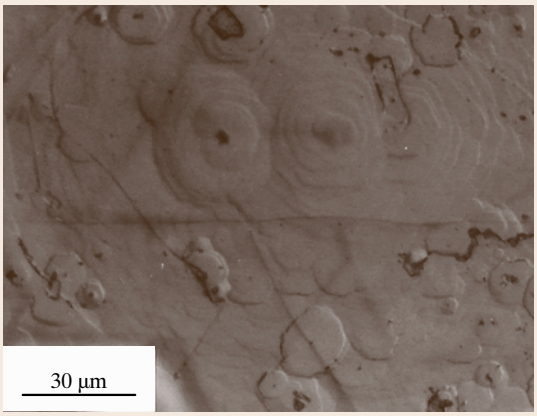


Fig. 31.16 Nomarski microphotograph of the surface of a GaN LPE film, grown on a HVPE MOVPE seed layer on sapphire (after [31.99])

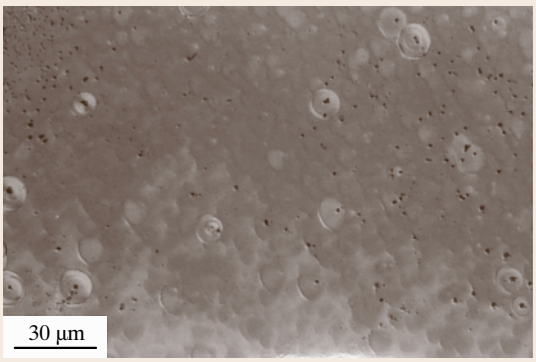


Fig. 31.17 InGaN film grown by LPE on *c*-sapphire, without a buffer layer. The lattice pulling effect can tentatively explain the (partial) suppression of the island growth mechanism that would be expected from the large lattice mismatch with the substrate

after LPE growth, we recognize the overgrown regions, with smooth regions that start to develop from the pre-existing islands. For longer growth time, such film may evolve into a smooth surface, with large flat areas.

Improved surface morphology can be obtained by using better-quality GaN templates, as shown in Fig. 31.16. The overgrown regions develop to rather flat areas already in the early stages of LPE film growth, when using a multilayer HVPE GaN template grown on a MOVPE GaN buffer layer on sapphire [31.99].

InGaN LPE films could be successfully grown [31.114] using the same growth system described above. However, the film morphologies that were obtained for films grown on sapphire, without a buffer layer, were quite surprising. In contrary to GaN LPE films, island growth was largely suppressed, and smooth surfaces did develop despite the lattice mismatch. Rather smooth, rounded hillocks developed and steps propagated over large surface area. This can be recognized in the Nomarski microphotographs of InGaN LPE film, which is shown in Fig. 31.17.

The InGaN film shown in Fig. 31.17 was grown from liquid Ga and ammonia reactive gas atmosphere by the procedure previously described. Growth conditions were similar to those applied for LPE of GaN [31.99], and film thickness was typically a few microns (GaN and InGaN). This significant difference in growth morphology can tentatively be explained by a possible lattice pulling effect that would minimize the lattice mismatch and hence reduce or suppress the island growth mechanism. Further investigations are needed to confirm this. However, this preliminary result is very promising, and LPE-grown high-quality InGaN solid-solution films may become a valuable approach in the fabrication of high-efficiency radiation-hard photovoltaic solar cells.

31.8 Conclusions

Liquid-phase epitaxy has been a major technique for the development of many materials, and for a variety of applications. The challenges of this technique, and especially the stringent requirements in all substrate and growth process parameters, have often been critical limiting factors. This has been shown especially in Sect. 31.5, where some aspects of LPE growth of high-temperature superconductors were presented. However, the structural perfection and homogeneity of the films that can potentially be achieved is the highest possible, when compared with any other melt or thin-film growth techniques. This is not only true for the case of semiconductors (for example, GaP; Sect. 31.1) but also in the case of complex oxides. There is a misconception that only thick films can be grown by LPE. Though this has been true for a majority of compounds investigated by LPE so far, the recent results presented in Sect. 31.6 on the LPE growth of langasites demonstrates that thin

films of the order of a few hundred nanometers can also be grown by LPE.

The technologically important field of nitrides could greatly benefit from LPE-grown low-defect films, and this was discussed in Sect. 31.3. Several nitride applications actually hinge on the high density of film defects, and the lack of lattice-matched substrates. Despite the use of sophisticated multistep deposition methods and epitaxial lateral overgrowth, and the significant amount of research by several groups in this area over the past 10 years, the lower limit for the dislocation density achieved is about $10^5\ \mathrm{cm}^{-3}$, which is two or three orders of magnitude too high for many applications. Especially here, the LPE technique could possibly lead to major advances, enabling material with lower dislocation density that could also be used as ideal substrates for growth of multilayer structures by VPE. In fact, LPE GaN films with low dislocation densities of the order of $10^2\ \mathrm{cm}^{-3}$ have already been demonstrated, and they can be grown without the need of any external patterning or the ELO technique. Besides GaN, AlN and InN could also benefit from LPE studies; for example, one would expect highest efficiencies for In-rich InGaN photovoltaic cells based on LPE-grown material.

Despite its challenges, the LPE technique offers the unique possibility to explore fundamental structure–property relations in a variety of doped and undoped materials grown under equilibrium conditions. This technique allows the study of incongruently melting materials with potentially interesting properties that are not investigated because they that cannot be grown by any other (melt) growth technique. In conclusion, there is significant margin for the discovery and study of novel materials using the LPE growth technique.

References

31.1 H. Nelson: Epitaxial growth from the liquid state and its application to the fabrication of tunnel and laser diodes, RCA Review **24**, 603 (1963)

31.2 L.R. Dawson: High-efficiency graded band-gap $Ga_{1-x}Al_xAs$ light-emitting- diodes, J. Appl. Phys. **48**, 2485–2492 (1977)

31.3 J.I. Nishizawa, Y. Okuno: Liquid-phase epitaxy of GaP by a temperature difference method under controlled vapor-pressure, IEEE Trans. Electron. Dev. **22**, 716–721 (1975)

31.4 J. Nishizawa, Y. Okuno, M. Koike, F. Sakurai: Bright pure green emission from N-free GaP LEDs, Jpn. J. Appl. Phys. **19**, 377–382 (1980)

31.5 M. Konagi, K. Takahashi: Formation of GaAs-(GaAl)As heterojunction transistors by liquid-phase epitaxy, Trans. IEEE J. **94-C**, 141 (1974)

31.6 E. Bauser, D. Kass, M. Warth, H.P. Strunk: Mater. Res. Soc. Symp. Proc. **54**, 267 (1986)

31.7 T. Nishinaga, T. Nakano, S. Zhang: Epitaxial lateral overgrowth of GaAs by LPE, Jpn. J. Appl. Phys. **27**, L964–L967 (1988)

31.8 Y. Suzuki, T. Nishinaga: Epitaxial lateral overgrowth of Si by LPE with Sn solution and its orientation dependence, Jpn. J. Appl. Phys. **28**, 440–445 (1989)

31.9 S. Kinoshita, Y. Suzuki, T. Nishinaga: Epitaxial lateral overgrowth of Si on non-planar substrate, J. Cryst. Growth **115**, 561–566 (1991)

31.10 T. Kochiya, Y. Oyama, T. Kimura, K. Suto, J.I. Nishizawa: Dislocation-free large area InP ELO layers by liquid-phase epitaxy, J. Cryst. Growth **281**, 263–274 (2005)

31.11 S.H. Shin, G.T. Niizawa, J.G. Pasko, G.L. Bostrup, F.J. Ryan, M. Khoshnevisan, C.I. Westmark, C. Fuller: P-I-N CdTe gamma-ray detectors by liquid phase epitaxy (LPE), Presented at the Nucl. Sci. Symp., Orlando (1984)

31.12 G. Bostrup, K.L. Hess, J. Ellsworth, D. Cooper, R. Haines: LPE HgCdTe on sapphire: Status and advancements, J. Electron. Mater. **30**, 560–566 (2001)

31.13 R.C. Linares: Epitaxial growth of narrow linewidth yttrium iron garnet films, J. Cryst. Growth **3/4**, 443–446 (1968)

31.14 H.J. Levinstein, S. Licht, R.W. Landorf, S. Blank: Growth of high-quality garnet thin films from supercooled melts, Appl. Phys. Lett. **19**, 486–488 (1971)

31.15 S.L. Blank: Crystal growth: Magnetic garnets by liquid-phase epitaxy, J. Educ. Modules Mater. Sci. Eng. **2**, 357–390 (1980)

31.16 M.H. Randles: Crystals for magnetic applications. In: *Crystals: Growth, Properties, and Applications*, Vol.1, ed. by C.J.M. Rooijmans (Springer, Berlin, Heidelberg 1978) pp. 71–96

31.17 D. Elwell, H.J. Scheel: *Crystal Growth from High-Temperature Solutions* (Academic, London 1975)

31.18 W. van Erk: The growth kinetics of garnet liquid-phase epitaxy using horizontal dipping, J. Cryst. Growth **43**, 446–456 (1978)

31.19 R. Ghez, E.A. Giess: Liquid phase epitaxial-growth kinetics of magnetic garnet films grown by isothermal dipping with axial rotation, Mater. Res. Bull. **8**, 31–42 (1973)

31.20 W.K. Burton, N. Cabrera, F.C. Frank: The growth of crystals and the equilibrium structure of their surface, Philos. Trans. R. Soc. Lond. Ser. A **243**, 299–358 (1951)

31.21 A.A. Chernov: Crystal growth from the solution and from the melt, Sov. Phys. Usp. **4**, 129 (1961)

31.22 J.A. Burton, R.C. Prim, W.P. Slichter: The distribution. of solute in crystals grown from the melt, J. Chem. Phys. **21**, 1987–1991 (1953)

31.23 G. Winkler: *Magnetic Garnets* (Vieweg, Braunschweig 1981)

31.24 C. Klemenz, I. Utke, H.J. Scheel: Film orientation, growth parameters and growth modes in epitaxy of $YBa_2Cu_3O_x$, J. Cryst. Growth **204**, 62–68 (1999)

31.25 D.P. Norton: Science and technology of high-temperature superconducting films, Annu. Rev. Mater. Sci. **28**, 299–347 (1998)

31.26 H. Hilgenkamp, J. Mannhart: Grain boundaries in high-T_c superconducting films, Rev. Mod. Phys. **74**, 485–549 (2002)

31.27 K.H. Wu, S.P. Chen, J.Y. Juang, T.M. Uen, Y.S. Gou: Investigation of the evolution of $YBa_2Cu_3O_{7-d}$ films deposited by scanning pulsed laser deposition on different substrates, Physica C **289**, 230–242 (1997)

31.28 C. Klemenz, H.J. Scheel: Flat $YBa_2Cu_3O_{7-\delta}$ layers for planar tunnel-device technology, Physica C **265**, 126–134 (1996)

31.29 C. Klemenz: Hollow cores and step bunching effects on (001)YBCO surfaces grown by liquid-phase epitaxy, J. Cryst. Growth **187**, 221–227 (1998)

31.30 A. Chakraborty, K.C. Kim, F. Wu, J.S. Speck, S.P. DenBaars, U.K. Mishra: Defect reduction in nonpolar *a*-plane GaN films using in situ SiN_x nanomask, Appl. Phys. Lett. **89**, 041903 (2006)

31.31 K.A. Dunn, S.E. Babcock, D.S. Stone, R.J. Matyi, L. Zhang, T.F. Kuech: Dislocation arrangement in a thick LEO GaN film on sapphire, MRS Internet J. Nitride Semicond. Res. **5**, W2.11 (2000)

31.32 B.A. Haskell, T.J. Baker, M.B. McLaurin, F. Wu, P.T. Fini, S.P. DenBaars, J.S. Speck, S. Nakamura: Defect reduction in $(11\bar{2}0)$ *m*-plane gallium nitride via lateral epitaxial overgrowth by hydride vapor phase epitaxy, Appl. Phys. Lett. **86**, L1197–L1199 (2005)

31.33 B.A. Haskell, F. Wu, M.D. Craven, S. Matsuda, P.T. Fini, T. Fujii, K. Fujito, S.P. DenBaars, J.S. Speck, S. Nakamura: Defect reduction in (1120) *a*-plane gallium nitride via lateral epitaxial overgrowth by hydride vapor-phase epitaxy, Appl. Phys. Lett. **83**, 644–646 (2003)

31.34 C. Klemenz, H.J. Scheel: Solubility of $YBa_2Cu_3O_{7-\delta}$ and $Nd_{1+x}Ba_{2-x}Cu_3O_{7+\delta}$ in the BaO/CuO flux, J. Cryst. Growth **200**, 435–440 (1999)

31.35 J.G. Bednorz, K.A. Müller: Possible high-T_c superconductivity in the Ba-La-Cu-O system, Z. Phys. B **64**, 189–193 (1986)

31.36 M.K. Wu, J.R. Ashburn, C.J. Torng, P.H. Hor, R.L. Meng, L. Gao, Z.J. Huang, Y.Q. Wang, C.W. Chu: Superconductivity at 93 K in a new mixed-phase Y-Ba-Cu-O compound system at ambient pressure, Phys. Rev. Lett. **58**, 908–910 (1987)

31.37 M. Hawley, I.D. Raistrick, J.G. Beery, R.J. Houlton: Growth mechanism of sputtered films of $YBa_2Cu_3O_7$ studied by scanning tunneling microscopy, Science **251**, 1587–1589 (1991)

31.38 C. Gerber, D. Anselmetti, J.G. Bednorz, J. Mannhart, D.G. Schlom: Screw dislocations in high-T_c films, Nature **350**, 279–280 (1991)

31.39 R.F. Belt, J. Ings, G. Diercks: Superconductor film growth on $LaGaO_3$ substrates by liquid-phase epitaxy, Appl. Phys. Lett. **56**, 1805–1807 (1990)

31.40 G. Balestrino, V. Foglietti, M. Marinelli, E. Milani, A. Paoletti, P. Paroli: Epitaxial films of BSCCO grown from liquid KCl solutions onto several substrates, IEEE Trans. Magn. **27**, 1589–1591 (1991)

31.41 G. Balestrino, V. Foglietti, M. Marinelli, E. Milani, A. Paoletti, P. Paroli, G. Luce: Structural and electrical properties of epitaxial films of BSCCO grown from liquid phase epitaxy, Solid State Commun. **76**, 503–505 (1990)

31.42 S. Narayanan, K.K. Raina, R.K. Pandey: Thin film growth of the 2122-phase of BCSCO superconductor with high degree of crystalline perfection, J. Mater. Res. **7**, 2303–2307 (1992)

31.43 H. Takeya, H. Takei: Preparation of high-T_c Bi-Sr-Ca-Cu-O films on MgO substrates by the liquid phase epitaxial (LPE) method, Jpn. J. Appl. Phys. **28**, L229–L232 (1989)

31.44 R.S. Liu, Y.T. Huang, P.T. Wu, J.J. Chu: Epitaxial growth of high-T_c Bi- Ca-Sr-Cu-O superconducting layer by LPE process, Jpn. J. Appl. Phys. **27**, L1470–L1472 (1988)

31.45 R.S. Liu, Y.T. Huang, J.M. Liang, P.T. Wu: Epitaxial growth of high-T_c superconducting Tl-Ca-Ba-Cu-O films by liquid phase epataxial process, Physica C **156**, 785–787 (1988)

31.46 A.S. Yue, W.S. Liao, H.J. Choi: LPE growth of high-T_c $Bi_2Sr_2Ca_1Cu_2O_x$ films, Cryogenics **32**, 596–599 (1992)

31.47 J.S. Shin, H. Ozaki: Superconducting Bi-Sr-Ca-Cu-O films prepared by the liquid phase epitaxial method, Physica C **173**, 93–98 (1991)

31.48 L.H. Perng, T.S. Chin, K.C. Chen, C.H. Lin: Y-Ba-Cu-O superconducting films grown on (100) magnesia and sapphire substrates by a melt growth method without crucible, Supercond. Sci. Technol. **3**, 233–237 (1990)

31.49 C. Klemenz, H.J. Scheel: Liquid-phase epitaxy of high-T_c superconductors, J. Cryst. Growth **129**, 421–428 (1993)

31.50 C. Dubs, K. Fischer, P. Görnert: Liquid phase epitaxy of $YBa_2Cu_3O_{7-x}$ on $NdGaO_3$ and $LaGaO_3$ substrates, J. Cryst. Growth **123**, 611–614 (1992)

31.51 H.J. Scheel, C. Klemenz, F.-K. Reinhart, H.P. Lang, H.-J. Güntherodt: Monosteps on extremely flat $YBa_2Cu_3O_{7-\delta}$ surfaces grown by liquid-phase epitaxy, Appl. Phys. Lett. **65**, 901–903 (1994)

31.52 T. Kitamura, M. Yoshida, Y. Yamada, Y. Shiohara, I. Hirabayashi, S. Tanaka: Crystalline orientation of $YBa_2Cu_3O_x$ film prepared by liquid-phase epitaxial growth on $NdGaO_3$ substrate, Appl. Phys. Lett. **66**, 1421–1423 (1995)

31.53 H.J. Scheel, F. Licci: Phase diagrams and crystal growth of oxide superconductors, Thermochim. Acta **174**, 115–130 (1991)

31.54 F. Licci, H.J. Scheel, P. Tissot: Determination of the eutectic composition by crystal growth and flux separation: example $BaCuO_2$–CuO_x, J. Cryst. Growth **112**, 600–605 (1991)

31.55 H.J. Scheel, F. Licci: Crystal growth of $YBa_2Cu_3O_{7-x}$, J. Cryst. Growth **85**, 607–614 (1987)

31.56 T. Aichele, S. Bornmann, C. Dubs, P. Görnert: Liquid-phase epitaxy (LPE) of $YBa_2Cu_3O_{7-x}$ high T_c superconductors, Cryst. Res. Technol. **32**, 1145–1154 (1997)

31.57 C. Krauns, M. Sumida, M. Tagami, Y. Yamada, Y. Shiohara: Solubility of RE elements into Ba-Cu-O melts and the enthalpy of dissolution, Z. Phys. B **96**, 207–212 (1994)

31.58 P. Bennema, J.P. van der Eerden: Crystal graphs, connected nets, roughening transition and the morphology of crystals. In: *Morphology of Crystals*, ed. by I. Sunagawa (Terra Scientific, Tokyo 1987), Chap. 1

31.59 G.H. Gilmer, K.A. Jackson: Computer simulation of crystal growth. In: *Crystal Growth and Materials*, ed. by E. Kaldis, H.J. Scheel (North Holland, Amsterdam 1977) pp. 80–114

31.60 D.S. Tsagareishvili, G.G. Gvelesiani, I.B. Baratashvili, G.K. Moiseev, N.A. Vatolin: Thermodynamic functions of $YBa_2Cu_3O_7$, $YBa_2Cu_3O_6$, Y_2BaCuO_5 and $BaCuO_2$, Russ. J. Phys. Chem. **64**, 1404–1406 (1990)

31.61 G.K. Moiseev, N.A. Vatolin, S.I. Zaizeva, N.I. Ilyinych, D.S. Tsagareishvili, G.G. Gvelesiani, I.B. Baratashvili, J. Sestàk: Calculation of thermodynamic properties of the phases in the Y-Ba-Cu-O system, Thermochim. Acta **198**, 267–278 (1992)

31.62 W. van Erk: The growth kinetics of garnet liquid-phase epitaxy using horizontal dipping, J. Cryst. Growth **43**, 446–456 (1979)

31.63 W. van Erk: A solubility model for rare-earth iron garnets in a PbO/B_2O_3 solution, J. Cryst. Growth **46**, 539–550 (1979)

31.64 W. van Erk: Growth of a mixed crystal from an ideal dilute solution, J. Cryst. Growth **57**, 71–83 (1982)

31.65 I. Utke, C. Klemenz, H.J. Scheel, M. Sasaura, S. Myazawa: Misfit problems in epitaxy of high-T_c superconductors, J. Cryst. Growth **174**, 806–812 (1997)

31.66 I. Utke, C. Klemenz, H.J. Scheel, P. Nüesch: High-temperature x-ray measurements of gallates and cuprates, J. Cryst. Growth **174**, 813–820 (1997)

31.67 N. Cabrera, M.M. Levine: On the dislocation theory of evaporation of crystals, Philos. Mag. **1**, 450–458 (1956)

31.68 H.P. Lang, H. Haefke, G. Leemann, H.J. Güntherodt: Scanning tunneling microscopy study of different growth stages of $YBa_2Cu_3O_{7-d}$ thin films, Physica C **194**, 81–91 (1992)

31.69 M. Hawley, I.D. Raistrick, J.G. Beery, R.J. Houlton: Growth mechanism of sputtered films of $YBa_2Cu_3O_7$ studied by scanning tunelling microscopy, Science **251**, 1587–1589 (1991)

31.70 C. Gerber, D. Anselmetti, J.G. Bednorz, J. Mannhart, D.G. Schlom: Screw dislocations in high-T_c films, Nature **350**, 279–280 (1991)

31.71 T. Nishinaga, H.J. Scheel: Crystal growth aspect of high-T_c superconductors. In: *Advances in Superconductivity VIII*, ed. by H. Hyakawa, Y. Enomoto (Springer, Tokyo 1996) pp. 33–38

31.72 A.A. Kaminskii, B.V. Mill, G.G. Khodzhabagyan, A.F. Konstantinova, A.I. Okorochkov, I.M. Silvestrova: Investigation of trigonal $(La_{1-x}Nd_x)_3Ga_5SiO_{14}$ crystals: I. Growth and optical properties, Phys. Status Solidi (a) **80**, 387–398 (1983)

31.73 Y.V. Pisarevski, P.A. Senushencov, P.A. Popov, B.V. Mill: New strong piezoelectric $La_3Ga_{5.5}Nb_{0.5}O_{14}$ with temperature compensation cuts, Proc. IEEE Int. Freq. Control Symp. (1995) pp. 653–656

31.74 S. Kück: Laser-related spectroscopy of ion-doped crystals for tunable solid-state lasers, Appl. Phys. B **72**, 515–562 (2001)

31.75 Y. Kim, A. Ballato: Force-frequency effects of Y-cut langanite and Y-cut langatate, IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 1678–1682 (2003)

31.76 J.A. Kosinski, R.A. Pastore Jr., X. Yang, J. Yang, J.A. Turner: Stress-induced frequency shifts in langasite thickness-mode resonators, Proc. IEEE Int. Freq. Control Symp. (2003) pp. 716–722

31.77 C. Klemenz, M. Berkowski, B. Deveaud-Pledran, D.C. Malocha: Defect structure of langasite-type crystals: a challenge for applications, Proc. IEEE Int. Freq. Control Symp. (2002) pp. 301–306

31.78 R. Fachberger, T. Holzheu, E. Riha, E. Born, P. Pongratz, H. Cerva: Langasite and langatate nonuniform material properties correlated to the performance of SAW-devices, Proc. IEEE Int. Freq. Control Symp. (2001) pp. 235–239

31.79 C.F. Klemenz, J. Luo, D. Shah: High-quality 2 inch $La_3Ga_{5.5}Ta_{0.5}O_{14}$ and $Ca_3TaGa_3Si_2O_{14}$ crystals for oscillators and resonators, Adv. Electron. Ceram. Mater. Ceram. Eng. Sci. Proc. **26**, 169–176 (2005)

31.80 B.V. Mill, Y.V. Pisarevshy, E.L. Belokoneva: Synthesis, growth, and some properties of single crystals with the $Ca_3Ga_2Ge_4O_{14}$ structure, Proc. Jt. Meet. Eur. Freq. Time Forum IEEE Int. Freq. Control Symp. (1999) pp. 829–834

31.81 B.V. Mill, M.V. Lomonosov: Two new lines of langasite family compositions, Proc. IEEE Int. Freq. Control Symp. (2001) pp. 255–262

31.82 V.I. Chani, K. Shimamura, Y.M. Yu, T. Fukuda: Design of new oxide crystals with improved structural stability, Mater. Sci. Eng. R **20**, 281–338 (1997)

31.83 C. Klemenz: Liquid-phase epitaxy of $La_3Ga_5SiO_{14}$, Proc. Eur. Freq. Time Forum (2001) pp. 42–45

31.84 C. Klemenz: High-quality langasite films grown by liquid-phase epitaxy, J. Cryst. Growth **237**, 714–719 (2002)

31.85 C. Klemenz: High-quality $La_3Ga_{5.5}Ta_{0.5}O_{14}$ and $La_3Ga_{5.5}Nb_{0.5}O_{14}$ LPE films for oscillators and resonators, J. Cryst. Growth **250**, 34–40 (2003)

31.86 C.F. Klemenz, A. Sayir: In-situ Al:Ti-co-doped $La_3Ga_{5.5}Ta_{0.5}O_{14}$ films for high-Q resonators, Proc. IEEE Int. Freq. Contol Symp. (2006) pp. 676–680

31.87 H. Amano, M. Kito, K. Hiramatsu, I. Akasaki: P-type conduction in Mg-doped GaN treated with low-energy electron beam irradiation (LEEBI), Jpn. J. Appl. Phys. **28**, L2112–L2114 (1989)

31.88 F.A. Ponce: Defects and interfaces in GaN epitaxy, MRS Bull. **22**, 51–57 (1997)

31.89 S. Nakamura, M. Senoh, S. Nagahama, N. Iwasa, T. Yamada, T. Matsushita, Y. Sugimoto, H. Kiyoku: Subband emissions of InGaN multi-quantum-well laser diodes under room-temperature continuous wave operation, Appl. Phys. Lett. **70**, 2753–2755 (1997)

31.90 T. Detchprohm, T. Kuroda, K. Hiramatsu, N. Sawaki, H. Goto: The selective growth in hydride vapor phase epitaxy of GaN, Inst. Phys. Conf. Ser. **142**, 859–862 (1996)

31.91 T.S. Zheleva, O. Nam, M.D. Bremser, R.F. Davis: Dislocation density reduction via lateral epitaxy in selectively grown GaN structures, Appl. Phys. Lett. **71**, 2472–2474 (1997)

31.92 V.Y. Davydov, A.A. Klochikhin, R.P. Seisyan, V.V. Emtsev, S.V. Ivanov, F. Bechstedt, J. Furthmüller, H. Harima, V. Mudryi, J. Aderhold, O. Semchinova, J. Graul: Absorption and emission of hexagonal InN. Evidence of narrow fundamental band gap, Phys. Status Solidi (b) **229**, R1–R3 (2002)

31.93 J. Wu, W. Walukiewicz, K.M. Yu, J.W. Ager, E.E. Haller, H. Lu, W.J. Schaff, Y. Saito, Y. Nanishi: Unusual properties of the fundamental bandgap of InN, Appl. Phys. Lett. **80**, 3967 (2002)

31.94 J. Wu, W. Walukiewicz, W. Shan, K.M. Yu, J.W. Ager, E.E. Haller, H. Lu, W.J. Schaff: Effects of the narrow band gap on the properties of InN, Phys. Rev. B **66**, 201403 (2002)

31.95 T.L. Tansley, C.P. Foley: Optical bandgap of indium nitride, J. Appl. Phys. **59**, 3241–3244 (1986)

31.96 J. Wu, W. Walukiewicz, K.M. Yu, J.W. Ager, E.E. Haller, H. Lu, W.J. Schaff: Small band gap bowing in $In_{1-x}Ga_xN$ alloys, Appl. Phys. Lett. **80**, 4741–4743 (2002)

31.97 J. Wu, W. Walukiewicz, K.M. Yu, W. Shan, J.W. Ager, E.E. Haller, H. Lu, W.J. Schaff, W.K. Metzger, S. Kurtz: Superior radiation resistance of $In_{1-x}Ga_xN$ alloys: Full-solar-spectrum photovoltaic material system, J. Appl. Phys. **94**, 6477–6482 (2003)

31.98 R.A. Logan, C.D. Thurmond: Heteroepitaxial thermal gradient solution growth of GaN, J. Electrochem. Soc. **119**, 1727–1734 (1972)

31.99 C. Klemenz, H.J. Scheel: Crystal growth and liquid-phase epitaxy of GaN, J. Cryst. Growth **211**, 62–67 (2000)

31.100 G. Sun, E. Meissner, P. Berwian, G. Müller, J. Friedrich: Study on the kinetics of the formation reaction of GaN from Ga-solutions under ammonia atmosphere, J. Cryst. Growth **305**, 326–334 (2007)

31.101 O. Glemser, M. Field, K. Kleine-Weischeide: Z. Anorg. Chem. **332**, 257–259 (1964)

31.102 R.M. Yonco, E. Veleckis, V.A. Maroni: Solubility of nitrogen in liquid lithium and thermal decomposition of solid Li_3N, J. Nucl. Mater. **57**, 317–324 (1975)

31.103 W.J. Wang, X.L. Chen, Y.T. Song, W.X. Yuan, Y.G. Cao, X. Wu: Assessment of Li-Ga-N ternary system and GaN single crystal growth, J. Cryst. Growth **264**, 13–16 (2004)

31.104 M. Yano, M. Okamoto, Y.K. Yap, M. Yoshimura, Y. Mori, T. Sasaki: Growth of nitride crystals, BN, AlN and GaN by using a Na flux, Diam. Relat. Mater. **9**, 512–515 (2000)

31.105 H. Yamane, M. Shimada, S.J. Clarke, F.J. DiSalvo: Preparation of GaN single crystals using a Na flux, Chem. Mater. **9**, 413–416 (1997)

31.106 F. Kawamura, T. Iwahashi, M. Morishita, K. Omae, M. Yoshimura, Y. Mori, T. Sasaki: Growth of transparent: Large size GaN single crystal with low dislocations using Ca-Na flux system, Jpn. J. Appl. Phys. **42**, L729–L731 (2003)

31.107 F. Kawamura, H. Umeda, M. Kawahara, M. Yoshimura, Y. Mori, T. Sasaki, H. Okado, K. Arakawa, H. Mori: Drastic decrease in dislocations during liquid-phase epitaxy growth of GaN single crystals using Na flux method without any artificial processes, Jpn. J. Appl. Phys. **45**, 2528–2530 (2006)

31.108 C.O. Dugger: The synthesis of aluminum nitride single crystals, Mater. Res. Bull. **9**, 331–336 (1974)

31.109 T. Fukuda, D. Ehrentraut: Prospects for the ammonothermal growth of large GaN crystal, J. Cryst. Growth **305**, 304–310 (2007)

31.110 R. Dwiliński, R. Doradziński, J. Garczyński, L. Sierzputowski, J.M. Baranowski, M. Kamińska: AMMONO method of GaN and AlN production, Diam. Relat. Mater. **7**, 1348–1350 (1998)

31.111 T. Hashimoto, K. Fujito, M. Saito, J.S. Speck, S. Nakamura: Ammonothermal growth of GaN on an over-1-inch seed crystal, Jpn. J. Appl. Phys. **44**, L1570–L1572 (2005)

31.112 B. Wang, M.J. Callahan, K.D. Rakes, L.O. Bouthillette, S.-Q. Wang, D.F. Bliss, J.W. Kolis: Ammonothermal growth of GaN crystals in alkaline solutions, J. Cryst. Growth **287**, 376–379 (2006)

31.113 D. Ehrentraut, N. Hoshino, Y. Kagamitani, A. Yoshikawa, T. Fukuda, H. Itoh, S. Kawabata: Temperature effect of ammonium halogenides as mineralizers on the phase stability of gallium nitride synthesized under acidic ammonothermal conditions, J. Mater. Chem. **17**, 886–893 (2007)
31.114 C. Klemenz: unpublished

32. Molecular-Beam Epitaxial Growth of HgCdTe

James W. Garland, Sivalingam Sivananthan

Epitaxial HgCdTe grown by molecular-beam epitaxy (MBE) is the material of choice for advanced infrared (IR) detection and imaging devices. Its bandgap is easily tunable over the entire IR range with only very small changes in lattice constant, offering the possibility of multilayer device structures and thus an unlimited choice of device designs, and it yields devices with quantum efficiencies as high as 0.99. Despite a number of unresolved challenges in achieving its ultimate promise for industrial application, the great achievements in the MBE growth of HgCdTe are made evident by its routine use in the industrial manufacture of focal-plane arrays (FPAs). MBE growth can be continuously monitored in situ by reflection high-energy electron diffraction, spectroscopic ellipsometry (SE), and other characterization tools, providing instantaneous feedback on the influence of growth conditions on film structure. This allows the growth of a large range of unique structures such as superlattices (SLs), quantum well devices, lasers, and advanced design devices such as multicolor and high-operating-temperature IR sensors and focal-plane arrays. This chapter considers the theory and practice of MBE growth of HgCdTe and HgTe/CdTe superlattices and the use of HgCdTe in IR devices, emphasizing such incompletely resolved issues as the choice and preparation of substrates, dislocation reduction, p-doping, and the uses of SE.

The theory of MBE growth is summarized briefly in Sect. 32.2, followed by a lengthy discussion of substrate-related issues in Sect. 32.3, including a summary of the relative merits and demerits of different substrate materials. The growth hardware is discussed very briefly in Sect. 32.4, followed by a discussion of the in situ characterization tools used for monitoring and control of the growth in Sect. 32.5 and of the growth procedure for HgCdTe in Sect. 32.6. A discussion of the doping of HgCdTe, including the serious issues still surrounding p-type doping, is given in Sect. 32.7. The properties achievable in MBE-grown HgCdTe are summarized in Sect. 32.8, with emphasis on the types of defects common in MBE-grown material, their effects on device performance, and possible methods to reduce the present defect densities. The use of MBE-grown HgTe/CdTe SLs for IR absorbers in lieu of HgCdTe alloy material is considered in Sect. 32.9. Finally, a brief discussion of the devices enabled by the MBE growth of HgCdTe and of their fabrication is given in Sects. 32.10 and 32.11, and a brief concluding summary of the chapter is given in Sect. 32.12.

32.1 **Overview** 1070
- 32.1.1 Why HgCdTe Is Important 1071
- 32.1.2 Why MBE Is the Preferred Method of Growth for HgCdTe IR Detectors and Imagers 1072
- 32.1.3 General Description of the MBE Growth Technique 1072

32.2 **Theory of MBE Growth** 1073
- 32.2.1 Pseudo-Equilibrium Theories 1074
- 32.2.2 Kinetic Theories 1075

32.3 **Substrate Materials** 1076
- 32.3.1 Substrate Orientation 1077
- 32.3.2 CdZnTe Substrates 1077
- 32.3.3 Si-Based Substrates 1084
- 32.3.4 Other Substrates 1087

32.4 **Design of the Growth Hardware** 1088
- 32.4.1 Mounting of the Substrate 1088
- 32.4.2 Valving of the Effusion Cells 1089

32.5 **In situ Characterization Tools for Monitoring and Controlling the Growth** 1090
- 32.5.1 Spectroscopic Ellipsometry (SE): Basic Theory and Experimental Setup 1090
- 32.5.2 SE Data Analysis 1092

32.5.3 SE Study of Hg Absorption and Adsorption on CdTe 1098
32.5.4 Correlation Between the Quality of MBE-Grown HgCdTe and the Depolarization and Surface Roughness Coefficients Measured by in situ SE 1099
32.5.5 Surface Characterization by in situ RHEED 1100
32.5.6 Other in situ Tools for Controlling the Growth 1101
32.6 Nucleation and Growth Procedure 1101
32.6.1 Nucleation and Growth of CdTe or ZnTe on Si 1101
32.6.2 Substrate Preparation and Growth of HgCdTe 1102
32.7 Dopants and Dopant Activation 1104
32.7.1 Extrinsic n-Type Doping 1104
32.7.2 Extrinsic p-Type Doping 1105
32.7.3 In situ Group I Dopant Incorporation 1106
32.7.4 In situ Group V Dopant Incorporation 1106
32.7.5 As Activation 1106
32.8 Properties of HgCdTe Epilayers Grown by MBE 1107
32.8.1 Electrical and Optical Properties 1107
32.8.2 Structural Properties 1110
32.8.3 Surface Defects 1110
32.9 HgTe/CdTe Superlattices 1112
32.9.1 Theoretical Properties 1113
32.9.2 Growth 1114
32.9.3 Experimentally Observed Properties 1114
32.10 Architectures of Advanced IR Detectors 1115
32.10.1 Reduction of Internal Detector Noise 1116
32.10.2 Increasing Detector Response 1116
32.10.3 High-Speed IR Detectors 1117
32.10.4 High-Operating-Temperature (HOT) IR Detectors 1118
32.11 IR Focal-Plane Arrays (FPAs) 1118
32.12 Conclusions 1119
References 1121

Molecular-beam epitaxy (MBE) is a process for growing thin epitaxial films of a wide variety of materials, ranging from oxides to semiconductors to metals. Because of their high technological value, its first and still most common application is for the growth of compound semiconductors and their alloys. In MBE beams of atoms or molecules are incident in an ultrahigh vacuum upon a heated crystal processed to produce a nearly atomically clean, smooth surface. The arriving atoms form a crystalline layer in registry with the substrate, i. e., an epitaxial film. MBE allows doping levels and alloy compositions to be controlled precisely and changed rapidly, producing almost atomically abrupt homojunctions or heterojunctions. In 1969 *Cho* [32.1] published landmark results reporting the first in situ observations of the MBE growth process using high-energy electron diffraction. He demonstrated that MBE growth could produce atomically flat, ordered layers; thus these studies marked the beginning of the use of MBE for practical device fabrication. This structural analysis capability proved to be crucial for characterizing MBE epitaxy because it provided instantaneous feedback on the influence of growth conditions on film structure. That allowed the growth of a large range of unique structures such as superlattices (SLs), quantum well devices, lasers, and advanced design devices such as multicolor and high-operating-temperature IR sensors and focal-plane arrays.

MBE-grown HgCdTe is the material of choice for advanced infrared (IR) detection and imaging devices. Its bandgap is easily tunable over the entire IR range with only very small changes in lattice constant, it offers the possibility of multilayer device structures and thus an unlimited choice of device designs, and it yields devices with quantum efficiencies as high as 0.99. Despite a number of unresolved challenges in achieving its ultimate promise for industrial application, the great achievements in the MBE growth of HgCdTe are made evident by its routine use in the industrial manufacture of focal-plane arrays (FPAs).

32.1 Overview

The MBE growth technique for HgCdTe, the in situ characterization techniques used in its growth, the present status of achievable HgCdTe material quality, and its applications to IR detection are reviewed here.

Particular emphasis is placed on the issues remaining to be solved before MBE-grown HgCdTe can reach its full potential as an ideal material for IR devices. These issues include substrate choice and preparation, the p-doping of HgCdTe, and the reduction of material defect densities. Recent progress in addressing those issues is discussed, as are techniques still under investigation for addressing those issues. Except for points not documented elsewhere, detailed discussions are left to the references.

32.1.1 Why HgCdTe Is Important

Infrared detectors fall into two broad categories: photon and thermal detectors. Photon detectors offer much higher detection speeds and sensitivities as well as the ability to differentiate between objects at different temperatures using multicolor or hyperspectral detectors and to form high-resolution images using FPAs. These properties are essential for almost all military and space applications. Mercury cadmium telluride has been of great interest as a material for infrared (IR) detection since it was first grown almost 50 years ago and has dominated the market for more than two decades [32.2]. It is the only semiconductor system in which the optical absorption edge can be made to vary across the entire IR spectrum with an almost constant lattice parameter. The 300 K lattice constant changes only from 6.4614 Å for HgTe with a -0.14 eV bandgap to 6.4825 Å for CdTe with a 1.56 eV bandgap. This allows the high performance of photodiodes over a much larger range than InGaAs, for example, and allows the epitaxial growth of layers with different optical cutoffs on top of one another with very little strain, $\approx 0.1\%$, which in turn, along with the low-temperature MBE growth method, allows the fabrication of advanced detectors and FPAs, such as multicolor and high-operating-temperature (HOT) FPAs, and allows the growth of HgTe/CdTe SLs, either for very long-wavelength infrared (VLWIR) detection or as buffer layers. In addition to this great advantage, HgCdTe offers two other overwhelming advantages that explain its dominance in the IR marketplace. Both the thermal generation rate per unit volume of material and the quantum efficiency for photon absorption in the IR are higher in HgCdTe than in any competing material. From a fundamental point of view only the diffusion-limited operation of a photovoltaic diode need be considered. In that case only these two properties are important in determining the maximum useful operating temperature, as discussed by *Kinch* [32.3] and outlined below.

The following four different types of materials systems have been serious players in the IR systems marketplace:

1. Direct-bandgap semiconductors (minority-carrier transport):
   a) Binary III–V alloys such as InSb-based alloys
   b) Ternary II–VI alloys (tunable-bandgap HgCdTe)
   c) II–VI tunable-bandgap type I SLs (HgTe/CdTe)
   d) Type II and III SLs such as InAs/GaInSb.
2. Extrinsic semiconductors such as Si:Ga or Ge:Hg (majority-carrier transport)
3. Type I SL quantum-well infrared photodetectors (QWIPs) such as GaAs/AlGaAs QWIPs (majority-carrier transport)
4. Silicon Schottky barriers such as PtSi or IrSi (majority-carrier transport).

For the reasons discussed in the previous paragraph, HgCdTe is the dominant direct-bandgap semiconductor material for applications in all regions of the IR spectrum except the VLWIR. Even there, either HgCdTe or HgTe/CdTe SLs are expected to become strong or even dominant players.

In the long-wavelength IR (LWIR) spectral band the normalized detector dark current is four orders of magnitude smaller for direct-bandgap semiconductors such as HgCdTe than for any of the other three types of systems, so that direct-bandgap semiconductors such as HgCdTe give the highest background-limited performance (BLIP) operating temperature by far. This is extremely important because cryogenic cooling is expensive and greatly increases the weight and bulk of IR detection and imaging systems. The sensitivity of a detector can be expressed in a variety of ways, but the limiting factor in all of them is the fluctuation in the relevant carrier concentration, as this determines the minimum observable signal. Ideally, neglecting the background incident flux, the most commonly used measure of the detectivity D^* depends only on the properties of the IR material in the detector. It is two or more orders of magnitude higher for HgCdTe than for any of the other three types of systems. Although material growth issues have prevented HgCdTe from reaching its full potential, it is the material of choice for most detector systems. The only important advantage of III–V alloys and SL QWIPs is their lower cost arising from the preexisting large industrial infrastructure in III–V materials and device growth, processing, and packaging. For military systems, which dominate the market, the increasing requirements of succeeding generations of IR detection and imaging systems will further enhance

the dominance of HgCdTe in the IR materials marketplace. No other material, even in principle, will be able to meet the requirements for higher operating temperatures (and ultimately totally uncooled operation), larger-area FPAs, faster frame rates, and multispectral operation with no increase, or even a reduction, in cost.

32.1.2 Why MBE Is the Preferred Method of Growth for HgCdTe IR Detectors and Imagers

Since the growth of HgCdTe by MBE was first reported in 1982 [32.4], more than 15 groups have been established worldwide for the growth of Hg-based alloys and SLs by MBE or metalorganic MBE (MOMBE) for IR device applications. Since that time enormous progress has been made in reducing defect densities, increasing growth areas, improving the control of alloy compositions and doping levels, and demonstrating increasingly sophisticated and powerful IR electrooptical devices, as discussed in later sections of this chapter. Molecular-beam epitaxy has many advantages over other methods for the epitaxial growth of HgCdTe [32.5]:

1. The MBE growth temperature for HgCdTe is only $\approx 185\,^{\circ}\mathrm{C}$, far below the liquid-phase epitaxy (LPE) and metalorganic vapor-phase epitaxy (MOVPE) (also called metalorganic chemical vapor deposition (MOCVD)) growth temperatures of $\approx 450\,^{\circ}\mathrm{C}$ and $300–400\,^{\circ}\mathrm{C}$, respectively. This offers several great advantages. It greatly reduces impurity outdiffusion from the substrate, it allows the formation of precisely controlled, atomically sharp or graded hetero- or homojunctions, and it allows the growth of SLs.
2. MBE is ideally suited for the deposition of multilayer structures; in addition to having only diffuse junctions, heterostructures grown by LPE usually are limited to two or three layers because different melts are required for the growth of layers having different compositions or dopings. MBE is the only technique suited to the growth of SLs and the only technique allowing in situ growth of CdTe and ZnS passivation layers for devices.
3. MBE growth allows the use of spectroscopic ellipsometry (SE), reflection high-energy electron diffraction (RHEED), and other in situ analytical characterization techniques to monitor and control the growth.
4. Unlike LPE, no aggressive medium is present in MBE growth. Combined with the low growth temperature and analytical facilities for surface control, this allows the use of alternative lower cost, more robust, larger size, more attainable substrates such as CdTe/Si with improved matching of the substrate and readout system thermal expansion coefficients.
5. In MBE no contamination of interfaces occurs because one can grow different compounds in the same growth run in the same system under ultrahigh vacuum.

In 1983 the first photovoltaic device fabricated from MBE-grown HgCdTe was reported [32.6]. By 1985 the quality of MBE-grown HgCdTe layers was comparable to that of the best layers grown by other techniques [32.7]. By 1997 the performance of IR FPAs made from MBE-grown material was comparable to that of those made from LPE materials [32.8]. Although materials-related HgCdTe problems remain, progress in MBE technology continues to outstrip that in more mature techniques such as LPE. Meanwhile, techniques closely related to conventional MBE and at first appearing to offer advantages have largely been abandoned: metalorganic MBE (MOMBE) [32.9, 10] because of its higher growth temperature and because of the ease of carbon contamination and void formation, and photo-assisted MBE (PAMBE) [32.11], which offered higher-quality growth and the as-grown activation of dopants with no self-compensation [32.12], because it was limited to extremely small-area growth.

32.1.3 General Description of the MBE Growth Technique

The reader is referred to [32.13, 14] for detailed and extensive reviews of the physical principles, equipment, and general procedures of MBE growth. Only a very brief general description is given here; specifics of the MBE growth of HgCdTe are discussed in Sects. 32.3–32.7. MBE uses localized beams of atoms or molecules in an ultrahigh-vacuum (UHV) environment as the source of the constituents to the growing surface of a substrate crystal. The UHV environment minimizes contamination of the growing surface. The beam atoms and molecules travel in nearly collision-free paths until arriving either at the substrate or at chilled walls of the chamber, where they condense and are effectively removed from the system. The substrate is kept at a moderately elevated temperature to provide sufficient thermal energy to the arriving atoms for them to migrate over the surface to lattice sites. When a shutter is interposed in a beam, the beam is effectively turned off almost instantly. These features make it possible

to grow the films very slowly without contamination, and to change the composition of the arriving stream of atoms very abruptly; in fact, the composition of the flux can be changed in times much shorter than that needed to grow a single atomic layer of the film.

The vacuum environment surrounding the growing crystal must be kept as low as possible to avoid contamination that might affect surface morphology and the electrical properties of the epilayer grown. Because the growth rate is low, $\approx 1-3$ ML/s, to keep the arrival rate of a background species at 10^{-6} of the growth rate would require a partial pressure of $\approx 10^{-12}$ Torr; however, fortunately many of the background species have low reactivities (low sticking coefficients). For III–V growth, the best background pressures achievable are $\approx 10^{-11}$ Torr (except in brand-new growth chambers), and pressures commonly are $\approx 10^{-10}$ Torr; for HgCdTe growth the achievable pressures are ≈ 100 times higher because of Hg contamination. Typically substrates are loaded into the growth chamber via a load-lock chamber, where they are outgassed while the growth chamber remains under vacuum. Then, the vacuum is broken only in the small loading chamber, so only it needs to be evacuated. II–VI and III–V semiconductors should not be grown in the same chamber because the residual pressure from one would dope the other.

The growth chamber must include source ovens, beam shutters and their actuating mechanisms, a substrate holder and heater, in situ growth characterization tools, beam flux monitors, and cryopanels to act as cryopumps and to condense unused beam flux. All of these must be designed to minimize outgassing. The chamber must be designed to keep the substrate at a uniform well-controlled temperature, to provide uniform well-controlled fluxes to the substrate, to avoid temperature transients on the growth surface due to the shuttering and unshuttering of the atomic beams. Variations in the flux ratios and/or the temperature of the growing surface lead to compositional variations and even non-stoichiometry, which can destroy the crystalline quality of the growth. Typically, chambers are custom-designed for each type of material to be grown.

32.2 Theory of MBE Growth

Theories of MBE growth can be classified either as quasi-equilibrium thermodynamic theories or as kinetic growth theories. Although MBE growth typically is a far-from-equilibrium process, quasi-equilibrium theories have been found to yield valuable results even in cases in which desorption is negligible, so that the solid and vapor phases are far from equilibrium with one another. Only a partial equilibrium of the solid phase, not a global quasi-equilibrium, is required for the validity of most quasi-equilibrium theories. On the other hand kinetic theories are necessary to understand important aspects of MBE growth such as the surface morphology, as well as in determining the limits of validity of quasi-equilibrium thermodynamic theories. The structural quality of MBE-grown films depends largely on whether partial, or at least local, thermodynamic equilibrium is established on a time scale appropriate to the growth. If local thermodynamic equilibrium is established, one may use the usual notions of absolute rate theory and the resulting rate expressions for kinetic processes when describing MBE growth [32.15, 16]. However, whether the combined effect of individual kinetic processes is able to move the growing film toward a partial equilibrium state and to reach this state in a time scale relevant to the growth depends on the growth conditions and the specific nature of the material being grown. Under some conditions, local thermodynamic equilibrium may exist during the growth, but the kinetics may not be fast enough to ensure partial equilibrium across the entire growth surface or even macroscopic areas of that surface. The microstructural and even the chemical properties of a film grown under only local equilibrium conditions can be very different from those of the film if it were grown under partial equilibrium conditions.

The basic MBE growth models have been reviewed in detail by *Tsao* [32.14]. The driving forces for MBE growth are the difference $\Delta\mu^{\alpha}$ between the chemical potential μ^{α} of the vapor of the α element and its equilibrium chemical potential for sublimation $\mu^{\alpha}_{\mathrm{eq}}$. For many elements desorption is negligible in MBE; on the other hand, the desorption rate of Hg is orders of magnitude higher than its absorption rate in the MBE growth of HgCdTe. This implies that $\Delta\mu^{\mathrm{Hg}} \ll k_{\mathrm{B}}T$ at the growth temperature. For growth on a planar surface there exist three basic MBE growth modes. First, there is the layer-by-layer, two-dimensional or Frank–van der Merwe growth mode, in which each atomic layer tends to be completed before the next layer is started. Second, there is the three-dimensional or Volmer–Weber mode, in which hillocks and cavities form on the surface. Third is Stranski–Krastanov growth, in which the

growth begins as two dimensional and then becomes three dimensional. In fact, MBE growth of HgCdTe occurs on vicinal surfaces, with a well-defined miscut to a surface with low Miller indices. In this mode the growth surface is characterized by an array of steps between smooth terraces. The adatoms preferentially bond at the bottoms of the steps and the crystal grows as the steps advance across the growth surface with a velocity proportional to $\Delta\mu$ for the controlling chemical species (Te for HgCdTe) (step-flow growth).

A phenomenon of particular interest is the self-assembly of islands to form a regular array of quantum dots. To a first approximation, this can be understood as a form of growth in the Stranski–Krastanov regime. According to this point of view, based on general thermodynamic arguments, the deposited material starts to grow as a homogeneous wetting layer, because it has a lower surface energy than the substrate. After a critical thickness, the growth of three-dimensional islands that allow for partial elastic strain relaxation becomes energetically preferable to the two-dimensional film growth. These three-dimensional islands grow until they have reached an optimum size [32.17, 18]. Kinetics enters in this theory only as an external parameter, the island density, which is determined during the nucleation phase and assumed to remain constant during three-dimensional growth of the islands.

32.2.1 Pseudo-Equilibrium Theories

The general principles of the quasichemical thermodynamic theory were presented as early as the 1930s [32.19–24]. The formalism usually used was first given in 1956 [32.25]. The validity of the theory, which rests almost solely on the law of mass action, is unquestioned for thermodynamic equilibrium. However, by definition, thermodynamic equilibrium does not exist during crystal growth. Therefore, it is interesting to consider the past successes and failures of the theory in modeling MBE growth in terms of the degree of equilibrium of the growth. In the theory, one derives from any existing equilibrium condition a corresponding relationship between the probability of occurrence of some configuration of atoms, vacancies, etc. in the solid and the partial pressures of the gases with which the solid is approximately in equilibrium. For approximate global equilibrium during growth two requirements must be met. First, the sticking coefficient of that species must be very small, so that the ratio of the incident flux to the exeunt flux is very nearly one. Second, any intermediate-state reactions must either be sufficiently fast not to hinder thermodynamic equilibrium or so slow that they take place only after absorption is complete, i. e., in the bulk below the growth surface. These two requirements suffice for the valid application of the theory to MBE growth. However, it has been shown that the application of the theory to MBE growth correctly describes the growth even in the absence of global equilibrium, provided that the second requirement holds.

As early as 1958, equilibrium thermodynamic arguments were used [32.26] to explain the different modes of epitaxial growth in terms of the surface energies at the epilayer–vacuum (σ), the epilayer–substrate (σ_i), and the substrate–vacuum (σ_s) layer interfaces. In partial equilibrium, layer-by-layer growth should be expected when $\sigma_s > \sigma + \sigma_i$, that is, when the change in the surface energy $\Delta\sigma = \sigma + \sigma_i - \sigma_s$ accompanying the deposition process is negative. When $\sigma_s < \sigma + \sigma_i$, or $\Delta\sigma > 0$, the film will grow as isolated islands. Stranski–Krastanov growth takes place when $\Delta\sigma$ changes sign from negative to positive because of strain energy after some characteristic thickness. This theory of course requires the atoms on the growing surface to have a high enough mobility to find the lowest-energy sites to occupy. Thus, it typically is valid only at the high end of the window of desirable growth temperatures.

Many authors have given first-principles thermodynamic treatments of the MBE growth of III–V compounds and alloys in agreement with experiment, even though their MBE growth is far from equilibrium for all chemical species involved. For example, the dependence of the desorption rate of Ga in $Ga_{1-x}Al_xAs$ on x and on the growth temperature was established [32.27], and it was shown [32.28–30] that dopant incorporation in the MBE growth of GaAs occurs in accord with the theory. Also, the growth rates and compositions of a series of III–V compounds and alloys have been calculated, and results in agreement with experiment were found [32.31]. Additionally, the In desorption rate in MBE-grown AlInAs alloys has been calculated as a function of the As beam flux and substrate induced strain, again in agreement with experiment [32.32, 33]. Finally, the surface segregation of base elements and impurities in MBE-grown II–V compounds has been described [32.34].

Of more interest for this review, several authors have applied equilibrium thermodynamic theory to the growth of HgCdTe with useful results. *Gailliard* [32.35] found that MBE growth of CdTe, HgTe, and HgCdTe

is correctly described by the theory, and successfully calculated growth rates and compositions, as well as determining the range of acceptable growth conditions, all with no adjustable parameters. *Colin* and *Skauli* [32.36] used the theory to construct a general model for the optimization of the MBE growth conditions for HgCdTe. They calculated an optimal growth temperature as a function of Hg flux, Cd concentration, and growth rate. They also suggested that the growth of HgCdTe is improved by growing on a Te-rich surface, which is now common practice. *Vydyanath* et al. [32.37] used equilibrium thermodynamics to determine the Cd composition in In-doped HgCdTe as a function of the In and Te source temperatures. *Piquette* et al. [32.38] modified the results of Colin and Skauli by forcing the population of Hg atoms on the growth surface to be proportional to the HgTe fractional growth rate, finding a new formula for the optimized growth temperature that experimentally led to consistently lower defect densities. *Chang* et al. [32.39] have used that result to show that all thermodynamically equivalent growth conditions near optimized conditions lie on a straight line with slope −5385 in a graph of $\ln[p_{\mathrm{Hg}}/\nu(1-x)]$ versus T^{-1}, where p_{Hg} is the pressure read by the Hg flux gauge, ν is the growth rate, and x is the Cd concentration. Equilibrium thermodynamics has also been applied to study the density of Te-condensation-related voids as a function of Hg flux [32.10] and to determine that As dopants from an uncracked source are incorporated primarily as As_4 molecules [32.40], rather than being dissociated on the growth surface.

32.2.2 Kinetic Theories

The MBE growth of semiconductors is performed at relatively low temperatures and slow growth rates in comparison with LPE and vapor-phase epitaxy (VPE). Low growth temperatures result in growth mechanisms that are strongly dominated by surface kinetics and chemistry. As the source molecular beams are introduced into the MBE chamber some of these molecular species reach the substrate and become physisorbed or weakly chemisorbed to the surface. The adsorbed molecules can then migrate to energetically favorable lattice sites, step edges, nucleated islands, etc. and bond there. A thermodynamic redistribution near the surface layers then establishes the final configuration.

MBE growth depends on the sticking coefficients and surface kinetics of atomic or molecular beams impinging on a suitable substrate. The importance of kinetics in the MBE growth of HgCdTe is demonstrated by the strong dependence of the Hg sticking coefficient [32.41–43] and the quality of growth [32.43, 44] on substrate orientation. One can identify a number of parameters relating to the kinetics of growth; they include the flux F of the impinging species, the sticking coefficient S, the desorption rate R, substrate orientation, substrate temperature, etc. A variety of kinetic rate equation atomistic models [32.45] have been explored for growth from the vapor phase on a planar substrate. In the case of epitaxial HgCdTe MBE growth, surface kinetics place strict boundaries on optimal deposition conditions. It has been found [32.46] that surface kinetics limits the Hg flux, the substrate orientation, and the substrate temperature during MBE HgCdTe growth.

Various kinetics problems have been treated; a few examples follow. A model for the evolution of the profile of a growing interface and a Langevin equation (the Kardar–Parisi–Zhang (KPZ) equation) for the local growth of the profile have been proposed [32.47]. It has been argued [32.48] that the KPZ equation is not the relevant continuum equation for MBE growth at intermediate temperatures, where overhangs and bulk defects can be neglected. A new nonlinear growth equation was suggested as the relevant dynamical equation for ideal MBE growth at intermediate to high temperatures. When desorption and overhangs occur, the KPZ equation remains the relevant equation describing MBE growth.

The effect of finite size on smooth layer-by-layer MBE growth has been examined, and it was concluded that, for finite system sizes, growth can always be smooth as long as the diffusion length is comparable to the system size even though the asymptotic growth may be rough in the thermodynamic limit [32.49]. The heteroepitaxy of InAs on GaAs(001) has been discussed, and first-principles density-functional theory calculations for In diffusion on the strained GaAs substrate have been presented [32.50]. In particular, the effect of heteroepitaxial strain on the growth kinetics of coherently strained InAs islands was considered. The strain field around an island was found to cause a slowing down of material transport from the substrate towards the island, and thus to help achieve more homogeneous island sizes.

32.3 Substrate Materials

As with all epitaxial systems and techniques, the structural quality of HgCdTe epilayers deposited by MBE is inextricably linked to the quality of the substrate. The absence of any substrate material satisfactory in all respects for the MBE growth of HgCdTe is perhaps the primary factor limiting the performance of IR detectors and FPAs. In 1993 *Triboulet* et al. [32.51] compared the basic properties and the advantages and disadvantages of different substrate materials. Here, we summarize and update that discussion and consider the question of substrate orientation. Special attention is paid to growth on CdZnTe and CdTe/Si substrates. For CdZnTe emphasis is placed on the nature of substrate bulk and surface imperfections, on their effect on MBE-grown epilayers, and on methods to ameliorate those effects and screen out bad substrate wafers before growth. For CdTe/Si, emphasis is placed on possible methods to overcome or ameliorate the effects of the large lattice mismatch between Si and CdTe, including new methods being developed or under consideration. Several characteristics of the substrates in common use are compared in Table 32.1.

The first MBE growth of HgCdTe [32.4] was performed on a CdTe(111) substrate; early growths also were performed on CdTe(100) [32.6]. However, even the small lattice mismatch between CdTe and $Hg_{1-x}Cd_xTe$ with $x < 0.5$ caused threading dislocations and other crystalline defects. Also the growth on high-symmetry planes allowed the formation of a high density of microtwins in the HgCdTe [32.6]. The twinning can be eliminated by growth on higher-index planes or miscut surfaces, but bulk CdTe was abandoned as a substrate material because of a combination of factors, including the lattice mismatch, its low thermal conductivity, its poor crystalline quality, its brittleness, and its cost. $Cd_{0.96}Zn_{0.04}Te$ has become the substrate of choice for the growth of the highest-quality HgCdTe epilayers because of its in-principle

Table 32.1 Characteristics of the types of substrates currently used for MBE growth of HgCdTe

	$Cd_{1-y}Zn_yTe$ ($y = 0.04$)	GaAs	Si
Cost per cm^2	US$ 400 (high-pressure Bridgman growth)	US$ 0.71 (vertical gradient freeze growth) (US$ 28.00 w/CdTe buffer)	US$ 0.56 (float zone growth) (US$ 27.86 w/CdTe buffer)
Largest commercially available size (cm^2)	49	182.4	5228
Robustness of material	Brittle	Moderately robust	Strong
Vickers hardness (kg/mm^2 at 300 K)	60	360	1150–1330
Thermal conductivity (mW/(cm K))	55	500	1235
Surface preparation	Difficult, sometimes poor	Standardized by large III–V industry	Standardized by large Si industry
Substrate surface defect density (cm^{-2})	10^4	5×10^3 (10^5 on CdTe buffer)	10^2 (10^5 on CdTe buffer)
Lattice mismatch with HgCdTe ($x = 0.2$ at 300 K)	< 1% (<1% w/CdTe)	13.6% (<1% w/CdTe)	19.47%
Substrate etch pit density (cm^{-2})		< 500 on CdTe buffer	< 5000 on CdTe buffer
Thermal mismatch with HgCdTe ($x = 0.2$ at 300 K)	3.53%	27.04% (1.85% w/CdTe)	51.85% (1.85% w/CdTe)
Valence issues	None	Minor	Severe
Other	Impurities, crystalline defects, compositional uniformity problem, large thermal mismatch with Si ROIC	In situ surface preparation may cause contamination	Allows monolithic integration Requires buffer layer

perfect lattice match to IR HgCdTe, but is plagued with many problems. Composite substrates consisting of epilayers of CdTe or $Cd_{0.96}Zn_{0.04}Te$ grown on InSb, sapphire, Ge, GaAs, and Si have been used because of their greater robustness, much lower cost, higher thermal conductivity, greater lateral uniformity, and larger possible areas in comparison with $Cd_{0.96}Zn_{0.04}Te$ substrates. Of these substrates, two are still in common use: CdTe/GaAs and CdTe/Si. However, the very large lattice and thermal mismatches between the underlying substrate material and the top epilayer in these substrates lead to serious problems. Great progress has been made in solving these and other problems associated with each type of substrate still in common use, but for no substrate have the problems yet been completely solved.

32.3.1 Substrate Orientation

Substrate orientation is of primary importance in MBE growth, as was recognized very early for growth on CdTe substrates by noting the strong dependence of the Hg sticking coefficient on orientation [32.41–43] and as is seen from the strong dependence of the quality of growth on substrate orientation [32.43, 44]. It was recognized from the beginning that growth on a low-index CdTe plane such as the (100) or (111) planes led to microtwinning. On alternative substrates such as GaAs or Si growth on the (100) plane led to twinning and double domains rotated by 90° from one another due to the equivalence of different CdTe orientations on such planes. Two approaches were taken to reduce the symmetry of the substrate surface and thus eliminate this problem:

1. To slightly miscut the substrate away from a low-index surface plane [32.52–57]
2. To grow on higher-index planes containing steps between (100) or (111) flat terraces [32.58–61].

Today almost all growth is carried out on (211) substrates. However, the question of the preferred substrate orientation is not as clear for lattice-matched substrates as for alternative substrates. A brief review of the results for growth on substrates with various orientations is given here. In 1988 a study [32.46] of growth on (100), (111)B, (211)A, and (211)B substrates revealed that (211)B growth gives the best surface morphology, suppresses twinning, and gives higher Hg sticking coefficients and thus better control of composition than any other orientation studied except (111)B. In 1994 simultaneous growth was conducted [32.62] on (111)B substrates miscut by 4° towards (110) (labeled (776)) and on (211)B-oriented substrates. These orientations were described as (111)B surfaces tilted 19° towards and 4° away from (100), respectively. The primary difference between these two orientations was the width of the (111) terraces between steps: 53 and 10.6 Å, respectively. It was found that the (776) orientation with broader terraces was more susceptible to twinning at lower growth temperatures. In 2000 it was found [32.63] that growth on (211)B and (311)B surfaces under optimized conditions gives material of comparable quality. Recently, simultaneous MBE growth was conducted [32.44] in the (511)B, (311)B, (211)B, (111)B, and (255)B orientations tilted 15.8, 25.2, 35.3, 54.7, and 74.2° away from the (100) surface in the (011) direction. Over a nominal range of growth temperatures from 180 to 187 °C the x-ray rocking curve full-width at half-maximum (FWHM), and etch pit density (EPD) were measured for all samples, and surface defect densities were measured using Nomarski microscopy. Under the best growth conditions comparable results were obtainable for all orientations, although the (255) orientation allowed the lowest-temperature growth.

32.3.2 CdZnTe Substrates

A golden rule of epitaxial growth is to match the lattice parameter of the substrate to that of the epitaxial layer in order to avoid strain and dislocation formation in the epilayer. Although, because of its lattice match to HgCdTe, bulk CdZnTe is the substrate of choice for the MBE growth of the highest-quality material, especially for LWIR and VLWIR detection, it suffers from both intrinsic limitations and quality concerns, as listed in Table 32.1. Many of the intrinsic limitations – the very high cost, limited area, brittleness and large thermal mismatch with the Si readout integrated circuit (ROIC) in devices – do not affect the quality of MBE growth but do impact the practicality of use of the HgCdTe layers grown for device applications. The poor thermal conductivity and quality concerns – crystalline defects, impurities, nonuniform composition, surface roughness, imperfect surface flatness, and stoichiometry – all impact the quality of the HgCdTe grown on the substrates. Furthermore, all are amplified in importance by great variations in quality from substrate to substrate. These variations introduce great variations in HgCdTe material and thus both significantly lower yield in device production and make it difficult to identify optimal

growth conditions; see, for example, [32.44]. For those reasons it is desirable to develop and use methods to screen substrates before the MBE growth of HgCdTe.

Effects of Poor Thermal Conductivity on HgCdTe Growth

Because the growth is highly noncongruent, the temperature window for the optimal growth of HgCdTe is very narrow, and the composition obtained for a given flux ratio is strongly temperature dependent. However, it is very important to have compositional and crystalline uniformity across HgCdTe epilayers to be used for FPAs. Thus, it is very important to keep the temperature as nearly uniform as possible across a growing HgCdTe layer. On the other hand, in MBE growth the outside edges, and especially the corners, of a growing epilayer become hotter than the central region. For that reason, it is highly desirable to use a substrate with a high thermal conductivity to minimize thermal gradients across a growing HgCdTe epilayer so as to minimize compositional gradients and pixel-to-pixel cutoff wavelength gradients in FPAs.

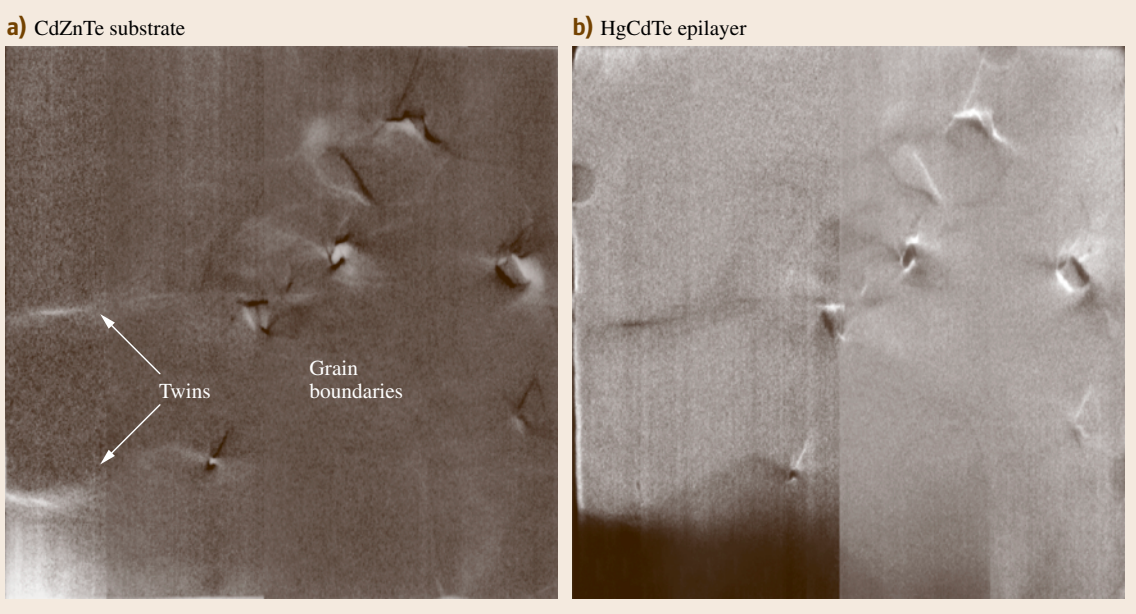


Fig. 32.1a,b Effects of (a) substrate twins and grain boundaries on defect formation in (b) HgCdTe epilayers

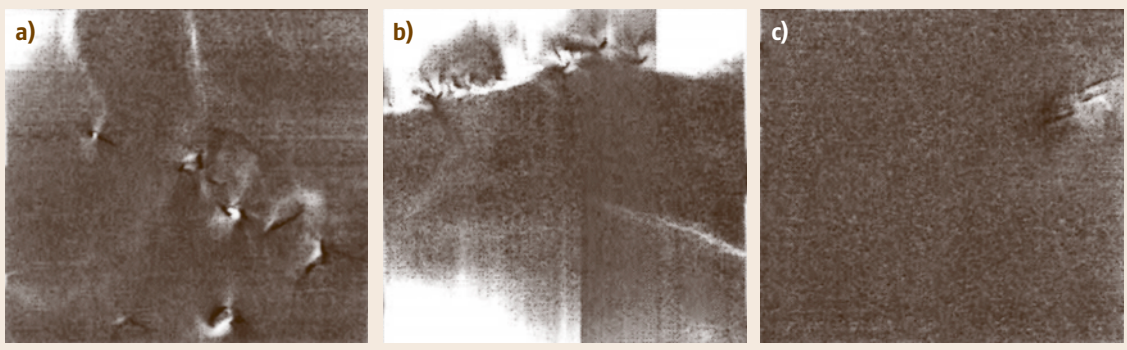


Fig. 32.2a–c Typical x-ray topographic mapping results from commercially purchased wafers with **(a)** and **(b)** large-area surface crystalline imperfections and **(c)** a moderate area of surface crystalline imperfection

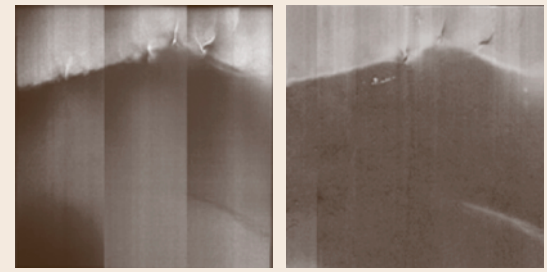


Fig. 32.3 X-ray topographic mapping results from HgCdTe epilayers grown on the substrates shown in Fig. 32.2a,b again showing that the defective features on the substrate are transferred to the HgCdTe epilayer

Effects of Substrate Crystalline Defects on HgCdTe Growth

Irrespective of the choice of substrate material, the crystalline quality of an MBE-grown HgCdTe epilayer depends on the crystalline quality of the substrate on which it is grown. The absence of a large quantity of substrate structural defects is a necessary, although not sufficient, condition for the MBE growth of a HgCdTe layer of high structural quality. The effects of substrate twins and grain boundaries on defect formation in HgCdTe epilayers is strikingly shown in Fig. 32.1; the HgCdTe surface defects form an exact map of the underlying CdZnTe defects. Of 12 commercially obtained substrates recently screened at the US Army Night Vision and Electronic Sensors Directorate, five showed large-area crystalline imperfections, as shown in Fig. 32.2a,b, three showed moderate areas of crystalline imperfections, as shown in Fig. 32.2c, and only four showed reasonably good surface crystalline perfection as revealed by x-ray topographic mapping, FWHM maps, and Fourier-transform IR (FTIR) transmittance maps. Results on others of these samples are given by *Carini* et al. [32.64]. When HgCdTe epilayers were grown on these substrates, the defective structures were transferred to the epilayers as shown in Fig. 32.3. Also, a high density of threading dislocations was found in the epilayers above the imperfections and very short carrier recombination lifetimes were measured [32.65]. These threading dislocations also can lead to microtwinning by allowing nucleation on the (111) terraces during growth as well as at the bottom of steps, thus giving three-dimensional rather than step-flow growth. This in turn leads to surface hillocks such as those shown in Fig. 32.4. More recently it was reported [32.44] that some substrates of a recently purchased batch had FWHM values of 15–25 arcsec, whereas others had values of 10–12 arcsec, with CdTe epilayers grown on the substrates having FWHM values limited by the substrate values.

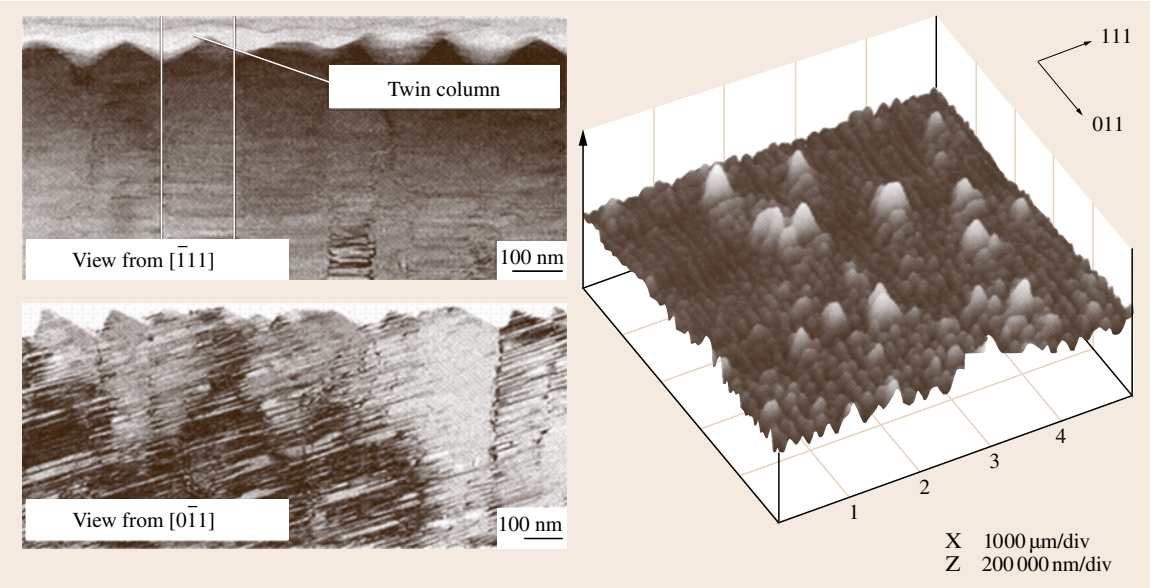


Fig. 32.4 Scanning electron microscopy (SEM) and atomic-force microscopy (AFM) images showing hillocks due to the formation of adjacent columns of twins

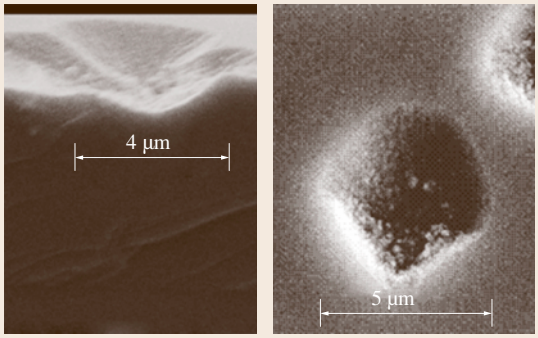


Fig. 32.5 SEM photographs of small voids

Another type of bulk defect found in CdZnTe substrates is Te precipitates. A large density of Te precipitates is unavoidable in as-grown CdZnTe boules because there is a large deviation from stoichiometry at the growth temperature. As the CdZnTe is cooled to room temperature, the excess Te forms precipitates from $< 1\,\mu m$ to $\approx 25\,\mu m$ in size. Much attention has been paid to elimination of these defects [32.66, 67], yet they remain a problem. These substrate precipitates create dislocation clusters [32.68] and induce Te precipitates [32.66, 69] in epilayers grown on the substrates and crater or void defects on the epilayer surfaces [32.70]. These voids can short p–n junctions and hence blind pixels in FPAs. Scanning electron microscopy (SEM) images of typical voids are shown in Fig. 32.5. The Te precipitates also give rise to high hole concentrations in HgCdTe layers grown by LPE on the substrates [32.69].

Effects of Substrate Impurities

For best performance of LWIR HgCdTe FPAs, it is necessary to minimize low-temperature leakage currents. At temperatures such as 40 K the major sources of dark current crossing the p–n junction are believed to be tunneling and generation in the depletion region. These mechanisms involve trap sites that most likely are due to impurities. Some impurities also act as unwanted dopants. For these reasons, it is necessary to minimize impurities in the HgCdTe layers. For growth on CdZnTe substrates, one origin of impurities is the substrate, from which they diffuse during layer growth [32.71] and device fabrication. A well-known example is Cu [32.72,73], which diffuses rapidly from substrate to epilayer, has been shown [32.74] to be preferentially attracted to HgCdTe over CdZnTe, and harms devices by shortening the carrier lifetime and converting weakly n-type material to p-type. Methods to extract mobile impurities from CdZnTe before MBE growth [32.75] and to getter and localize Cu at Te precipitates in CdZnTe [32.60] have been proposed, but Cu impurities remain a possible problem.

Effects of Nonuniform Composition

Zn segregation along the growth axis in the vertical Bridgman growth of CdZnTe leads to a gradual decrease in Zn composition from the bottom of a boule to the top, so that different wafers do not have the same lattice constant, allowing only part of a boule to be lattice-matched to a given composition of HgCdTe. More importantly, the Zn composition typically varies by about 1% from the center to the outside of a nominally 4% wafer, and that variation creates a variation in lattice constant equal to that caused by a 0.2% change in the composition of HgCdTe. Although small, that variation can cause a lateral variation in the density of threading dislocations in HgCdTe epilayers grown on CdZnTe substrates that is sufficient to cause significant lateral variations in the responsivity of LWIR and VLWIR FPAs. It also induces cross-hatch defects [32.76, 77] on the surface of MBE-grown HgCdTe epilayers, such as those shown in Fig. 32.6. Assuming CdZnTe and HgCdTe to relax with $\boldsymbol{a}/2\,\langle 011\rangle$ Burger's vectors which propagate on the $\{111\}$ family of glide planes $\boldsymbol{g}$, the resultant dislocation lines would be $\{\boldsymbol{u}\} = \{[01\bar{1}], [\bar{2}13], [0\bar{1}1], [2\bar{3}\bar{1}]\}$. Misfit dislocations from the small lattice mismatch

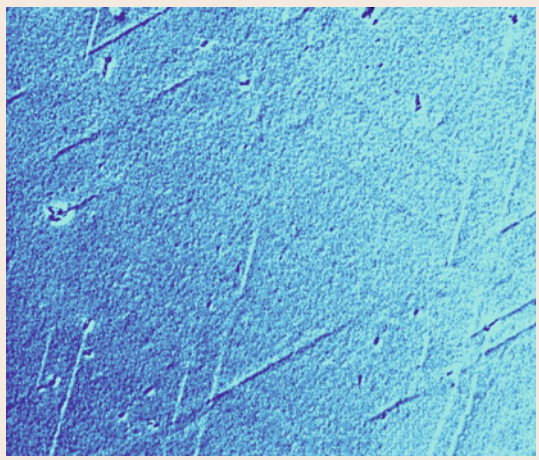


Fig. 32.6 Nomarski micrograph showing crosshatches on defect etched MBE-grown HgCdTe epilayers

and later from the thermal mismatch would allow threading dislocations to propagate on the {111} glide planes and close their dislocations along the ***u*** directions, resulting in the observed cross-hatch defects. Thus, these cross-hatch defects result from dislocations, but only when the dislocation density is low.

Effects of Substrate Roughness

Substrate surface roughness may be present on commercial substrates as received or may arise from imperfect surface preparation by the MBE grower. AFM surface roughness measurements on HgCdTe layers grown on substrates with smooth or rough surfaces show a direct correlation between substrate surface roughness and the surface roughness of the HgCdTe epilayer (Table 32.2). A high epilayer surface roughness on any lateral scale leads to imperfections in device fabrication and can lead to dead pixels in FPAs. High-frequency substrate roughness on a lateral scale of ≈ 100 Å or less inhibits the mobility of atoms on the epilayer growth surface and thus can lead to three-dimensional growth and poor crystalline quality. It has been observed that, for growth on moderately rough substrates (> 10Å of roughness as measured by SE), the epilayer surface roughness increases in the initial growth stage [32.78] consistent with three-dimensional growth.

It also has been observed that high densities of elongated *needle* defects oriented in a preferred direction occur on the surfaces of HgCdTe epilayers grown on rough CdZnTe substrates. As shown in Fig. 32.7, the density of these defects was found to increase exponentially with increasing substrate surface roughness as measured by SE [32.78]. However, the MBE deposition of an 80 Å CdTe buffer layer before HgCdTe growth reduces the final needle defect density by an order of magnitude or more, as shown in Fig. 32.7. Needle defects occur only for growth on CdZnTe, their density does not correlate to the CdZnTe bulk quality, and their connection to the surface roughness is not well understood. It also is not known how they affect device quality or how they correlate with surface preparation technique. It is possible that they could be eliminated by using plasma etching rather than wet chemical etching followed by thermal desorption of contaminants, the common practice. The length of the needle defects varies from a few micrometers to $> 10\,\mu$m in different samples, their width is only $0.2-0.3\,\mu$m, and their height is ≈ 100 Å. Figure 32.8 shows (a) a Nomarski picture (1000×) of needle de-

Table 32.2 Correlation of substrate surface roughness and epilayer surface roughness

	Smooth substrate (Å)	**Rough substrate** (Å)
Substrate surface roughness	8	50
HgCdTe epilayer surface roughness	8	58

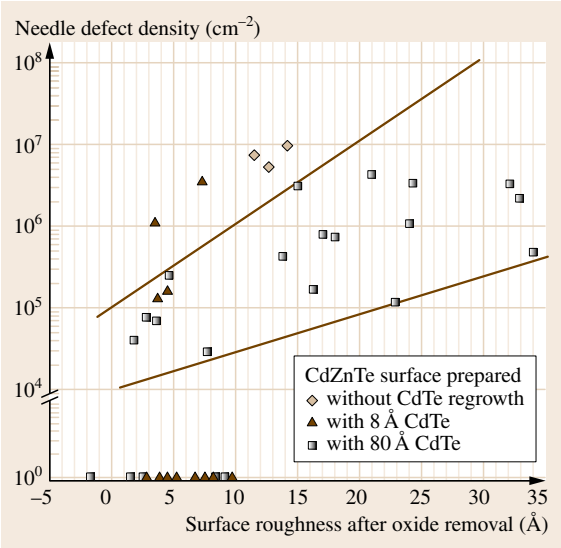


Fig. 32.7 Plot of the HgCdTe needle defect density versus surface roughness as measured by in situ SE with and without CdTe buffer layers. The *straight lines* are upper and lower limits of the defect density with an 80 Å CdTe buffer layer

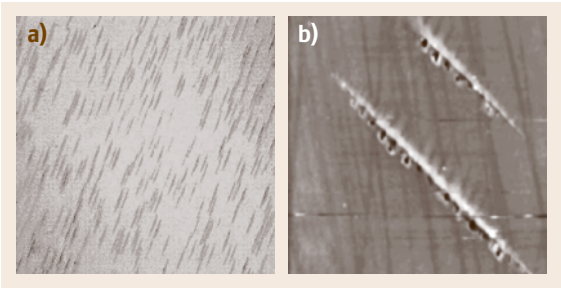


Fig. 32.8 **(a)** A Nomarski picture (1000×) of needle defects on a HgCdTe epilayer and **(b)** an AFM image of two needle defects ($5\,\mu$m × $5\,\mu$m)

fects on a HgCdTe epilayer and (b) an AFM image of two needle defects. The HgCdTe epilayers grown on these substrates also were characterized by temperature-dependent, photoconductive-decay lifetime data. Fits to the data indicate the presence of mid-gap recombination centers, which were not removed by 250 K/24 h annealing under a Hg-rich atmosphere. The strong correlation observed between the substrate roughness measured by SE and the density of these centers leads us to believe that they originate from bulk defects related to the substrate roughness rather than from Hg vacancies.

Effects of Substrate Surface Curvature

Substrate surface curvature can have two deleterious effects. First, any deviation in surface flatness $\approx 0.5\,\mu m$ or more can give problems in bonding to the flat Si ROIC surface for device fabrication. Second, the curvature will yield a surface orientation which is not constant across the wafer; that, in turn, may lead to epilayer MBE growth problems for a substrate which is only slightly miscut off a principal crystal axis, although it should have no effect on the usual growth on (211) surfaces.

Effects of Surface Nonstoichiometry and Contaminants

The surfaces of substrates must be prepared for MBE growth, i.e., polished and cleaned. The surfaces must be atomically clean, ordered, and have, at least approximately, the stoichiometry of the underlying wafer. The surfaces of commercially obtained wafers typically are covered with ≈ 1.5 ML of contaminants (mostly oxygen, carbon, and chlorine) [32.79]. The usual wet chemical preparation of the surface typically employs a 0.5% bromine in methanol solution that leaves behind an amorphous Te layer and residual carbon contamination, both of which must be removed for satisfactory nucleation of HgCdTe growth, although improved chemical polishing and cleaning procedures are under investigation [32.80, 81]. The substrate is then heated to 300–340 °C to remove the amorphous Te and residual contaminants such as carbon and oxygen to obtain a crystalline surface. This surface preparation procedure has been shown to leave $\approx 1/4$ of a monolayer of carbon as well as either excess Te or excess Zn [32.79, 82]. Incompletely removed amorphous Te leads to three-dimensional growth and microtwinning, and residual contaminants degrade the electrical properties of any HgCdTe epilayer grown on the substrate. Alternative wet chemical treatments lead to similar problems. An alternative electron cyclotron resonance Ar/H_2 plasma cleaning technique has been shown to leave the CdZnTe surface free of contaminants and with a roughness of only 4 Å as measured by interferometric microscopy [32.79]. However, HgCdTe epilayers grown by MBE on plasma-cleaned substrates were not found to be noticeably better than those grown on substrates cleaned by conventional wet chemical methods followed by annealing [32.79].

Characterization and Screening of CdZnTe Substrates

Many experimental methods have been used to characterize the crystalline quality, purity, and surface characteristics of CdZnTe substrate wafers. Poor crystalline quality – the presence of grain boundaries, low angle microcrystals, microtwinning, precipitates, and voids – can be examined in detail using cross-sectional SEM and TEM; however, those measurements are expensive, destructive, and can only be used to image microscopic sections of a substrate. X-ray topographic mapping measurements, which produce images such as those shown in Figs. 32.1 and 32.2, also require very expensive equipment and are not as detailed, but are nondestructive and capable of mapping an entire substrate wafer [32.64]. More detailed structural information, such as small variations in lattice spacing and the existence of regions of the substrate tilted only 0.011°, can be obtained by reciprocal-space mapping (RSM) and triple-axis x-ray diffractometry [32.83]. A simple gross quantitative measure of the substrate crystal structural perfection is given by the full-width at half-maximum (FWHM) of the x-ray double crystal rocking curve and by the existence or nonexistence of multiple peaks. The IR transmittance at wavelengths $< 6\,\mu m$ [32.64] and χ, the ratio of channel counts to random counts in Rutherford backscattering spectroscopy [32.66], also give qualitative measures of the crystal structural perfection, with the IR transmittance being higher and χ being smaller for better crystals.

Individual Te precipitates and other extended defects can be seen by relatively simple nondestructive optical measurements: IR transmission microscopy [32.66, 69, 70], light-scattering IR tomography [32.84], and scanning photoluminescence (PL) [32.84]. Also, the presence of Te precipitates or excess Te atoms is revealed by an A(Te) peak in the Raman scattering spectra at $\approx 127\,cm^{-1}$ [32.67] and IR transmittance gives at least a qualitative measure of the density of Te precipitates, falling off rapidly with increasing wavelength for wafers having a high density of Te precipitates [32.67] A general rule of thumb is that, for wavelengths well above the cutoff wavelength,

the transmittance should be $> 63\%$ in the absence of antireflection coatings. Reflection high-energy electron diffraction (RHEED), which is installed on all HgCdTe MBE growth chambers and is discussed in Sect. 32.5, gives yet another nondestructive qualitative description of the substrate crystalline perfection. Etch pit density (EPD) measurements reveal threading dislocations and Te precipitates at the substrate surface and are nondestructive except for the requirement of a surface etch, usually either an Everson [32.85] or a Nakagawa etch [32.86]; EPDs in the mid $10^4\,\mathrm{cm}^{-2}$ are desirable.

Quantitative measurements of substrate impurity dopant concentrations require destructive techniques. The most common are glow-discharge mass spectroscopy [32.73, 75, 87], sputter-initiated resonance ionization spectroscopy [32.70, 74], and secondary-ion mass spectroscopy (SIMS) [32.67, 88–90]. However, a qualitative measure can be obtained from the IR transmittance; for high impurity dopant levels the transmittance decreases by an amount proportional to $\ln(N_\mathrm{D} - N_\mathrm{A})$ as the wavelength increases in the LWIR and VLWIR [32.84].

The Zn concentration and its lateral variations usually are measured by FTIR or spectral photometry, measuring the cutoff in the FTIR transmission in the near IR, because this is a very simple nondestructive measurement. However, the shape of the FTIR transmission curve is affected by crystalline imperfections, in particular by Te precipitates, in a way that can make the determination of the Zn concentration uncertain by as much as ± 0.015 [32.81]. It can be determined somewhat more accurately by SE using a high-resolution ellipsometer such as the M2000, as discussed in Sect. 32.5. A much more accurate method of determining the Zn concentration and its lateral variation is low-temperature PL mapping. Measurement of the reabsorption dip at the free exciton energy is capable of determining the Zn concentration within a precision of ± 0.0005 [32.81, 82]. The depth variation of the Zn concentration also can be measured, but only by destructive techniques: Auger electron spectroscopy (AES) or atomic absorption spectroscopy (AAS) on beveled edges or vertical slices of a substrate [32.88].

The surface roughness can be nondestructively and easily measured directly by Nomarski microscopy, Zygo interferometry, and AFM with increasing resolution but over decreasing areas, and indirectly by SE, as discussed in Sect. 32.5. Surface contamination also can be measured nondestructively by x-ray photoelectron spectroscopy (XPS), directly by measuring the photoelectron peak intensities corresponding to the different possible contaminants or indirectly by measuring the decrease in the intensity of the CdTe peak [32.82]. In screening substrates, the substrate should first be thermally cleaned in vacuum in the XPS chamber and then transferred in vacuum to the growth chamber after the XPS measurement. AES also can be used to verify the absence of contaminants on substrate surfaces [32.79] but is not as easily used to measure amounts of contamination.

Use of Buffer Layers on Substrates

The use of buffer layers to block the propagation of defects originating in substrates or at the substrate–epilayer interface has proven useful and is being optimized. The various commercial manufacturers of FPAs use buffer layers in the growth of HgCdTe, but the details of their growth procedures remain proprietary. Thin interfacial CdTe layers have proven effective in reducing the surface roughness, evening out variations in strain, and improving the quality of HgCdTe layers grown on the substrates [32.78]. However, in order to be fully effective in smoothing out the surface and reducing the dislocation density, these layers must be grown to a thickness exceeding the critical layer thickness, so that the lattice match to HgCdTe is lost. This suggests the use of HgTe/CdTe SLs as buffer layers. They are more effective at blocking dislocations and can be grown more nearly lattice matched to the HgCdTe epilayer to be grown. Cross-sectional TEM micrographs such as those shown in Fig. 32.9 show that the substrate microscopic roughness is smoothed out after several periods of SL growth (a) and that threading dislocations terminate (b) or bend (c) at the HgTe/CdTe interfaces in the SL, despite the small (0.32%) lattice mismatch between HgTe and CdTe which would normally be considered insufficient to induce the motion of threading dislocations [32.91, 92]. After growth, the SLs are interdiffused by annealing to produce a short-wavelength IR (SWIR) HgCdTe layer which serves as a window for backside-illuminated devices and typically has a lattice mismatch of $\leq 0.09\%$ with the final HgCdTe epilayer to be grown. It has been shown [32.91, 92] that the use of HgTe/CdTe SL buffer layers leads to reproducibly high-quality MBE-grown HgCdTe layers with EPDs $\approx 10^5\,\mathrm{cm}^{-2}$ with values as low as $4\times10^4\,\mathrm{cm}^{-2}$ and minority-carrier lifetimes close to the theoretical limits for perfect material. It also has been found that, even aside from the improvement of material characteristics, the presence of an SWIR buffer layer improves device quality by reducing the back-surface recombination velocity.

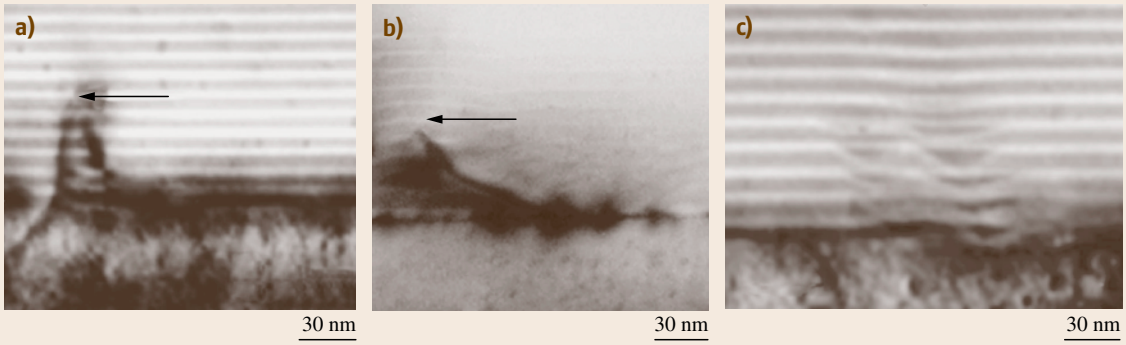


Fig. 32.9a–c Cross-sectional TEM micrographs showing the (a) bending and (b) termination of threading dislocations by HgTe/CdTe SL buffer layers, and (c) the smoothing of substrate roughness by SL buffer layers

32.3.3 Si-Based Substrates

There has been great interest in alternatives to CdZnTe as a substrate for the reasons outlined in Table 32.1 and detailed in Sect. 32.3.2 above. Of the many possible alternative substrate materials for the MBE growth of HgCdTe, Si has captured the greatest interest. HgCdTe was first grown on Si substrates in 1989 [32.93], and the growth of HgCdTe on Si having properties as good as those of HgCdTe grown on CdZnTe was reported in 1995 [32.94]. Although many factors favor Si over all other possible substrate materials, one factor in particular singles out Si uniquely – only Si offers the possibility of fabricating truly monolithic detectors and FPAs. Because the readout integrated circuit (ROIC) in all IR devices is based on Si, the use of any material other than Si for the substrate necessitates a hybrid-array technology wherein the detector and ROIC arrays are fabricated on separate manufacturing lines and subsequently indium-bump bonded to each other pixel by pixel. The functions of detection and charge readout would be monolithically integrated onto a single Si wafer in the ultimate IRFPA. This has been demonstrated by the epitaxial deposition of HgCdTe in selected areas of a Si ROIC [32.95].

Aside from the ultimate goal of monolithic integration, Si-based substrates offer many short-term benefits over all other possible substrate materials. First, Si wafers are commercially available in far larger sizes. This allows either much larger focal planes for greater resolution or greater numbers of smaller FPAs for reduced production costs. Second, in conjunction with that advantage, Si offers perfect thermal matching of its lattice to that of the ROIC, and thus no thermal mismatch limit on the size of FPAs. With other substrates, the thermal mismatch between substrate and Si ROIC creates a strain upon cycling down to cryogenic operating temperatures which could destroy the bump bonding for very large-area FPAs. Third, Si substrates allow the use of automated Si processing technology for increased efficiency in device fabrication. Fourth, Si is the most robust and durable possible substrate material, leading to less breakage and a higher device yield. Fifth, Si has the highest thermal conductivity, leading to the highest lateral uniformity of MBE-grown HgCdTe composition and thickness. Sixth, Si offers the lowest level of impurity migration into the HgCdTe. Finally, it has by far the lowest density of surface defects, although that advantage is overwhelmed by the problems associated with the large lattice and thermal mismatches between Si and HgCdTe. In summary, in every way that CdZnTe is a poor substrate, Si is clearly the best.

On the other hand, Si has the worst lattice and thermal mismatches with HgCdTe of any substrate material considered. The lattice parameter of HgCdTe (≈ 6.46 Å) is $\approx 19\%$ larger than that of Si (≈ 5.43 Å). The lattice mismatch is so extreme that HgCdTe does not even grow epitaxially on Si, so that a buffer layer is required. The buffer layers chosen have been $Cd_{0.96}Zn_{0.04}Te$ [32.96], CdSeTe [32.97], CdZnSeTe [32.98], and CdTe. Also, a first buffer layer of ZnTe, originally found to be necessary to preserve the substrate orientation [32.93], usually is grown before the deposition of the final buffer layer. Although $Cd_{0.96}Zn_{0.04}Te$ and CdZnSeTe buffer layers offer the advantage of an almost perfect lattice match to the HgCdTe to be grown, they have largely been abandoned because of the difficulty of obtaining epilayers of sufficient crystalline quality; the best values obtained for the x-ray diffraction (XRD) FWHM for lattice

match to HgCdTe have been > 150 arcsec. Even CdTe buffer layers have a relatively poor crystalline quality because of their large lattice mismatch with Si. Even with special growth procedures involving the predeposition of ZnTe films on As-terminated Si surfaces [32.99, 100], CdTe buffer layers exhibit etch pit densities of 2×10^5–$5\times10^5\,\mathrm{cm}^{-2}$, indicative of threading dislocation densities of the order of 2×10^5–$5\times10^5\,\mathrm{cm}^{-2}$. This in turn leads to dislocation densities in the mid $10^6\,\mathrm{cm}^{-2}$ in the HgCdTe layers grown on these substrates by MBE. Although otherwise excellent material quality has been achieved [32.101–104], little success was achieved in lowering the dislocation density in HgCdTe grown on Si substrates without undesirable high-temperature cyclic annealing.

For dislocation densities $< 10^5\,\mathrm{cm}^{-2}$ the dislocation density has little effect on the minority-carrier recombination time for $T > 77$ K, even for LWIR HgCdTe, because carrier recombination is dominated by intrinsic Auger mechanisms. However, for dislocation densities above the mid $10^5\,\mathrm{cm}^{-2}$ range and $T \leq 77$ K, carrier recombination is dominated by Shockley–Read–Hall (SRH) recombination due to the dislocations, and the minority-carrier recombination time for LWIR HgCdTe for is roughly inversely proportional to the dislocation density [32.105]. Also, dislocations can locally short out p–n junctions in devices, considerably increasing the dark current [32.106]. The MBE growth of HgCdTe on CdTe/Si substrates yields SWIR and mid-wavelength infrared (MWIR) photodiodes having performances similar to those grown on CdZnTe substrates at operating temperatures above 77 K [32.102, 103, 107, 108]. However, to date it has been incapable of yielding LWIR or VLWIR devices having similar performances (similar R_0A values) to the best devices grown on CdZnTe substrates [32.103, 109], except for devices with very small junction areas ($< 180\,\mu\mathrm{m}^2$) [32.110] or for high background operation at temperatures ≥ 77 K [32.104]. LWIR FPAs fabricated from HgCdTe grown on Si give near-ideal performance even at 40 K if the junction sizes are sufficiently small to have no more than ≈ 5 threading dislocations passing through each junction area. This limits the ideal pixel pitch for LWIR FPAs to pixel sizes < 25–$30\,\mu\mathrm{m}$ [32.110].

There exist in principle two general ways to ameliorate the effects of the large lattice and thermal mismatches between Si and HgCdTe:

1. Reduce the as-grown dislocation density
2. Passivate the dislocations so as to reduce their effect on device properties.

Buffer Layers to Reduce the As-Grown Dislocation Density

If the as-grown dislocation density could be reduced by an order of magnitude, growth on composite Si-based substrates would be the clear choice for the MBE growth of HgCdTe. The use of improved buffer layers is expected to yield quantitative, but not qualitative, improvements. The authors have briefly explored the use of strained-layer CdTe/ZnTe SLs to lower the dislocation density and improve the crystal quality of the final CdTe buffer layer, but with only moderate success. Also, the growth of CdSeTe buffers on top of the usual CdTe buffers has been suggested. More recently, preliminary results [32.111] have shown that nanometer-thick perovskite oxide buffer layers tend to smooth the surface of the final CdTe buffer layer and may offer a promising approach to reducing the dislocation density. In particular, under certain growth conditions, an amorphous silicate layer forms between the perovskite and the Si, making the perovskite a potentially compliant buffer and thus potentially greatly reducing the dislocation density for growth on Si [32.112]. Another approach [32.113] would be the epitaxial growth of insulating Be chalcogenides and then BeCdTe or BeMgTe with a graded lattice parameter, initially lattice matched to Si ($BeSe_{0.45}Te_{0.55}$) and ending with a layer lattice matched to CdTe or HgCdTe. These layers could be grown in the same MBE reactor as a CdTe buffer layer, and the final HgCdTe epilayer would be grown in as part of one continuous growth run. They should grow with relatively few defects because the Be chalcogenides have very high stacking fault energies – more than ten times that of ZnS, which suppress the formation of dislocations. Also, the Be chalcogenides have a much higher thermal stability than CdTe or ZnTe. Alternatively, one could just use a single BeTe layer, which shows [32.114] good two-dimensional growth on Si:As and has an intermediate lattice constant. However, Be would be a very difficult contaminant to remove from the MBE growth chamber.

Thermal Cycling to Reduce the Dislocation Density

It has been shown that the thermal cycling of HgCdTe grown by MBE or MOCVD on CdTe/GaAs and CdTe/Si substrates can greatly reduce the HgCdTe dislocation density. Dislocation densities as low as 2.3×10^5, almost as low as those obtained for growth on CdZnTe, have been obtained by thermal cycling between 300 and 490 °C [32.115]. The reduction in the

HgCdTe dislocation density was accompanied by an increase in the CdTe dislocation density by two to three orders of magnitude. This effect was explained as due to the large thermally induced stresses between the CdTe and the Si or GaAs. The problem with this approach is that high-temperature annealing can cause substantial diffusion and destroy the sharpness of the hetero- and homojunctions required by many modern detector designs.

Growth on Si-Based Compliant Substrates

Growth on compliant substrates was first suggested by *Lo* [32.116]. Compliant substrates either take up most of the strain associated with a lattice mismatch or getter the dislocations formed at the epilayer–substrate interface. A compliant substrate is made much thinner than the epilayer grown on it and is free to slide against a thicker *handle* substrate. The most common example of a compliant substrate is a Si-on-insulator (SOI) structure with a Si layer $\approx 10\,\text{nm}$ thick floating on an SiO_2 layer on a Si handle. Modification of the relaxation behavior in lattice-mismatched films due to growth on SOI compliant substrates is well documented. TEM measurements for SiGe grown on such films have shown dislocation densities reduced by more than five orders of magnitude [32.117]. AFM measurements have shown a factor of four reduction in surface roughness, and the XRD FWHM was shown to be reduced from 270 to 155 arcsec [32.118]. It has been shown that in this case the compliant thin Si layer does not deform elastically to absorb the strain due to the lattice mismatch, but rather facilitates a net downward image force on the dislocations so that they grow into the thin Si layer rather than the SiGe epilayer [32.117, 118]. In preliminary studies CdTe has been grown on an SOI substrate 20 nm thick. As opposed to the tensile strain found for growth on the usual Si substrates, the CdTe was found to be under compressive strain due to the difference between its thermal contraction and that of Si during the cooldown from its growth temperature to room temperature. However, the fact that CdTe has lower energies of formation for stacking faults and dislocations than does SiGe will make it more difficult to grow dislocation free CdTe on an SOI substrate. $Ge/SiO_2/Si$ would be an attractive alternative compliant substrate because Ge is more nearly lattice matched to CdTe than is Si and has even been used as a bulk substrate for the growth of HgCdTe. This substrate and a method for its growth are suggested in [32.113].

Selective-Area Growth on Patterned Substrates

There are two basic approaches to patterned heteroepitaxial growth. The first involves selective-area growth on a prepatterned substrate followed by lateral epitaxial overgrowth (LEO). The second involves the standard planar growth of a buffer layer followed by patterning of the buffer, annealing and then LEO. Both have been used with great success. The first approach has been used with MOCVD to grow GaAs on Si (4% lattice mismatch) [32.119] and GaN on Si (22% lattice mismatch) [32.119–121], as well as CdTe on Si (19% lattice mismatch) [32.122, 123], with dislocation densities much lower than those for planar growth in all cases. Patterned growth and LEO by MBE have been published only for GaAs on a patterned GaAs substrate [32.124] and for GaN on patterned Si [32.125] and has not been considered possible for CdTe. However, using nanopatterning with a CdTe seed layer, Bommena now has achieved the selective growth of CdTe on Si and LEO with MBE. Although much remains to be demonstrated, this work may open the door to finally overcoming the lattice mismatch between Si and CdTe. If nanopatterning does not sufficiently reduce the dislocation density at the CdTe surface, a further reduction could be achieved by patterning compliant SOI rather than Si.

The second approach has been used to remove almost all threading dislocations from patterned ZnSe epitaxial layers grown on GaAs by MOCVD by bending them over along glide planes to the pattern sidewalls [32.126, 127]. The activation energy for the glide process ($\approx 0.7\,\text{eV}$) requires a high-temperature anneal, 400–600 °C, making this approach less desirable for MBE growth of CdTe.

Passivation of Dislocations

A possible alternative to achieving an order of magnitude reduction in the dislocation density would be to achieve a partial reduction in the dislocation density and attempt to passivate the remaining dislocations, i. e., eliminate the trap states normally introduced by dislocations. The hydrogen passivation of both shallow and deep levels has been well studied in many major semiconductors, particularly Si and GaAs. Perhaps, its most prominent application is in α-Si : H solar cells. Notably, it has been found to significantly reduce leakage currents through dislocation cores in GaN [32.128]. Atomic hydrogen has been found to passivate shallow donors and acceptors in virtually all semiconductors. The exact mechanisms by which hydrogen passivates

deep levels are poorly understood, in part due to the lack of a clear understanding of the microscopic nature of many deep levels. The passivation of shallow and deep levels in CdTe and has received some attention; hydrogen was found to be an effective passivant when introduced in either molecular or atomic form, improving both electrical and optical properties [32.129–131].

The effects of hydrogen in HgCdTe still are not clear. Early research on the hydrogen passivation of HgCdTe involved electrochemical methods of introducing hydrogen into $Hg_{0.5}Cd_{0.5}Te$, followed by deep-level transient spectroscopy to determine its influence on deep levels [32.132, 133]. The dominant deep levels at approximately $E_g/2$ and $3E_g/4$ were significantly reduced in concentration after hydrogenation. Unfortunately, annealing at 70 °C restored their activity. Two more recent papers [32.134, 135] examine the effects of the hydrogenation of mercury-vacancy-doped MWIR and LWIR HgCdTe. Reference [32.134] concludes that electron cyclotron resonance (ECR) plasma hydrogenation is effective in passivating surface trap states, whereas [32.135] concludes that its only role is to passivate mercury vacancies.

Recently [32.136], attempts have been made to passivate the dislocations in HgCdTe grown on Si with hydrogen introduced via an ECR plasma. The dangling bonds formed along the dislocations give rise to states in the bandgap that act as Shockley–Read–Hall recombination centers. However, the attachment of hydrogen to the dangling bonds would be expected to eliminate the gap states, forming bonding states in the valence band and antibonding states in the conduction band. Hall mobility and minority-carrier lifetime data indicated that the incorporated hydrogen did passivate both scattering and recombination centers. It also was found that dislocations act as atomic diffusion channels both for outdiffusion, observed during annealing, and for indiffusion during ECR hydrogenation. Therefore it is reasonable to conclude that dislocations would be among the first type of defects to be passivated when samples are exposed to ECR plasmas, and that this passivation of dislocations was primarily responsible for the observed increases in mobilities and minority-carrier lifetimes. Preliminary stability studies indicated that the beneficial effects of hydrogenation remain after 3 months shelf storage and after heating to 80 °C.

Te and As Monolayers on Si

The growth of an As monolayer on Si before the growth of a CdTe or ZnTe buffer layer has been shown to be crucial [32.137]. First-principles density-functional calculations of Te adsorption on Si(100) [32.138] and pseudopotential density-functional calculations of the electronic and atomic structure of monolayers of Te and As on Si(211) [32.139] have been performed. Te-covered surfaces were found not to have any definitive reconstruction, explaining conflicting experimental results for the reconstruction [32.138]. Both As and Te surfaces were found to be metallic [32.139]. More importantly, the effects of Te passivation and As passivation of the Si surface before CdTe growth have been compared experimentally [32.137]. The CdTe film grown on a substrate treated with a Te flux was found to exhibit a rough film–substrate interface and to have very poor crystalline quality with a ≈ (111)A orientation. In contrast, the CdTe film grown under identical conditions except that the Si substrate was treated with an As flux was found to have an atomically abrupt film–substrate interface and a single-domain structure in the technologically preferred (111)B orientation. On the As-passivated surface the CdTe growth is initiated with Cd atoms having one bond to an As atom and three bonds with Te atoms from the incoming Te flux, producing an initial smooth Te stabilized growth surface. Recently, the structure of the As and Te atoms on As-passivated Si(211) surfaces after the early stages of ZnTe growth has been investigated [32.140] by x-ray photoelectron spectroscopy (XPS) and ion-scattering spectroscopy (ISS). It was found that the Si(111) terraces are completely covered by a monolayer of As, but that the step edges were left free to be covered by the first layer of Te atoms, allowing optimum nucleation of either a ZnTe or a CdTe buffer.

32.3.4 Other Substrates

As seen in Table 32.1, the characteristics of GaAs are intermediate between those of CdZnTe and Si. GaAs was the first substrate commonly used as a substitute for CdTe, is still the most commonly used substrate for the LPE growth of HgCdTe, and is still used for the MOMBE and MBE growth of HgCdTe by some groups. Although the lattice mismatch between GaAs and HgCdTe is 13.6%, high-quality HgCdTe can be grown on CdTe/GaAs substrates. However, growth on GaAs does limit the growth area to ≈ 1/30 of that allowed for growth on Si, does not allow monolithic integration, and allows the outdiffusion of Ga and As into the HgCdTe, which can seriously affect the electrical properties of the HgCdTe. To bring the diffused Ga and As in the HgCdTe down to background levels, one must grow either CdTe/ZnTe SL buffer layers [32.51]

or thick (3–8 μm) CdTe buffer layers [32.51, 141] The dislocation density of HgCdTe grown on CdTe/GaAs is similar to that of HgCdTe grown on CdTe/Si : As.

Ge also is still used as a substrate for the MBE growth of HgCdTe with CdTe buffer layers. As in the case of Si, large-area Ge substrates are available. Two-color MWIR FPAs fabricated from HgCdTe grown on Ge substrates displayed responsivities, noise characteristics, and operabilities matching those of detectors fabricated from HgCdTe grown on CdZnTe substrates, although the HgCdTe grown on Ge substrates was of lower quality [32.142, 143]. Because Ge is a more homogeneous substrate than CdZnTe, the XRD FWHM of the HgCdTe grown on CdTe/Ge was more uniform; however, it was ≈ 50% larger than that of the HgCdTe grown on the better CdZnTe substrates. Also, the dislocation density was two orders of magnitude higher in the HgCdTe grown on CdTe/Ge, indicating that CdTe/Ge substrates would not be suitable for LWIR or VLWIR material. In summary, Ge would appear to have most of the advantages and disadvantages of Si as a substrate material, with easier to prepare growth surfaces, but without quite as large a possible area, quite as high a crystal quality, or the great potential advantage of being used for monolithic detectors.

In the past, sapphire also has been extensively used as a substrate for the MBE, VPE, and LPE growth of HgCdTe and the fabrication of MWIR photodiodes [32.144], although it is no longer used for MBE growth. It has high crystalline quality, high electrical resistivity and thermal conductivity, and low cost, and is robust. However, like GaAs, sapphire substrates are available only with areas ≈ 1/30 of that allowed for growth on Si. Also, CdTe buffers grown on sapphire exhibit a high density of dislocations, although that density is rather uniform, allowing the fabrication of highly uniform devices. Also, sapphire is not transparent in the LWIR and VLWIR, and thus cannot be used as a substrate for back-illuminated devices in those spectral ranges.

The final possible substrate material we consider is InSb. The physical properties of InSb suggest it as an almost ideal substrate for the epitaxial growth of HgCdTe. InSb and IR HgCdTe have almost identical lattice constants and coefficients of thermal expansion, and HgCdTe has been grown on InSb with XRD FWHM values as low as 18–22 arcsec [32.145]. Also, InSb wafers are available with diameters up to 4 inches. However, a number of difficult issues must be solved before InSb can be developed as a substrate material for the growth of HgCdTe. First, InSb has a low melting point (527 °C) and a tenacious native oxide, precluding thermal desorption as a wafer cleaning technique. The oxide can be removed in situ by Ar sputtering, and the induced damage partially removed by thermal annealing, but some surface and subsurface damage remains. Second, In_2Te_3 forms readily at the interface, creating structural defects that degrade any following CdTe or HgCdTe epitaxial growth, although this effect can be partially suppressed by using a Cd-enhanced flux during the initial CdTe growth. Finally, both In and Sb diffuse rapidly into CdTe and HgCdTe, where they strongly affect the electrical properties. Thus, appropriate diffusion blocking buffer layers would be required with the use of InSb substrates unless the substrates are removed before any annealing.

32.4 Design of the Growth Hardware

The reader is referred to [32.13] for a comprehensive review of the growth hardware for MBE growth. Here we discuss only two more recent growth hardware issues of importance for the MBE growth of HgCdTe: the mounting of the substrate on which the HgCdTe is to be grown, and the valving of the effusion cells which supply the growth materials.

32.4.1 Mounting of the Substrate

The standard mounting geometries for MBE substrates include contact and contactless configurations. In the contact configuration the substrate is bonded to a diffusion plate using a mounting or bonding medium such as In or Ga. Contactless configurations were introduced to reduce the risk of contamination from the mounting media such as In or Ga; thus they are referred to as free or In-free mounting configurations. Another advantage of contactless configurations is the smaller thermal mass, allowing more rapid substrate temperature adjustments. In contactless configurations, the substrate is sandwiched between a diffuser plate and either a Mo retaining spring plate for a square substrate such as CdZnTe or a Mo ring for round substrates such as Si. The diffusion plate can be just a Mo plate; however, to achieve a constant growth temperature with only

a Mo plate one must severely ramp the nominal thermocouple temperature during growth [32.146]. Moreover, for a fixed thermocouple reading, the actual growth temperature depends strongly on the exact position of the thermocouple and on the mounting geometry, and its run-to-run variation was found to be $> 10\,^{\circ}\mathrm{C}$, as measured by optical pyrometry and by SE. As an improvement, the diffuser plate can be replaced by a Mo plate and a graphite disk in contact with the substrate. It also was found to be beneficial to put a sapphire disk between the Mo plate and the graphite disk because the flat sapphire disk offers better mechanical support to the soft graphite plate. The use of a diffuser plate with a graphite plate in contact with the sample increases thermal conduction, decreasing the importance of irradiation, thereby allowing much better temperature run-to-run reproducibility. Another mounting problem is that the substrate must be rotated at a constant rapid angular velocity to ensure a uniform beam flux across a growth wafer, and with a wobble as small as possible to avoid excessive noise in the SE data.

The temperature ramping recipe that has to be employed to keep the growth temperature constant depends on the substrate area compared with the area of the diffusion plate. For example, for the growth of large-area samples, the substrate can be either (i) directly exposed to the heater and heated by irradiation, or (ii) juxtaposed to a diffuser plate and heated mainly by conduction. Case (iii) is that of smaller-area substrates such as CdZnTe, which are mounted using a molybdenum spring plate that holds the substrate in contact with a diffuser plate facing the heater. The heating power is determined by a closed-loop feedback involving a thermocouple, which floats in the gap between the heater and the substrate. Its readings are related to both the heater and substrate temperatures, and in some cases even to the temperature of the cell walls. Thus, the grower of HgCdTe faces a serious temperature stabilization problem during growth, since it has a very small growth window in comparison with other MBE-grown materials. It has been experimentally observed that in case (i) the heater temperature must be ramped down [32.146], whereas in cases (ii) and (iii) (which have some similarity to mounting on a full molybdenum block) a temperature increase is necessary to keep the growth temperature constant [32.147]. Case (i) has been the most widely investigated, especially for III–V MBE-grown materials [32.148, 149]. Here the temperature transient is believed to stem from the fact that the substrate is largely transparent to infrared radiation, whereas the growing material is not. Thus the thermal absorptivity increases during growth, which results in an increase of the temperature. If the sample heating mechanism is conduction rather than irradiation, as in cases (ii) and (iii), material deposition results in a temperature decrease. This is because the emissivity of the front face increases without any substantial change in the heat conducted through the sample, as the substrate is much thicker than the epilayer being grown. Also radiative energy from the effusion cells can increase the substrate temperature, especially at nucleation, when temperature stability is essential to achieving good crystalline quality [32.150].

For these reasons, several optical methods have been used to measure the HgCdTe temperature in situ, such as (a) SE [32.151], (b) in situ FTIR spectroscopy [32.152], (c) dynamic reflectance spectroscopy (DRS) [32.153], and (d) absorption-edge spectroscopy (ABES) [32.154], on both II–VI and III–V materials. It has been shown that good reproducibility can be obtained from SE even though it gives no direct information on the growth temperature (Sect. 32.5.1). In the case of HgCdTe, DRS measurements must be carried out from the back face of the sample and generally require the deposition of a reflecting intermediate layer. For FTIR and ABES the HgCdTe must be optically thick to obtain reliable readings, so they are not applicable during the early stages of growth. However, information provided by RHEED spectra also can be used to indirectly analyze the growth surface temperature stability. In fact, RHEED and SE so far are the only measurements capable of characterizing growth nucleation in situ.

The employment of spring plates coated with materials with infrared emissivities similar to that of the HgCdTe to be grown allows the temperature ramping to be dramatically reduced or even eliminated for the growth of small-area samples. By employing a graphite plate in addition to a diffusion plate and coating the spring plates with an appropriate material, a dramatic improvement in growth yield was found. So long as substrates with similar characteristics were used, thermocouple run-to-run errors as low as $\pm(2-3)\,\mathrm{K}$ were found for samples mounted with graphite plates and appropriately coated spring plates.

32.4.2 Valving of the Effusion Cells

Valved cells were introduced in the MBE growth of III–V materials as early as 1993 [32.155]. For the growth of HgCdTe a valved Hg source was introduced at the Microphysics Laboratory (MPL) of the Univer-

sity of Illinois at Chicago in 1998 and a valved Te source was introduced at the Rockwell Science Center in 1999 [32.156]. The use of a valved As cracker cell also recently has become common. The different valved cells share the same design concept, but with the Hg cell having a tube connected to the Hg reservoir to ease the Hg supply and the As cell having a high-temperature cracker to increase the ratio of As and As_2 to As_4 in the exeunt flux. The Hg and As fluxes can be precisely controlled and optimized by adjusting the computer-controlled needle valve for nearly continuous or abrupt flux changes with rapid flux stabilization, quick shutoff, and improved growth reproducibility. Also, the exit orifices of the valves can be customized to obtain excellent flux uniformity in any MBE systems. All these superior flux control capabilities are extremely desirable to maintain the optimized MBE growth of HgCdTe heterojunction material with complicated design architectures, which is impossible to achieve using conventional thermally controlled sources.

An As-valved cracker cell provides the convenience of a gas source with the safety of a solid source, which is beneficial to optimize flux profiles in time, reduce As consumption, and eliminate unintentional As incorporation into heterojunction structures. The source's cracking zone temperature can be adjusted to generate either As_4- or As_2-dominated flux beams. Such a valved cracker design also offers great flexibility to adjust the charge capacity and obtain optimized cracking efficiency.

32.5 In situ Characterization Tools for Monitoring and Controlling the Growth

Because the sticking coefficient of Hg is very much lower than those of Cd and Te, the growth window for MBE growth of HgCdTe is very small. The window in growth temperature for the growth of optimal epilayers has been reported to be as low as $\pm 0.5\,^{\circ}C$, and that for the Hg flux $\pm 2.5\%$, which corresponds to $\pm 0.5\,^{\circ}C$ in the temperature of the Hg cell [32.71]. This precise a degree of control is not possible, but it is very desirable to be able to control the growth temperature within $\pm 3\,^{\circ}C$ and the Hg flux within $\pm 5\%$. Furthermore, to obtain a desired cutoff wavelength or cutoff energy $\pm 6\%$ in a device, one must control the Cd composition x within ± 0.002. Thus, excellent in situ monitoring and controlling of the growth are necessary, requiring multiple methods for monitoring and controlling the growth.

32.5.1 Spectroscopic Ellipsometry (SE): Basic Theory and Experimental Setup

Due to its simplicity, intrinsic accuracy and noninvasiveness, spectroscopic ellipsometry (SE) is widely used as an in situ tool to monitor, control, and characterize the MBE growth of $Hg_{1-x}Cd_xTe$ alloys. The growth temperature, alloy composition, and presence of defects, overlayers or roughness at the surface in principle all are measurable by in situ SE, if the appropriate data analysis technique is used. Here, greatest attention is given to:

1. The use of SE to determine the temperature and surface roughness of substrates and the composition and surface roughness of the HgCdTe during growth
2. The principal methods of analysis of SE data, pointing out the strengths and weaknesses of each method.

Various specialized applications of SE in the study of the MBE growth of $Hg_{1-x}Cd_xTe$ also are given.

Since shortly after its introduction [32.157] as an in situ characterization tool in 1991, ellipsometry has been applied primarily to find the $Hg_{1-x}Cd_xTe$ alloy composition x. Single-wavelength ellipsometers were first used for this purpose [32.158, 159]. In 1964 44-wavelength SE was used to determine x to within ± 0.01 during MOCVD growth of $Hg_{1-x}Cd_xTe$ and feedback control was established to force x to converge faster to its targeted value [32.160, 161]. Growth rate control and monitoring of the CdZnTe preparation procedure, its surface roughness, and its temperature were reported in 1996 [32.162]. By 1998 SE was mature enough to be introduced in a production environment, as the run-to-run precision in x had improved to ± 0.001 or ± 0.002 [32.163]. Now SE has entered the realm of standard use for compositional control, even in industry. In actuality, the capabilities of SE go well beyond the determination of x, although little work has been done to explore other applications.

Introduction to Ellipsometry

Ellipsometry is a polarimetric technique that uses the change of polarization of light upon reflection for

the characterization of surfaces, interfaces, and thin films [32.164]. If the analyzed surface is homogeneous, isotropic, and not contaminated by the presence of overlayers, and neglecting the difference between the electronic structure at the surface, a single ellipsometric measurement with light of wavelength λ yields the complex optical dielectric function $\varepsilon_s(\lambda)$ of the material at that wavelength

$$\varepsilon_s(\lambda) = \varepsilon_a(\lambda)\left[\sin^2\phi + \sin^2\phi\tan^2\phi\left(\frac{1-\rho}{1+\rho}\right)^2\right], \tag{32.1}$$

where $\varepsilon_a(\lambda)$ is the dielectric function of the ambient, ϕ is the angle of incidence, and ρ is the complex reflectance ratio $r_s/r_p = \tan\Psi\, e^{i\Delta}$, where r_s and r_p are the Fresnel coefficients, i. e., the ratios between the incident and reflected electric fields for s- and p-polarized light, respectively. If $n > 1$ layers are optically accessible, $\varepsilon_s(\lambda)$ is replaced by a pseudodielectric function $\langle\varepsilon_s(\lambda)\rangle$ that depends on the dielectric function of each layer and the thickness of the first $n-1$ layers.

Multiwavelength ellipsometry (SE) provides the frequency-dependent dielectric function $\varepsilon_s(\omega)$, which is intimately connected to the semiconductor band structure and in particular to its critical-point (CP) energies. At these energies E_i there are discontinuities or infinite first derivatives in the joint density of states and thus, in the absence of line broadening, in the probability of photon absorption and hence in $\varepsilon_2(\omega)$, the imaginary part of $\varepsilon(\omega)$. Great effort has been devoted to the precise determination of the CP energies and line widths because they reflect any changes in temperature, composition, and other physical parameters. Given an ellipsometer with sufficient resolution and low noise, accurate determinations of the CP energies can be achieved by higher-order differentiation of the experimentally measured dielectric function [32.165], which greatly enhances the CP structure.

The actual measured quantities are the Fresnel coefficients of the sample, which usually are organized into Jones matrices of the form

$$\mathbf{J} = \begin{pmatrix} r_{pp} & r_{ps} \\ r_{sp} & r_{ss} \end{pmatrix} = r_{ss}\begin{pmatrix} \rho_{pp} & \rho_{ps} \\ \rho_{sp} & 1 \end{pmatrix}. \tag{32.2}$$

The Jones matrix of a stack of layers is just the product of the Jones matrices of the layers in the stack. For isotropic materials, such as HgCdTe and its substrates, the Jones matrices reduce to diagonal form. Since no existing in situ ellipsometer can measure all four Fresnel coefficients in one reading at a single azimuthal angle, surface anisotropy due to oriented surface defects or the morphology of high-index surfaces can only be treated in an approximate fashion as a source of uncertainty.

The polarizing properties of any sample are given by its Mueller matrix, which relates the Stokes vector $\boldsymbol{S}$ of the exiting reflected, transmitted or scattered light to that of the incident light $\boldsymbol{S}_0$. For a homogeneous, isotropic semiconductor, the Mueller matrix can be represented as the product of an ideal linear polarizer with transmission axis ψ and an ideal retarder of retardance Δ

$$\begin{aligned}\mathbf{M} &= \begin{pmatrix} 1 & -\cos 2\psi & 0 & 0 \\ -\cos 2\psi & 1 & 0 & 0 \\ 0 & 0 & \sin 2\psi\cos\Delta & \sin 2\psi\sin\Delta \\ 0 & 0 & -\sin 2\psi\sin\Delta & \sin 2\psi\cos\Delta \end{pmatrix} \\ &= \begin{pmatrix} 1 & -N & 0 & 0 \\ -N & 1 & 0 & 0 \\ 0 & 0 & C & S \\ 0 & 0 & -S & C \end{pmatrix}. \end{aligned} \tag{32.3}$$

The final form in (32.3) holds even with depolarization; in general, $C^2 + S^2 + N^2 = p^2$, where $p = 1$ for nondepolarizing samples and $p = 0$ for completely depolarizing samples. C, S, and N completely characterize all isotropic, homogenous materials, even depolarizing ones. If the further assumption is made that a sample is nondepolarizing, then the knowledge of any two of C, S, and N suffices to calculate the pseudodielectric function.

Ellipsometer Designs

The primary ellipsometer designs in use for in situ SE are the rotating compensator ellipsometer (RCE) and the rotating analyzer ellipsometer (RAE). These designs differ in how fully they characterize the Mueller matrix. In the RAE [32.166], light generated by the source is first linearly polarized by a polarizer, then reflected by the sample, goes through an analyzer, and finally enters the photodetector. The presence of the first polarizer makes the measurement immune from polarization by the source optics. The analyzer is continuously rotated to generate a modulation of the detected signal, which is Fourier-analyzed to extract C and N. This design cannot measure S because that would require circularly polarized light. A simple calculation shows that the irradiance at the detector is given by

$$I = I_{dc} + I_s\sin 2\omega t + I_c\cos 2\omega t\,,$$

where

$$\frac{I_s}{I_{dc}} = \frac{C \sin 2\theta_p}{1 - N \cos 2\theta_p}, \quad \frac{I_c}{I_{dc}} = \frac{\cos 2\theta_p - N}{1 - N \cos 2\theta_p}. \tag{32.4}$$

The azimuthal angle of the polarizer θ_p can be determined through a calibration procedure [32.167, 168]. Because S does not appear in these equations, even in the absence of depolarization only S^2 is determined, and the sign of Δ is undetermined. Also, the sensitivity of the RAE ellipsometer is zero for $\varepsilon_2(\omega) = 0$, hence, at or below the optical energy gap of semiconductors. In particular, RAE ellipsometers designed for the characterization of HgCdTe in the region around the E_1 CP (2–3.5 eV) do not perform well for CdTe or Si in the same energy range.

The RCE was introduced to circumvent these limitations. A compensator is inserted between the polarizer and the sample. In the most common design, the compensator is rotated continuously, whereas the analyzer and the polarizer remain fixed. The output irradiance at the detector has extra Fourier components, of angular frequency 4ω,

$$I(t) = I_{dc} + I_c^{(2)} \cos(2\omega t) + I_s^{(2)} \sin(2\omega t) + I_c^{(4)} \cos(4\omega t) + I_s^{(4)} \sin(4\omega t),$$

which allows C, S, and N to be calculated without ambiguity. Thus, the RCE determines the sign of Δ, gives accurate readings even for $\varepsilon_2(\omega) = 0$, and can measure depolarization. Several other ex situ designs which have been proposed can measure the anisotropic dielectric function and the depolarization at a single azimuthal angle, taking only two separate measurements, for example that by *Jellison* and *Modine* [32.169], but no applications to in situ measurements of HgCdTe have been proposed.

Retrofitting the MBE Chamber

The setup for in situ experiments is complicated by many constraints, and the data are subject to more sources of uncertainty than in an ex situ setup. Constraints are for example posed by the availability of ports for the windows, the window type, the intrinsic characteristics of the substrate manipulator, etc. The effect of windows on accuracy is very large for most ellipsometers [32.170]. Different ellipsometers measure different linear combinations of the Mueller matrix elements, and therefore differ in how fully they characterize and hence can correct for window imperfections. The RAE cannot measure even the out-of-plane component of birefringence during an experiment, while the RCE can. It is not possible with any instrument to separate the in-plane window retardation from that of the sample, even when the entire Mueller matrix is measured. To determine the in-plane window retardation one must resort to measurements with well-known samples, such as Si oxide of known thickness. Low-strain bakeable windows, guaranteed to have a birefringence of less than 1°, are commercially available, but even run-to-run variations of the order of 0.1°, which may occur when the windows are baked out, can be important. Also, the measured $\varepsilon_2(\omega)$ is extremely sensitive to changes in the angle and plane of incidence, so that much care must be taken to minimize wobble using wobble-free manipulators and a proper substrate mounting procedure. Unfortunately, substrate rotation cannot be avoided. Thus, it is common to equate the data acquisition time to the total rotation time of the substrate.

32.5.2 SE Data Analysis

To interpret ellipsometric results, the complex reflection coefficients must be computed from a model of the sample surface. In the general case, one constructs a model for the dielectric function which depends on the parameters to be determined. This is fitted to the experimental data using the standard Levenberg–Marquardt algorithm. The function itself and possibly its first, second or third derivatives, or a combination of them, are fitted. There are two principal approaches to parameterizing the optical constants of a bulk semiconductor, based either on the construction of calibrated libraries of optical dielectric functions or on the determination of their CP parameters. Library models are commonly used to find the composition and surface roughness of HgCdTe during growth, and the temperature and surface roughness of CdZnTe or CdTe before growth. Their main advantage is their ease of use. Unfortunately, when growth conditions are changed, such models can suffer from large run-to-run uncertainties, due to changes in surface morphology, small changes in temperature, and changes in window retardation and birefringence. The use of CP models for the dielectric function has been introduced to reduce these uncertainties.

Multilayer Models and the Virtual Interface Approximation

The most common multilayer model for HgCdTe, CdZnTe, and CdTe characterization is a three-layer model, comprised of epilayer or substrate, a surface roughness layer, and the ambient, normally vacuum. If the optical constants of the growing layer are un-

known, the virtual interface approximation [32.171, 172] and dynamic algorithms can be used to extract them, the growth rates, and in some cases even the surface roughness of the growing layer at the beginning of growth. Exact expressions for the dielectric function ε_0 of a virtual surface located at the physical surface can be found as a function of the measured pseudodielectric function $\langle\varepsilon\rangle$. These expressions can be fitted to the experimental data to obtain estimates for both the near-surface composition and the growth rate, as shown [32.150] for MOCVD-grown HgCdTe. The virtual interface approximation has attained great importance for the characterization of HgCdTe, as it provides a way to monitor growth nucleation.

Surface Roughness

Surface roughness results in reflectance loss, polarization of the specularly reflected light, and the introduction of stray light reflected at near-specular angles. Most of the specularly reflected light is reflected coherently and stays polarized. Part is reflected incoherently and is depolarized, but the depolarization caused by a small degree of microscopic roughness is minimal and needs to be taken into account only for moderately to very rough surfaces. In that case, near-specular scattering may also play a role [32.173]. For roughness on a scale smaller than the wavelength of the reflected light, the changes in the coherent, polarized part of the reflected beam can be described by the Bruggeman effective medium theory (EMA) [32.174,175]. That theory is derived from the Clausius–Mossotti expression [32.176] for the dielectric response of a heterogeneous medium of point-like microstructures embedded in a host for the case of an aggregate of two types of microstructures in which the volume fractions of the two types are comparable, so that the host must be chosen as some self-consistently chosen effective medium. In the EMA the dielectric function of the effective medium is chosen to be

$$\varepsilon = \frac{\varepsilon_a\varepsilon_b + \bar{\varepsilon}(f_a\varepsilon_a + f_b\varepsilon_b)}{\bar{\varepsilon} + (f_a\varepsilon_b + f_b\varepsilon_a)} , \qquad (32.5)$$

where ε_a and ε_b are the dielectric functions of the two types of microstructures, $\bar{\varepsilon} = (q^{-1} - 1)\varepsilon_b$, and q is a screening parameter between zero and one. The thickness t of the roughness layer is related to the mean-square roughness by $t = 2\sqrt{2}\sigma_z$ [32.177]. The void fraction in the roughness layer is commonly taken to be 50%, so that (32.5) reduces to the form $\varepsilon = [2\varepsilon_b + \bar{\varepsilon}(1 + \varepsilon_b)]/[2\bar{\varepsilon} + (1 + \varepsilon_b)]$.

Analysis Based on Dielectric Function Libraries

Libraries of dielectric functions are acquired at a set of values of one or more parameters, e.g., the temperature and/or the alloy composition, and interpolated between those values. In the simplest case, libraries are used in conjunction with multilayer models for the dielectric function. A typical model valid only for optically thick layers consists of a composition-dependent epilayer or temperature-dependent substrate, surface roughness, and ambient. The surface roughness is described in the EMA. Hence, a fit to the experimental data yields the thickness t of the roughness layer and either temperature or composition. In most in situ environments, it is also necessary to leave the angle of incidence ϕ free in the fit, thus having three independent fitting parameters.

To obtain a meaningful library, one's ellipsometer must be accurately calibrated, the windows must be free of birefringence, the angle of incidence must be known, and the sample composition and temperature must be known. Also the surface one is measuring must be exactly known, and no undesired overlayer or unknown roughness should be present. These conditions are very hard to attain in situ, but they can be attained at least approximately. A full calibration of the ellipsometer before data acquisition can be performed ex situ, with the angle of incidence as well as the sample type perfectly known. Then, standard calibration procedures can be performed. For most commercially available RAEs or RCEs, these calibration procedures are executed automatically by software algorithms. The window birefringence must be minimized using *strain-free* windows. Residual in-plane birefringence can be estimated in situ using a thermally grown SiO_2/Si calibration sample, the optical constants of which are well known. Since the window birefringence is approximately constant on any given MBE system, if low-strain windows are used, a precision of about $\pm 0.1°$ in the measured values of Δ is expected. The angle of incidence can be measured in situ in most cases, exploiting databases of room-temperature optical constants available in the literature.

Unfortunately, exact knowledge of a sample surface is very difficult to attain even in situ; submonolayer overlayers and roughness are always present, even in the cleanest in situ environment. Fortunately, they can be estimated by comparison with ex situ data at room temperature, and subtracted mathematically using the Jones formalism for multilayer analysis. The caveat here is that usually surface roughness and oxide layers are essentially indistinguishable in treating the effects

of a thin overlayer of unknown thickness [32.178–180]. Usually an oxide overlayer is subtracted from the data to obtain the bulk dielectric function.

Since all in situ SE measurements of epilayer and substrate properties are based on the dielectric function library used, the measuring accuracy and precision largely depends on that library. Bearing in mind the caveats discussed above, *Johs* [32.181] developed a procedure to measure dielectric function libraries of growing materials, in particular of HgCdTe during MBE growth. He confronted with great success the problem of determining a library of dielectric functions for different Cd molar fractions, from $x = 0.2$ to 0.5, at the ideal growth temperature T_g. His paper discusses the difficulties that one encounters in measuring a compositional library $\langle\varepsilon(\omega; x, T_g)\rangle$ in great detail. Because $\varepsilon(\omega)$ depends strongly on T as well as x, the very narrow growth window for HgCdTe turns out to be very helpful in obtaining $\langle\varepsilon(\omega; x, T_g)\rangle$. Because of the very narrow growth window, the measured thickness of the surface roughness layer increases rapidly as the growth temperature departs from the ideal. This allows one to adjust the temperature to T_g even without knowing its value and thus to hold the uncertainty in the measured $\langle\varepsilon(\omega; x, T_g)\rangle$ to a minimum.

The compositional control of HgCdTe attained using the library developed by Johs has been excellent, as reported by a large number of authors [32.182–188]. Attention has not been focused on the real-time variations of the composition (which are inaccessible by routine FTIR measurements in any case), but only on its average value over the entire layer. This was obtained as follows: a set of 8–10 measured dielectric functions was suitably selected to span the entire growth process. A global fit was performed, treating the angle of incidence as a global parameter and leaving composition and surface roughness free to attain different values for each spectrum. Finally, the obtained x-values were averaged to get the average composition.

A fit to $\langle\varepsilon_2\rangle$ yields average compositional values that are in good agreement with those obtained by FTIR. The fact that a fit to the measured values of $\langle\varepsilon_2\rangle$ gives the best ellipsometric measurement of x is due to two factors:

1. $\langle\varepsilon_2\rangle$ contains much more direct information on the joint density of states (JDS) than either $\langle\varepsilon_1\rangle$ or Δ and Ψ
2. $\langle\varepsilon_2\rangle$ is much less affected by variations in the window birefringence.

These two factors also assure that $\langle\varepsilon_2\rangle$ gives a very accurate measurement of x when proper care is taken and when the library used is appropriate for the growth temperature and growth and SE measurement configurations used. However, it should be emphasized that the excellent results found for compositional monitoring and control require fully reproducible growth conditions. Changes in the geometry of the growth setup, the growth rate, the sample mounting, etc. have been shown to lead to substantial errors in the HgCdTe compositions measured using the library function of Johs. Also, in situ measurements of the composition y in CdZnTe as a function of temperature using a library of $Cd_{1-y}Zn_yTe$ dielectric functions have been performed for three samples with $z = 0.0$, 0.045, and 0.125, starting at room temperature, going up to 350 °C, and coming back down to 150 °C. He found that the composition measured upon coming back down in temperature from 350 °C was very different from than that measured going up in temperature, ≈ 0.05 higher over the range 150–300 °C for the CdTe sample. He also measured the temperatures using the same library, and found errors in the measured temperatures from −10 to −30 °C for the CdTe sample upon coming back down in temperature.

The acquisition of a two-dimensional (2-D) library $\langle\varepsilon(\omega; x, T)\rangle$ of HgCdTe pseudodielectric functions as a function of the two variables x and T is much more difficult than might be expected a priori. This is because high-quality HgCdTe only grows within a ±5 °C window, so that to get a temperature library one must interrupt the growth and ramp the temperature to different values. This must be done leaving the layer under Hg flux to avoid irreparable damage to the surface. We have found no systematic dependence of the CP energies on T, in data taken with this procedure, probably because of Hg evaporation, which alters the composition near the surface, producing random changes in the dielectric function and in any measured CP energies. No alternative procedure has been proposed to date.

These difficulties may be overcome in the near future, for example by a more accurate control of the Hg flux. However, even if a two-dimensional library were obtained, one must still conclude that one could not obtain x and T simultaneously from such a library without SE data greatly improved over that provided by the 88-wavelength RAE now commonly used, which does not provide the resolution necessary to allow the accurate determination of CP energies through multi-

ple differentiation of the data. It has been shown that a fit to the undifferentiated data leaving four parameters free (x, T, t, and ϕ) yields a reasonably good estimate for T, but also substantially increases the uncertainty in x and t. A better approach would be based on the interplay of a 2-D library for $\langle\varepsilon(\omega; x, T)\rangle$ and a knowledge of the energy $E_1(x, T)$ obtained by multiply differentiating that library with respect to ω. The energy E_1 could be obtained very precisely and would be almost independent of any surface overlayer. One could find T as a function of x from a measured E_1, reducing a two-dimensional fitting problem to a one-dimensional problem using the undifferentiated library. One cannot simultaneously determine x and T by measuring two CP parameters without an ellipsometer capable of simultaneously measuring E_0 and E_1 accurately. No such ellipsometer is available at present. Measurements of the other E_i do not give sufficient information; $E_1 + \Delta_1$ and $E_0 + \Delta_0$ cannot be measured sufficiently precisely, and $E_1 + \Delta_1$ moves almost in parallel with E_1 with changes in x or T. Also, the line widths Γ_i depend strongly on material quality and cannot be measured as precisely as the E_i.

A new set of libraries of $\varepsilon(\omega; T)$ for $Cd_{1-z}Zn_zTe$ for $z = 0.0$, 0.045, 0.125, and 1.0 is available from MPL. That library ranges from 75 to 300 °C in increments of 25 °C with increments of 15 Å in wavelength from 1.2 to 5.0 eV and increments of 30 Å from 0.7 to 1.2 eV. The library was obtained by ramping the temperature of each CdZnTe sample up in 25 °C steps, pausing at each step to allow the sample to come to a steady state, and then measuring $\varepsilon(\omega)$ at that temperature. The following procedure [32.151] was used. Clean $Cd_{1-z}Zn_zTe$ was introduced in the MBE chamber after etching in a 0.5% bromine and methanol solution, and rinsing in methanol and deionized (DI) water. The surface oxide was desorbed at ≈ 250 °C, the substrate was ramped to different temperatures, and its pseudodielectric function $\langle\varepsilon(\omega; T, \phi, t)\rangle$ was measured at each temperature T. The temperature at the surface was measured with a thermocouple, and the temperature measurements were calibrated according to the melting points of Sn and In. A fit to room-temperature data gave the angle of incidence ϕ and the thickness t of the surface roughness overlayer, using calibrated data acquired ex situ. This procedure allowed us to obtain a set of spectra $\varepsilon_{CZT}(\omega; T_1, \ldots, T_n)$ with overlayer effects removed. The library was obtained using an M-2000 ellipsometer, which has 220 wavelengths from 1000 to 1700 nm, allowing the differentiation of the data to determine the E_0 and E_1 CP energies using a CP model, as discussed immediately below.

One can never ensure that one is measuring the optical constants of an abruptly terminated, clean, and smooth semiconductor surface. All that can be ensured is that our optical constants are reproducible, under the same conditions, on the same system. One cannot ensure the absolute accuracy of a library of measured optical constants, because of window birefringence and the presence of overlayers, but one can expect one's measurements to be reproducible from run to run on the same system. However, system-to-system portability of a library is never certain.

Analysis Based on Critical Point (CP) Models

An analysis of the dielectric function which includes a CP analysis in principle is superior to an analysis based on libraries, for the following reasons. First, much of the information obtainable from the derivatives of the pseudodielectric function is complementary to that obtainable from fits to the function itself and is not available from fits to the function itself. That information consists of the energies E_i and intrinsic line widths Γ_i of all band-structure CPs within the spectral range of the SE data. Second, the use of a parameterized theoretical model for the dielectric function allows us to measure directly the phase shift of the incoming light in SE as it passes though a surface overlayer and thus to measure directly the optical thickness of the overlayer, which is not otherwise possible. The E_i are virtually independent of surface overlayers and, given sufficiently good SE data, can be obtained very precisely.

High accuracy in the determination of the E_i and Γ_i is paramount in in situ applications. The problems with CP analyses are that the data fitting is more difficult and that low-noise, high-resolution SE data is required to perform the higher-order differentiation needed to obtain accurate values for the CP parameters. It has been found [32.189] that it is impossible to perform a useful CP analysis using only the 88 wavelengths available with an M-88 ellipsometer with the typical signal-to-noise ratio (SNR) of in situ data. However, the use of an M-2000 ellipsometer in situ has been found to yield sufficient resolution and a sufficiently small noise level to allow the profitable use of a CP model for data analysis.

Several CP models have been proposed, but only the Kim model [32.190] is capable of simultaneously fitting the optical dielectric function and its derivatives. The optical dielectric function is given by the fundamental

expression [32.191]

$$\varepsilon(\omega) = 1 + \frac{8\pi^2 e^2}{m^2} \times \sum_{c,v} \int W_{cv}(E)\,dE \times \left\{ \left[\Phi(\hbar\omega - E) - \Phi(\hbar\omega + E) \right] \right\} . \quad (32.6)$$

In this expression, $W_{cv}(E) = P_{cv}(E)^2 J_{cv}(E)$ represents the product of the Brillouin-zone-average transition probability at the energy E and a joint density of states (JDS), $J_{cv}(E)$. $\Phi(\hbar\omega \pm E)$ is the broadening function and is equal to $\left|\hbar\omega \pm E + i\Gamma\right|^{-1}$ for Lorentzian broadening; in general, the broadening is intermediate between Lorentzian and Gaussian [32.192]. The JDS is given by the formula

$$J_{cv}(E) = \int_{BZ} \delta\left[E_c(\boldsymbol{k}) - E_v(\boldsymbol{k}) - E \right] d\boldsymbol{k} , \quad (32.7)$$

and counts the number of states per unit energy between a given pair of conduction (c) and valence (v) bands that differ by an energy E. The JDS and hence the $W_{cv}(E)$ are smoothly varying functions of energy away from the critical points.

The Kim model parameterizes the fundamental expression (32.6) by approximating the $W_{cv}(E)$ by low-order polynomials in energy multiplied by functions of energy having the proper analytic behavior at the CPs. Thus the model has two parameters for each CP, E_i and Γ_i, along with amplitude parameters with respect to which the dielectric function is linear. $W_{cv}(E)$ is divided into seven contributions, one for each pair of connected CPs. The resultant JDS is shown schematically in Fig. 32.10. The expression for each region

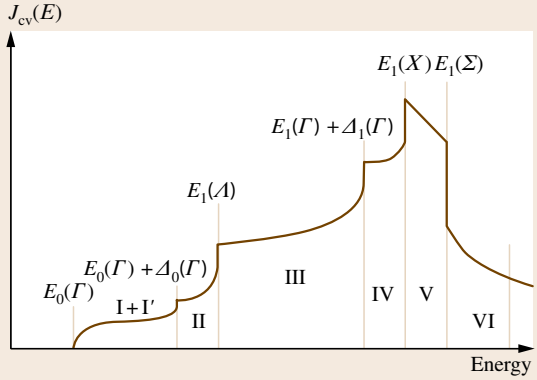


Fig. 32.10 Schematic diagram of the joint density of states of a zincblende semiconductor

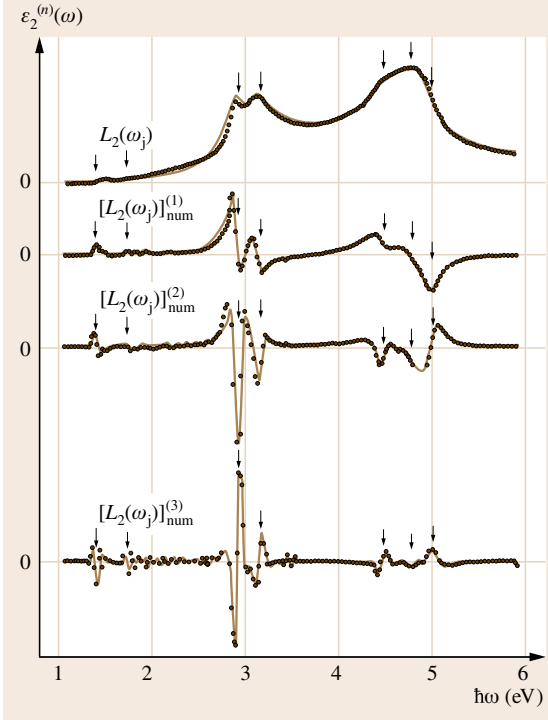


Fig. 32.11 Simultaneous fit of the Kim, et al. model to the experimental $\varepsilon_2(\omega)$ for GaAs and its first three numerical derivatives allowing a mixture of Lorentzian and Gaussian broadening for the four lowest-energy critical points. The *arrows* show the energies of the seven critical points used in the fit. The *black dots* show the experimental $\varepsilon_2(\omega)$ and its numerical derivatives; the *red curves* show the modeled $\varepsilon_2(\omega)$ and its numerical derivatives

in Fig. 32.10 is given in [32.190] and is fitted independently. The expressions for the seven contributions are substituted in (32.6) and integrated analytically. A proper harmonic oscillator is also added to account for out-of-range contributions to $\varepsilon_1(\omega)$.

The Kim model allows:

1. Predictions for the region below the fundamental gap, but above the onset of phonon transitions
2. The measurement of the true JDS
3. The measurement of the thickness of overlayers using the measured phase shifts at the CPs.

It was initially tested by simultaneously fitting $\varepsilon_2(\omega)$ for GaAs and its first three derivatives; the results are shown in Fig. 32.11; the fit to $\varepsilon_2(\omega)$ was good to 0.4%. It was later used to fit the dielectric functions of CdTe [32.193] and ZnSe [32.194] and their derivatives and to obtain

functional forms for the temperature-dependent dielectric functions of CdTe [32.195] and HgCdTe [32.196] and the composition dependent dielectric function of AlGaAs [32.197]. It was found that, with a total of 75 free parameters, the measured $\varepsilon_2(\omega)$ for $Al_{1-x}Ga_xAs$ and its derivatives for all compositions could be fitted by functions of x and ω within a fractional root-mean-square (rms) error of only 1.2% in $\varepsilon_2(\omega; x)$. When measuring in situ, fewer parameters should be used because the data are affected by much larger uncertainties.

One value of using the Kim model is its failure to fit the measured $\varepsilon_2(\omega)$ for materials having poor crystalline quality or having overlayers not explicitly included in the fitting through the use of a multilayer model. This means that it cannot give incorrect CP energies to compensate for the effects of overlayers or poor crystalline quality, so the CP energies given are very reliable. This is illustrated by its failure to fit the alloy series $Hg_{1-y}Zn_yTe$ and $Cd_{1-y}Zn_yTe$. Fits to samples having different alloy compositions [32.198] failed to yield a systematic dependence of the room-temperature Γ_i on y over the compositional range $0 < y < 1$. The poor material quality of these alloy systems led to fluctuations in the CP line widths with alloy composition that made it impossible to obtain functional forms for their composition-dependent dielectric functions. That result highlights the sensitivity of this model to even small changes in CP parameters. To be able to monitor such changes opens new horizons for the in situ monitoring of HgCdTe growth.

Comparison of Results from Fitting to a Library and from CP Analysis Using the Kim Model

Consider a few cases in which both library functions and a CP analysis have been used in fitting data obtained from an M-2000 ellipsometer to measure composition or temperature. Abad found that the CP energies found for CdZnTe from the derivatives of the measured dielectric function are essentially unaffected by surface changes caused by annealing, even though those changes cause significant changes in the dielectric function itself. Also, *Badano* et al. [32.199] recently have studied the effect of surface roughness on the ellipsometric response from CdTe epilayers having up to ≈ 200 Å-thick surface roughness with a correlation length of several micrometers. SE data were taken ex situ at room temperature to reduce the noise in the data and the CP line widths, so as to improve the precision of measurement of the CP energies. Although the surface roughness introduced noticeable changes in the dielectric function, the CP energies found were unaffected by the roughness. Since the sample surface temperature can be defined solely by the CP energy E_1, this confirmed that the sample surface does not affect the temperature measured using the Kim model. These results, which argue in favor of using CP analysis, are not unexpected, because the dielectric function of a microscopically disordered surface does not have sharp derivatives.

Also, *Badano* et al. [32.200] recently have compared the use of library functions and the use of the Kim model for the measurement of the temperature of CdTe epilayers grown on Ge. The layers were ramped up and down in temperature as in the study by Abad, but only up to ≈ 250 °C, and were measured under a Te flux and under no flux. The reproducibility obtained using the two different data analysis methods, upon ramping the temperature up and then back down, were similar, except that the Te flux affected the values obtained for T using library functions but not those obtained using the Kim model. An error analysis of the CP fitting procedure suggested that noise in the data should introduce a random error of approximately ±5 K in the value of T found using the Kim mode. However, the values of T obtained using the Kim model were reproducible within ±2 K, an outstanding result. The values for T found using a library were systematically higher than those found using the Kim model. Since values obtained using a library are much more subject to systematic error while those obtained from the Kim model are much more subject to random error, it would be reasonable to ascribe this systematic difference to systematic error arising from the use of library functions.

On the other hand, even with the use of an M-2000 ellipsometer, the noise in SE data induced random dispersions in y and T of the order of ±0.005 in y and ±4 K in T for $Cd_{1-y}Zn_yTe$. Those uncertainties are somewhat too large, although substantially better than those found using a library of dielectric functions. On the other hand, it suggests that by measuring y ex situ at room temperature and then, knowing y, measuring T in situ, one could obtain more precise values than are now available. The uncertainties would be substantially reduced by not having to determine y and T simultaneously, and the uncertainty in y would be further reduced by measuring it ex situ. However, the large variations in y which sometimes exist from point to point on a CdZnTe substrate would make this method problematic, because the in situ and ex situ measurement points on the CdZnTe may not be identical. Also, Abad found the Kim model to be substantially more difficult to use than library fitting with the software provided with the ellipsometer. This result is discouraging

with regard to the possibility of being able to measure x and T in situ simultaneously for HgCdTe, although the method proposed for HgCdTe is different, involving the determination of x from a library of dielectric functions and T from the determination of the CP energy E_1.

In conclusion, the use of the Kim model eliminates errors arising with the use of library functions due to changes in incident flux, surface damage and surface states, window coatings, etc., but introduces uncertainties due to the relatively large amount of noise in in situ SE data. Although for measuring temperature those uncertainties appear to be smaller than the errors which easily can occur with the use of library functions, they may need to be reduced further for SE to become a valuable tool for measuring T. Also, the use of the Kim model entails fits that are not as robust and may converge slowly, a serious problem for the in situ control of growth. The use of the Kim model for the in situ determination of substrate temperature with an M-2000 ellipsometer is promising, but would require a more robust fitting procedure to be justified in a well-controlled production environment, in which the use of library functions determines T reproducibly even if not accurately.

However, the use of the Kim model would have two other advantages, namely the determination of the CP line widths and, unlike fitting to library functions, the determination of absolute values for the surface roughness or the thickness of other overlayers. Results from Badano et al. show that the use of a CP model can pinpoint the presence of overlayers on a sample, and does not fit SE data unless a proper model for such an overlayer is adopted. We envision the possibility of acquiring libraries of true bulk dielectric functions, by eliminating the roughness overlayer from the data, and possibly also the effect of other overlayers due to the presence of the impinging fluxes. More work is being done in this field.

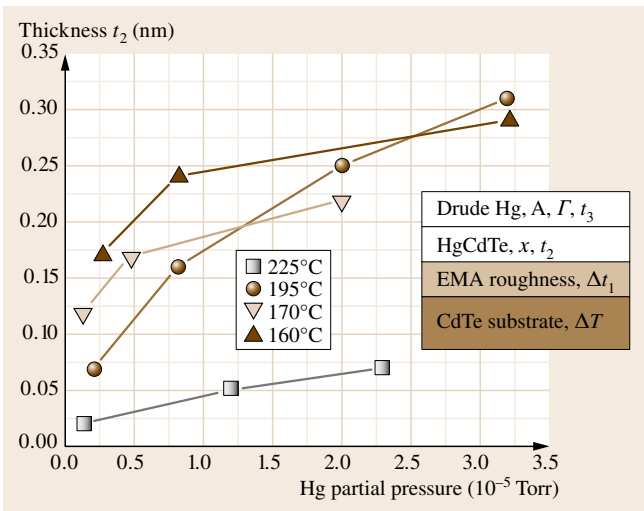


Fig. 32.12 Calculated thickness t_2 of the chemisorbed Hg layer, obtained as a function of the Hg flux for different substrate temperatures. The uncertainty in each thickness obtained from the correlation matrix is approximately ±0.4 Å. The values for the Hg partial pressure are uncorrected N_2 pressure readings. The data at 195 °C were acquired during a separate run than the data taken at 225, 160, and 170 °C. In the 195 °C run, the sample was exposed to Te flux down to a lower temperature, thus it was likely to have a higher Te coverage. The *inset* shows the assumed structure and the fitting parameters used in fitting the SE data

32.5.3 SE Study of Hg Absorption and Adsorption on CdTe

SE has been shown [32.201,202] to measure the relative abundance of chemisorbed and physisorbed monolayers of Hg on the CdTe substrate surface prior to growth nucleation. Data with submonolayer resolution were obtained. The substrate dielectric function was measured after the substrate was prepared using a standard procedure, but before subjecting it to Hg flux. Then, the Hg flux was opened and the dielectric function was remeasured. In the analysis of the change in the pseudodielectric function due to Hg deposition, two major approximations were made. First, the surface roughness and the chemisorbed and physisorbed Hg were treated as separate layers with the physisorbed Hg on top of the chemisorbed Hg, and, second, the three-dimensional (3-D) form of $\varepsilon(\omega; x)$ for $Hg_{1-x}Cd_xTe$ for large x was used to mimic the effect of chemisorbed Hg in the surface region. Thus, the data were interpreted using a five-layer model comprised of CdTe substrate, surface roughness, chemisorbed Hg, physisorbed Hg, and ambient. The quantities left free in the fit to the data were the thicknesses of the two Hg layers, that of the roughness layer, and the plasma frequency ω_p in the Drude form assumed for the dielectric function of the physisorbed Hg layer. The values found for the relative thicknesses of the chemisorbed and physisorbed layers are physically reasonable and show the expected dependence on the temperature and the Hg flux, as shown in Fig. 32.12. Although the absolute numbers provided by the analysis undoubtedly are not accurate, the results confirm our general ideas regarding the nature of the CdTe(211) surface under a Hg flux, and is valuable at least techno-

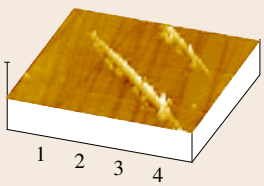
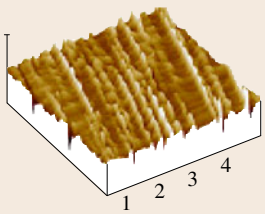
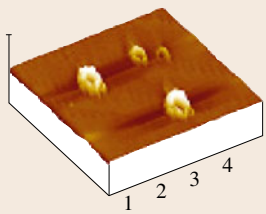
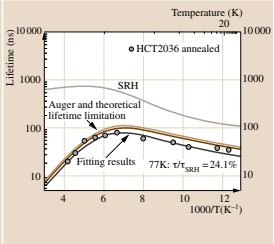
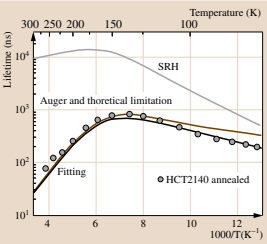
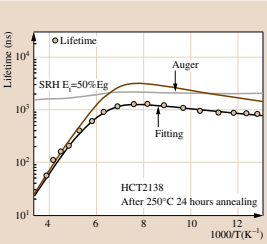
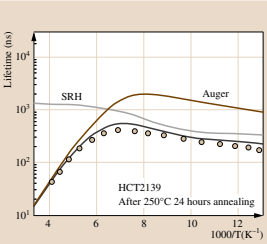
logically to obtain a reliable run-to-run characterization of the surface before growth.

32.5.4 Correlation Between the Quality of MBE-Grown HgCdTe and the Depolarization and Surface Roughness Coefficients Measured by in situ SE

The SE surface roughness thickness parameter t measured using an effective medium approximation (EMA) for the roughness is routinely used to monitor growth. The surface roughness measured by SE during growth is a good indicator of the final quality of HgCdTe layers. Substantive increases in t are always associated with the onset of 3-D growth and the degeneration of material quality, mostly associated with the formation of twins. In addition, the substrate submicroscopic roughness as measured by SE correlates well with the qualitative roughness inferred from reflection high-energy electron diffraction (RHEED). Due to the nature of the EMA model, it is sensitive to roughness on an atomic scale and can give information about how small-scale defects form. However, it does not model roughness on a micrometer scale.

The evolution of surface roughness with increasing epilayer thickness for CdTe grown under nonideal conditions has been studied [32.199]. Both the use of the Kim model and the use of library functions gave surface roughness values which initially increased with increasing epilayer thickness and then decreased. It was found that the apparent decrease arose from the onset of macroscopic roughness, which could not be fitted by the EMA. For rough samples with roughness on much larger than a nanometer scale, the measured dielectric function departs significantly from the form

Table 32.3 Correlation between the depolarization coefficient $\bar{u}$, the surface morphology, the SRH contribution to the minority carrier recombination rate, the EMA surface roughness, and the etch pit density

Sample#	HCT2036	HCT2140	HCT2138	HCT2139
Surface morphology				
Depolarization coefficient u	u_0 u_{end} Δu –5 –5 ±5	u_0 u_{end} Δu –5 5 ±4	u_0 u_{end} Δu –5 18 ±4	u_0 u_{end} Δu –5 25 ±4
EMA t (Å)	t_0 t_{end} 0 2	t_0 t_{end} 0 2.5	t_0 t_{end} 0 5	t_0 t_{end} 0 5
Temperature dependent lifetime				
τ/τ_{Auger} @77K (ns)	35/38	199/328	794/1361	188/778
τ/τ_{SRH}@77K	24.1%	36.7%	56.2%	70.1%
EPD (cm^{-2})	2.0×10^5	9.1×10^5	2.7×10^6	4.9×10^6

it takes with only atomic-scale roughness because the EMA is not suitable for describing larger-scale roughness with lower lateral spatial frequency. Because the RAE is very sensitive to depolarization effects, we have introduced a model for the analysis of RAE SE data, which contains a depolarization coefficient parameter.

The depolarization coefficient u for the RAE is defined by the equations

$$\cos 2\psi = (1+u)\cos 2\psi' \quad \text{and}$$
$$\cos\Delta = \cos\Delta'\left(1+\frac{1}{\sin^2\psi'}u\right), \tag{32.8}$$

where ψ' and Δ' are the values of ψ and Δ measured by SE in the presence of depolarization, and ψ and Δ are the values in the absence of depolarization [32.203]. The quantity u is the ratio of the intensity of the light scattered incoherently to that of the specularly reflected light [32.204]. Obviously, u is a function of wavelength or frequency, $u = u(\omega)$, but no measurements of $u(\omega)$ are available in the literature for rough HgCdTe or CdZnTe. We define $\bar{u}$ as the value of u averaged over the wavelength range of our SE data. It measures surface roughness on a scale larger than that of the EMA surface roughness, a lateral scale more nearly of order the wavelength of the incident polarized light.

We have discovered correlations between the minority-carrier recombination lifetime τ and $\bar{u}$. It correlates well enough with τ to be regarded as a useful predictor during growth. Typical results showing the correlation between $\bar{u}$ measured in situ, the surface morphology, the EPD, and the relative importance of Shockley–Read–Hall recombination processes are shown in Table 32.3. This suggests that the measurement of $\bar{u}$ can be developed as an in situ early warning of the degeneration of material quality, so that the MBE growth conditions can be reoptimized to stop the degeneration of layer quality, and hence improve the material growth yield.

32.5.5 Surface Characterization by in situ RHEED

In general, reflection high-energy electron diffraction (RHEED) has been the most important analytical tool available to the MBE grower, although SE has now taken over that role for the growth of HgCdTe. RHEED employs electron guns and display units such as fluorescent screens and/or recording devices such as charge-coupled device (CCD) cameras to obtain information on the structure and/or morphology of crystal surfaces. It is installed on virtually all MBE systems, and various authors have given detailed discussions of applications of RHEED in MBE growth [32.205]. It is particularly well suited for the in situ characterization of the growth surface during MBE growth. MBE provides the high vacuum and clean surface required for RHEED measurements; in return, RHEED does not interfere with the incoming particle beams during growth and is particularly sensitive to the top few layers or even the top monolayer. It uses relatively high electron energies (5–100 keV, typically 20 keV for HgCdTe) and very low impact angles ($< 5°$). The high energy of the electron beam makes the reflected image from an ordered surface sharp, and the low impact angle allows the electrons to pass only a few atomic layers into the crystal, giving an image that represents the structure of the surface instead of that of the whole crystal. RHEED signals from single-crystal surfaces that are flat on a nanometer scale produce a RHEED pattern with lines rather than dots because a 2-D array has a one-dimensional Fourier transform. Thus, the RHEED image of a smooth ordered surface shows sharp one-dimensional (1-D) rods, rather than zero-dimensional (0-D) dots. In contrast, an oxide substrate surface layer, being amorphous,

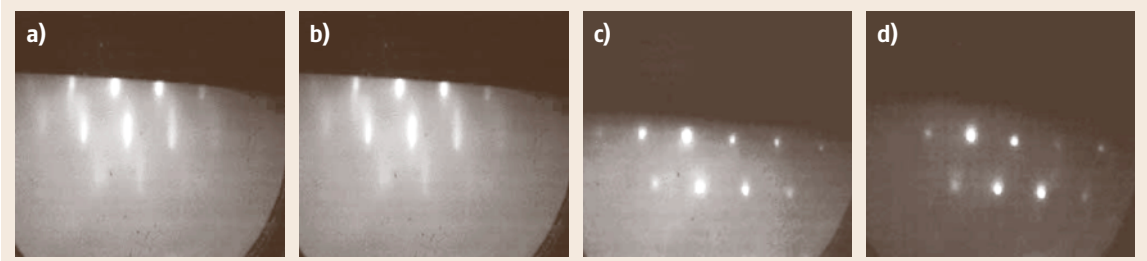


Fig. 32.13a–d RHEED patterns from CdZnTe substrates for different values of the surface roughness as measured by SE: **(a)** 3 Å, **(b)** 10 Å, **(c)** 20 Å, and **(d)** 30 Å

gives rise to a diffuse diffraction pattern, and relatively rough surfaces produce a RHEED pattern having dots rather than rods. The effect of CdZnTe substrate surface roughness is shown in Fig. 32.13. The HgCdTe (211)B MBE surface normally has several atomic layers of surface roughness, so the actual RHEED image normally has reflection and transmission contributions, which makes the quantitative analysis of the image very difficult. On the other hand, the analysis of RHEED data from 2-D growth on low-index surfaces has attracted great interest, with over 500 publications on oscillations in the RHEED intensity alone. Such oscillations, with a period equal to the growth period, were first observed in the early 1980s in the growth of GaAs [32.206–208], and in 1989 were observed in the (100) growth of HgTe/CdTe SLs [32.209]. They were first explained quantitatively in 1998 [32.210] and remain the subject of theoretical study even today [32.211].

32.5.6 Other in situ Tools for Controlling the Growth

Other commonly used in situ tools for controlling the growth include a heater and thermocouple behind the substrate to control the growth temperature and a thermocouple and/or ion flux gauge on each effusion cell to control the flux rates. The difficulty in determining the growth temperature from the thermocouple reading was discussed above in Sect. 32.4.1. Because of that difficulty, an optical pyrometer sometimes is used either to calibrate the thermocouple or to directly control the growth temperature. Because of the extreme sensitivity of the fluxes on effusion cell temperatures, the use of flux gages (and valved cells) is recommended for any fluxes which need to be well controlled. Finally, in situ FTIR, although not commonly used, can be used for temperature monitoring [32.152].

32.6 Nucleation and Growth Procedure

The nucleation of HgCdTe growth always occurs on either CdTe, CdZnTe with $\approx 4\%$ Zn or CdSeTe with $\approx 4\%$ Se. The more difficult nucleation process is the nucleation of CdTe, CdZnTe, CdSeTe or ZnTe on Si or another severely non-lattice-matched substrate material. We consider first the nucleation of CdTe or ZnTe on Si and then the nucleation and growth of HgCdTe.

32.6.1 Nucleation and Growth of CdTe or ZnTe on Si

The MBE growth of CdTe on Si has been studied for over a decade [32.54, 93, 97, 99, 212–215]. The primary issues to be faced when growing Te-based II–VI compounds on Si(211) are the high density of surface steps, which exacerbates the difficulty of removing oxygen and other contaminants from the surface, controlling the polarity of the II–VI layer to be grown, and finally reducing the dislocation density in that layer. Substrate preparation is a crucial step in the growth of Te-based II–VI compounds on Si. The two most important techniques for preparing silicon surfaces for low-defect epitaxy are oxidation and hydrogen passivation. Surface oxidation primarily serves to oxidize the adventitious surface contaminants to facilitate their low-temperature removal. Wet chemical techniques such as the RCA technique [32.216] or a modified RCA technique [32.217], which leave an oxide film ≈ 12 Å thick on the surface, are effective in reducing contamination by metals and surface carbon contamination. One such technique is to clean the Si surface in a mixture of $NH_4OH : H_2O_2 : H_2O$ (1 : 1 : 5, 85 °C), etch the oxide in dilute HF (2%), and reoxidize the surface in a solution of $HCl : H_2O_2 : H_2O$ (1 : 1 : 5, 65 °C). However, these techniques require the oxide layer to be thermally desorbed in situ at temperatures ≥ 850 °C. Hydrogen passivation describes the formation of a self-protective layer on silicon surfaces upon exposure to HF acid. The hydrogen passivation layer has the important advantage of being volatile at ≈ 560 °C. A hydrogen passivation treatment has been developed [32.218] that greatly reduces the oxygen and carbon contamination of the Si surface without an extensive prebaking or the high-temperature thermal desorption required after RCA cleaning. This treatment allows the growth of epitaxial layers with reduced dislocation densities, but is substantially more difficult than the RCA treatment.

After thermal desorption of the passivant layer, the substrate is quickly cooled to ≈ 450–500 °C under an As_4 or cracked arsenic flux and then cooled to the nucleation temperature of ≈ 300–340 °C with either a Te_2 flux or no incident flux. This leaves a monolayer of As on the surface of the Si. The absorption of a precursor As layer has been shown to greatly reduce the sticking coefficient of Te on the Si surface and increase its surface mobility [32.219–221] and to

be essential to the MBE growth of ZnTe [32.219] or CdTe [32.220] with good growth morphology. A model has been proposed [32.221] for the effect of the As, and density-functional calculations of the geometrical and electronic structures of As and Te on Si(211) have been performed [32.222].

Different groups have developed a variety of recipes for the remainder of the growth procedure. Either a thin ZnTe buffer layer (≈ 60–500 Å thick) is grown by migration-enhanced epitaxy [32.96, 97, 99] at 230–300 °C or a thicker layer (0.1–1 μm) is grown by MBE at ≈ 220 °C and then annealed at ≈ 360–500 °C under a Te_2 flux before the final thick CdTe layer is grown at ≈ 300 °C. The growth of an As precursor and a ZnTe buffer ensures the growth of CdTe (211)B on Si (211)B. To further enhance the quality of the final CdTe epilayer, a ZnTe/CdTe strained-layer SL buffer could be grown before the CdTe growth. It also has been reported [32.97, 223] that the quality of the CdTe growth is enhanced by periodically interrupting the growth and flash annealing at ≈ 530 °C or periodically annealing ≈ 10 min at ≈ 360–380 °C under a Te_2 flux. Reported CdTe growth rates vary from 0.7 μm/h at 300 °C to 1.0 μm/h at 270 °C. Both the desorption of the Si passivation layer and the subsequent growth are monitored by RHEED and SE.

32.6.2 Substrate Preparation and Growth of HgCdTe

The following nucleation and growth procedures are those followed at MPL and at EPIR Technologies; similar procedures are followed at other laboratories and production facilities. A 12 keV RHEED system and an SE system are used routinely to monitor the substrate preparation and HgCdTe growth in situ. All substrates are first degreased in two separately heated trichloroethylene baths followed by rinses in two methanol baths. They are then prepared for HgCdTe growth by etching in a bromine/methanol solution, followed by several methanol and DI water rinses. This leaves a Te-rich surface. For CdZnTe substrates, the etch is performed for 20 s in a 0.5 vol. % solution to remove the top 1000–2000 Å of the CdZnTe, which may have polishing damage. For composite Si-based substrates, it is performed for 5 s in a 1 vol. % solution to remove 500–1000 Å of the CdTe, CdSeTe or CdZnSeTe top layer. Finally, the substrate is dried with nitrogen and loaded into the MBE system. Then it is slowly heated to 300 °C with no incident flux. At ≈ 135 °C the surface Te becomes mobile, decreasing the intensity of the rods in the RHEED pattern and causing dots to appear. At 195–205 °C the Te begins to evaporate, causing further changes in the RHEED pattern. At ≈ 220 °C the surface Te has evaporated, and the RHEED pattern again becomes streaky with rods. RHEED patterns taken during CdZnTe substrate preparation are shown in Fig. 32.14a–c. For the CdZnTe substrates, the Te shutter is opened at 300 °C, and the temperature is increased to ≈ 320 °C. The substrate is maintained at ≈ 320 °C for 5–10 min and then is slowly cooled down to 300 °C, where 200–500 Å of CdTe is grown to create a smooth surface. At 300 °C an additional thin CdTe layer is grown on the Si-based

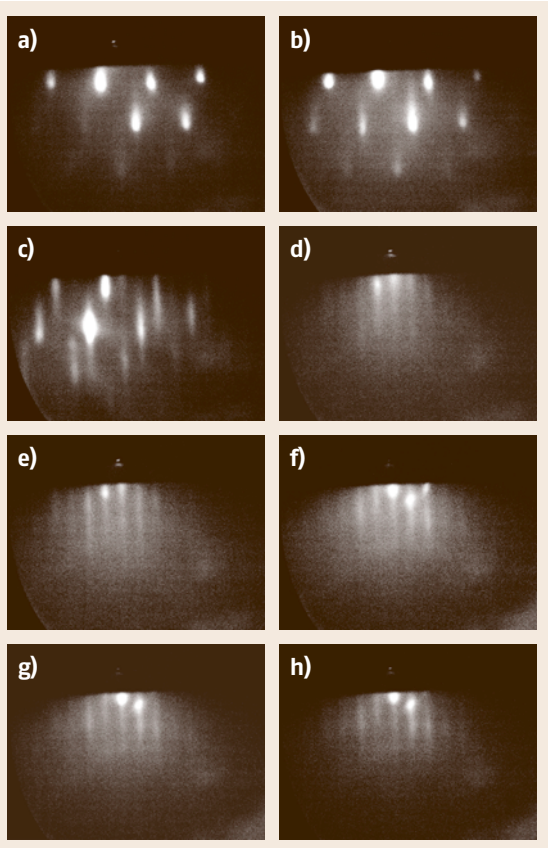


Fig. 32.14a–h In situ RHEED pictures of a HgCdTe sample grown on a CdZnTe substrate during MBE growth: **(a)** CdZnTe substrate as loaded, **(b)** after removal of Te and its oxide, **(c)** after CdTe regrowth and partial cooling towards the growth temperature, **(d)** at the moment HgCdTe growth is started, **(e)** 30 s of growth, **(f)** 1 h of growth, **(g)** about 2 h of growth, **(h)** after 3 h 45 min (end of growth)

substrates and is annealed at $\approx 320\,^{\circ}\mathrm{C}$. This procedure ensures a smooth and clean initial surface prior to HgCdTe nucleation.

After the final CdTe growth, all substrates are cooled to the HgCdTe growth temperature, $185 \pm 5\,^{\circ}\mathrm{C}$, as measured by calibration of the thermocouple behind the substrate. Four temperatures can be used to calibrate the thermocouple readings:

1. The onset of Te surface mobility on CdTe at $\approx 135\,^{\circ}\mathrm{C}$
2. The onset of Te evaporation at $\approx 195\,^{\circ}\mathrm{C}$
3. The melting temperatures of In droplets on the surface of calibration substrates at $156\,^{\circ}\mathrm{C}$
4. The melting temperatures of Sn droplets on the surface of calibration substrates at $232\,^{\circ}\mathrm{C}$.

The growth of HgCdTe is then begun. A Hg flux is present even before the growth shutters are opened, so that the growth starts with a monolayer of Hg. The growth proceeds at a rate of 2–3 μm/h with $\approx 3\times10^{-4}$–3.5×10^{-4} Torr Hg flux, 10^{-6} Torr Te_2 flux, and 2×10^{-7} Torr CdTe flux. The growth temperature is continuously adjusted to minimize the roughness of the growing surface as measured by SE, and the CdTe flux is continuously adjusted to maintain the SE-measured composition at the targeted value. The evolution of the RHEED pattern during the MBE growth of a HgCdTe on CdZnTe substrate is shown in Fig. 32.14d–h. If the growth temperature and/or growth rate are too high, Te precipitates and crater defects (also called voids) form due to the greatly reduced Hg sticking coefficient; if too low, microtwinning occurs and surface hillocks appear. To obtain lateral uniformity of the growth, the sample is continuously rotated and the growth fluxes are incident off center at an angle to the normal to the growth surface.

Due to initial island formation, the roughness of the growth initially increases up to a knee point corresponding to a critical growth thickness. Both the SE surface roughness and the RHEED intensity of a specular spot were studied at MPL during the early stages of growth at temperatures from 170 to 196 °C. The knee was found to occur in the first minute of growth. At the lowest temperatures the increase in roughness continued beyond the knee, but much more slowly, at intermediate temperatures a very slow recovery was observed beyond the knee, and at higher temperatures an exponential decay with a time constant ≈ 3 min was observed. After the HgCdTe growth was paused, a recovery was observed under Hg flux at all temperatures. The critical thickness is plotted for two representative samples in Fig. 32.15, hct2352 having a higher growth rate than hct2387 at the same temperature. The critical thickness of layer hct2387 shows a much weaker dependence on temperature than layer hct2352. Their critical thicknesses are similar above 185 °C, but layer hct2352 has higher critical thicknesses below 185 °C. Those differences can be described in terms of the dimensionless ratio $\alpha = Jw^2/Dn_0$. This is the ratio of the input adatom flux J times the square of the (111) terrace width w to

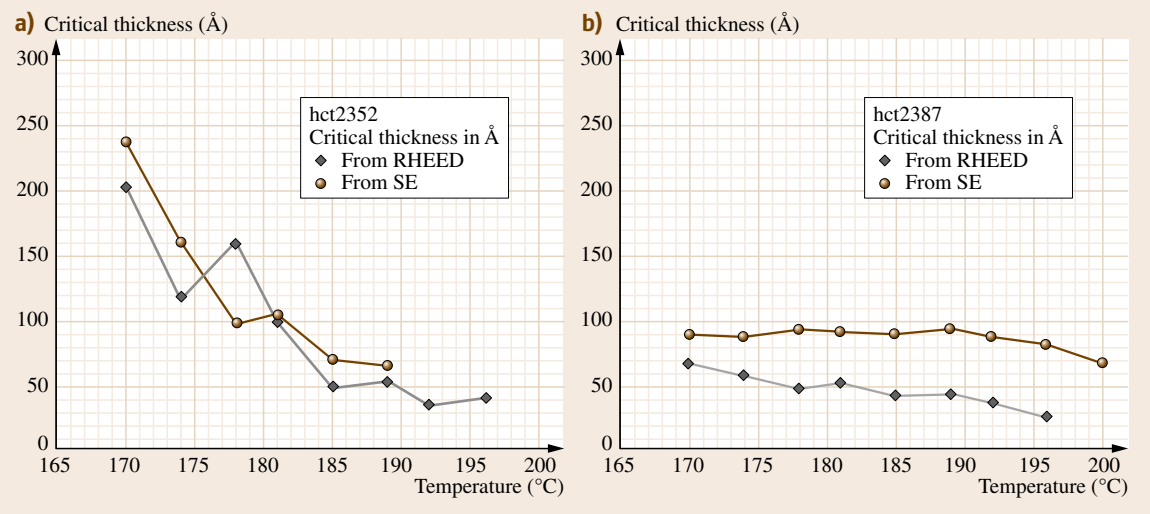


Fig. 32.15a,b Temperature dependence of the critical thickness during the initial stages of growth. The results shown in **(a)** are from a sample with a higher growth rate than that in **(b)**

the diffusion constant D times n_0, the concentration of lattice sites for island formation. The results shown in Fig. 32.15 can be explained as follows:

1. At higher temperatures (small α), the rate of clustering is low, and growth proceeds largely by the incorporation of mobile adatoms into step edges. Because most nucleation occurs at the bottoms of step edges, the activation energies for nucleation and island formation are predominantly mechanical, temperature-insensitive energies The critical thicknesses would be predicted to be of the same order as the step heights, ≈ 20 Å.
2. At lower temperatures (large α), the adatoms are less mobile, so the island concentration builds up until the islands largely overlap. The lower the temperature, the lower the adatom mobilities, and thus the steeper the islands and the greater the island heights before overlapping becomes important. Therefore, the critical thickness increases rapidly as the temperature is decreased. The islands and step edges merge, and less densely populated fresh roughened terraces are formed (Stranski–Krastanov growth). The build-up of islands is then repeated, but to a lesser extent, until a steady-state island concentration is reached when the nucleation rate is balanced by the capture of islands due to convection.
3. At very low temperatures (very large α), thermal diffusion is very low and hence the steepness of islands on the terraces is so large that the surface becomes extremely rough and continues to grow rougher even after reaching the critical thickness, eventually saturating and becoming time independent (3-D growth).

With a relatively small increase in temperature, 3-D growth turns into terraced hillock growth (modified S-K-M) and then into step-flow growth, as explained by the above discussion. It also explains why, as the input flux or the substrate roughness increases, step-flow growth requires a higher substrate temperature (lower α) due to a higher level of nucleation.

32.7 Dopants and Dopant Activation

The active components of both currently used and advanced IR photodetectors being developed, as well as other optoelectronic devices, use photovoltaic p–n junctions. In order to form sharp, precisely controlled interfaces and p–n junctions and to obtain well controlled doping levels, it is essential to grow such junctions by MBE with extrinsic n- and p-type doping. Although the stable, well-controlled n-type extrinsic doping of as-grown HgCdTe and HgTe/CdTe SLs has been demonstrated by many authors, the p-type extrinsic doping of these materials has remained a problem. In particular, the reproducible p-type activation of As, the best low-diffusivity p-type dopant in these materials, has repeatedly been reported to occur reproducibly only at annealing temperatures sufficiently high to vitiate many of the advantages of MBE as a growth technique. Even though almost all device fabricators have demonstrated the ability to fabricate high-performance HgCdTe-based photovoltaic photodiodes using various p-doping techniques, interdiffusion across interfaces, dopant diffusion, and the formation of Hg vacancies all become problems at the high annealing temperatures required for the reproducible p-type activation of group V dopants. The absence of a reproducible p-type doping technique which does not require high-temperature annealing remains a major stumbling block to successful implementation of high-yield fabrication of MBE-grown HgCdTe IR FPAs. This problem, attempts to solve it in the past, current progress, and alternative p-type dopants are reviewed briefly here.

32.7.1 Extrinsic n-Type Doping

Stable, well-controlled n-type extrinsic doping has not proven to be a major problem. The incorporation of In during MBE growth has been used to achieve this goal for as-grown HgCdTe [32.224–226] over a wide range of dopant concentrations (10^{14}–10^{19} cm^{-3}) and for HgTe/CdTe SLs [32.227–229] over the somewhat smaller range from 10^{15} to 10^{18} cm^{-3}. A low-temperature postgrowth anneal can be used to optimize the structural and electronic properties of the doped material, but is not required for the n-type activation of the dopant In atoms. The only problem in achieving well-controlled n-type doping arises at very low doping levels from the intrinsic p-type doping which arises from Hg vacancies. For very small In concentrations, these vacancies can compensate the n-type conduction, lowering the mobility and destroying the otherwise precise control of the doping. However, Hg diffuses easily through interstitial sites in HgTe and IR-absorber HgCdTe material, even in perfect material, having an

activation energy of only about 0.67 eV for diffusion via interstitial sites [32.230], and diffuses even more rapidly along dislocation cores [32.231]. This allows the Hg vacancy density to be reduced to values below $10^{13}\,cm^{-3}$ by low-temperature annealing under a high Hg vapor overpressure, essentially eliminating this problem, although it could reoccur during subsequent device processing steps or during long-term storage.

32.7.2 Extrinsic p-Type Doping

Extrinsic doping is necessary:

1. Because intrinsic p-type doping by Hg vacancies is not stable against diffusion [32.230, 232, 233] from p- to n-type regions
2. Because the density profile of Hg vacancies cannot be well controlled
3. Because Hg vacancies significantly degrade carrier mobilities [32.234] and act as SRH recombination centers for minority carriers, shortening the carrier lifetimes [32.225, 235–239] and hence decreasing the quantum efficiency of detectors having p-type absorption layers.

On the other hand, achieving the stable, well-controlled extrinsic p-type doping of MBE-grown HgCdTe and HgTe/CdTe SLs without negating the advantages of MBE growth offers difficult and fundamental challenges. The primary problem arises because the MBE growth of HgCdTe is optimally performed under Te-rich conditions [32.226, 233, 240–242], so that Hg vacancies are the dominant native point defects in as-grown HgCdTe. Although their density typically [32.239] is only $\approx 10^{14}\,cm^{-3}$, it is many orders of magnitude higher than the density of Te vacancies in as-grown HgCdTe [32.230,232,243–245]. Therefore, in equilibrium, group V atoms in as-grown HgCdTe are less likely to be located on anion (Te) sites than on cation (Hg or Cd) sites, where they form neutral complexes or act as donors rather than acceptors. This has been confirmed by both theory [32.230, 232, 243–245] and experiment [32.224, 225, 233, 241, 242, 246–251]. (Of course, even for very high As concentrations, the number of cation vacancy sites remaining to be filled by As atoms is essentially independent of the number already filled by As atoms.) On the other hand, group I atoms, which also are located primarily on cation sites, and thus are p-type activated, suffer from high diffusivities.

Postgrowth As Incorporation

Extrinsic p-type doping can be achieved either by the ex situ implantation or diffusion of dopant ions into n-type layers or by their in situ incorporation during MBE growth. As is the preferred p-type dopant for implantation or in situ incorporation during MBE growth. As implantation [32.252–255] is the key procedure of the planar device fabrication process, and has been employed in most IR FPA fabrications. However, ion implantation induces considerable damage, which reduces both minority-carrier lifetimes and carrier mobilities, and therefore is detrimental to device performance. A high-temperature ($\geq 425\,°C$) anneal is usually performed after ion implantation. This anneal serves a dual purpose:

1. It activates the As as a p-type dopant
2. It reduces the residual implantation-induced damage.

This technology has been used successfully for the fabrication of mid-wave IR (MWIR) FPAs. However, such a high-temperature anneal itself is detrimental in several ways: it leads to the possibility of introducing a much higher density of Hg vacancies – at 500 °C the equilibrium density of Hg vacancies has been calculated to be above $10^{18}\,cm^{-3}$ even under a 1 atm Hg partial pressure [32.244, 245] – and both interdiffusion across interfaces and dopant diffusion become problems, vitiating primary advantages of the MBE growth technique. Furthermore, even the residual damage remaining after annealing is severely detrimental to the performance of long-wavelength IR (LWIR) devices, due to their narrower bandgaps – bandgap states induce significant tunneling in devices with very narrow bandgaps, resulting in large dark currents. An alternative method [32.256, 257] for the postgrowth incorporation of As is to diffuse it in from the surface. However, that requires an even more deleterious thermal treatment. In addition, ion implantation and As indiffusion can be performed only in-plane, which dramatically limits the possible detector structure designs for which these techniques are applicable. The next generation of IR detectors, including single-focus multicolor detectors and high-operating-temperature LWIR detectors, cannot be fabricated in-plane. They will require in situ incorporation of extrinsic p-dopants during MBE growth and their activation as-grown or after only low-temperature ($< 250\,°C$) annealing.

32.7.3 In situ Group I Dopant Incorporation

In principle, either group I or group V elements could act as acceptors in HgTe or HgCdTe – group I elements on cation (Hg or Cd) sites or group V elements on anion (Te) sites. However, unlike liquid-phase epitaxial growth, the MBE growth of HgTe and narrow-gap HgCdTe must be performed under Te-rich conditions [32.224, 233, 241, 242], so that during growth the number of cation vacancies is several orders of magnitude higher than the number of Te vacancies [32.230, 232]. Thus, group I dopant atoms in these materials reside almost entirely on cation sites and act as singly ionized p-dopants in the as-grown material, unlike group V dopant atoms. For this reason group I elements were the first to be tried as extrinsic p-dopants. The incorporation of Cu, Ag, Au, and Li as p-type dopants in HgCdTe epilayers and their diffusivities have been studied by many authors [32.224, 227, 233, 239, 258–269]. Even though the doped materials showed close to 100% activation and excellent transport properties and minority-carrier recombination times, all of these dopants proved to be unstable in HgCdTe, diffusing out of the doped region during further MBE growth or during the low-temperature postgrowth anneal. This behavior would appear to prevent their use in FPA technology and hence limit their usefulness as an acceptor dopant, especially for the abrupt junctions required for IR detectors. Thus, they suffered a period of relative neglect in the late 1980s. However, Au in particular is again being studied intensively as a p-type dopant, as its diffusivity may be able to be held within acceptable bounds. We have found no perceptible diffusion of Au dopants in a HgTe/CdTe SL after 3 years of storage at room temperature.

32.7.4 In situ Group V Dopant Incorporation

In recent years, because of the high diffusivities of group I elements and the damage caused by ion implantation, studies on the p-type doping of HgCdTe and of HgTe/CdTe SLs have focused on in situ doping with group V elements, especially with As [32.224, 229, 262, 270–286], because it is more easily activated than the other group V elements. The larger size of the group V atoms, and hence their lower diffusivity, allows the growth of stable, well-controlled p–n junctions, and thus sharp interfaces and/or precisely controlled dopant profiles. The incident As flux in most older experiments was uncracked and consisted almost entirely of As_4 tetramers, which are physisorbed as tetramers or clusters of tetramers, not as single atoms [32.40], and which calculations suggest [32.270] remain as tetramers in as-grown HgCdTe. The use of cracked As has become common, so that the As is mostly chemisorbed in atomic or diatomic form, as happens with the use of a CdAs flux. This gives an As sticking coefficient orders of magnitude higher but appears not to ease the problem of obtaining p-type activation of the As without a high-temperature anneal.

32.7.5 As Activation

Because the As atoms in as-grown MBE-grown HgCdTe reside primarily on cation sites rather than Te sites, annealing under a Hg overpressure is required to activate the As as a p-dopant by transferring it to Te sites. Several authors have proposed models for the activation process and performed calculations [32.242, 244, 245, 271, 272], all assuming As incorporation as isolated atoms. The standard activation anneal is 10 min at $\approx 425\,^{\circ}\mathrm{C}$ followed by 24 h at $\approx 250\,^{\circ}\mathrm{C}$ under a Hg overpressure to fill the Hg vacancies created during the previous anneal. The anneal at a temperature well above $350\,^{\circ}\mathrm{C}$ induces Te evaporation, which allows the As atoms to move to the resultant Te vacancies in the second-stage anneal. This annealing procedure yields almost complete activation of the As up to concentrations of order 10^{18}–$10^{19}\,\mathrm{cm}^{-3}$ depending on the temperature of the first-stage anneal [32.244, 245, 273, 274], with the doping abruptly saturating at that level, presumably due to all of the Te vacancies created during the first-stage anneal being filled.

Unfortunately, this annealing procedure introduces undesirable broadening of interfaces and dopant diffusion, and thus limits many of the advantages of MBE as a low-temperature growth technique. Therefore, a variety of approaches have been used in attempts to obtain the p-type activation of As either as-grown or after only a low-temperature anneal. Although each of these approaches has shown promise, none have consistently given full activation of the As. One approach [32.242, 247, 275, 276] is to use a two-stage anneal with $T \leq 300\,^{\circ}\mathrm{C}$ in both stages, with Te vacancy–Te interstitial pairs formed in the first stage and As atoms filling the Te vacancies in the second stage. Another approach [32.250] is to force As atoms out of cation vacancy sites by using a high Hg overpressure in both stages of a two-stage low-temperature anneal. The lack of consistently reproducible results obtained using these approaches is not understood. A rather thorough study

of As incorporation, activation, and diffusion is given in [32.250].

Other approaches have involved either modified MBE growth methods or different methods of As incorporation. One approach [32.277–280] was to irradiate the growing HgCdTe with a laser beam (photo-assisted MBE, PAMBE), shifting the growth toward cation-rich conditions by causing Te desorption and thus allowing As incorporation on the vacant Te sites. This was combined with the idea of growing HgTe/CdTe SLs and then forming HgCdTe by annealing the SLs to interdiffuse the Hg and Cd. That allowed the As to be incorporated only during the CdTe growth, when the number of cation vacancies is much lower so that it is easier to incorporate the As on Te sites. This will yield almost complete p-activation of the As, but at a high price. PAMBE has been abandoned because it limits the growth to impractically small areas. Also, annealing the SL broadens any device interfaces. Another approach [32.281, 282] was to use a Te cracker cell so as to increase the Hg sticking coefficient and obtain As activation as-grown. Success was reported for (111) and (211) growth, but appears not to have been consistently reproducible.

An approach in which only the method of As incorporation is changed, not the growth technique, is that of planar doping. In planar doping the As flux is shut off most of the time, but periodically the CdTe and Te fluxes are shut off and the As flux is turned on, along with the Hg flux, which is never shut off. This approach was originated [32.283] in 1995, has been used many times [32.255, 283–286] (and in unpublished results of the authors) and has been observed to give full p-activation of the As in the as-grown material [32.283, 284]. Although full activation in the as-grown material is not observed consistently, at least partial activation is observed, and full activation appears to be consistently obtainable by combining planar doping with anneals at $\leq 300\,^{\circ}\mathrm{C}$. More consistent activation is obtainable in the planar doping of CdTe or of the CdTe layers of a HgTe/CdTe SL.

In conclusion, although progress is being made in solving the problem of obtaining well-controlled fully reproducible p-type doping of HgCdTe without the use of undesirable high-temperature anneals, this remains one of the major problems in the growth of material for advanced HgCdTe-based IR detectors and focal-plane arrays. The most promising avenue of attack on this problem appears to be planar doping with As coupled with the development of an optimal sequence of low-temperature anneals.

32.8 Properties of HgCdTe Epilayers Grown by MBE

The most important epilayer properties for detector and FPA performance are the minority-carrier lifetime, the carrier concentration and mobility, the perfection of the epilayer surface (to ensure the absence of dead pixels), and the lateral uniformity of the epilayer (to ensure uniformity of pixel response). Substrate imperfections, poor lattice matching and thermal matching between the substrate and the epilayer, and nonoptimal growth conditions all can degrade these properties. We consider first the electrical and optical properties.

32.8.1 Electrical and Optical Properties

Minority-carrier recombination processes in HgCdTe directly decrease photon detection and increase noise generation in photovoltaic infrared devices. Hence, it is especially important to maximize the minority-carrier lifetime τ. The three commonly recognized recombination mechanisms – Auger, radiative, and Shockley–Read–Hall (SRH) – are characterized by recombination lifetimes that are usually measured by fitting photoconductive decay data, which measure the total recombination lifetime $\tau = (\tau_{\mathrm{Auger}}^{-1} + \tau_{\mathrm{radiative}}^{-1} + \tau_{\mathrm{SRH}}^{-1})^{-1}$. One can determine each recombination rate separately by fitting the measured values of τ as a function of T. Of these three mechanisms, two (Auger and radiative recombination) are determined by intrinsic material properties, such as the composition and doping level, which can be precisely measured by other techniques, e.g., the Hall effect and infrared absorption mapping [32.287]. The Auger and radiative lifetimes can be calculated from those measured quantities using sets of semi-empirical formulas developed by *Beattie*, *Landsberg*, and *Blackmore* [32.288, 289] based on Kane's $\boldsymbol{k}\cdot\boldsymbol{p}$ model with several key parameters determined experimentally [32.290]. These intrinsic mechanisms establish upper limits on the lifetime. The differences between the calculated and measured lifetime values are attributed to SRH recombination, which arises from impurities or defects and can be fitted within the theoretical framework given by *Shockley*, *Read*, and *Hall* [32.291, 292]. A primary criterion for a high-

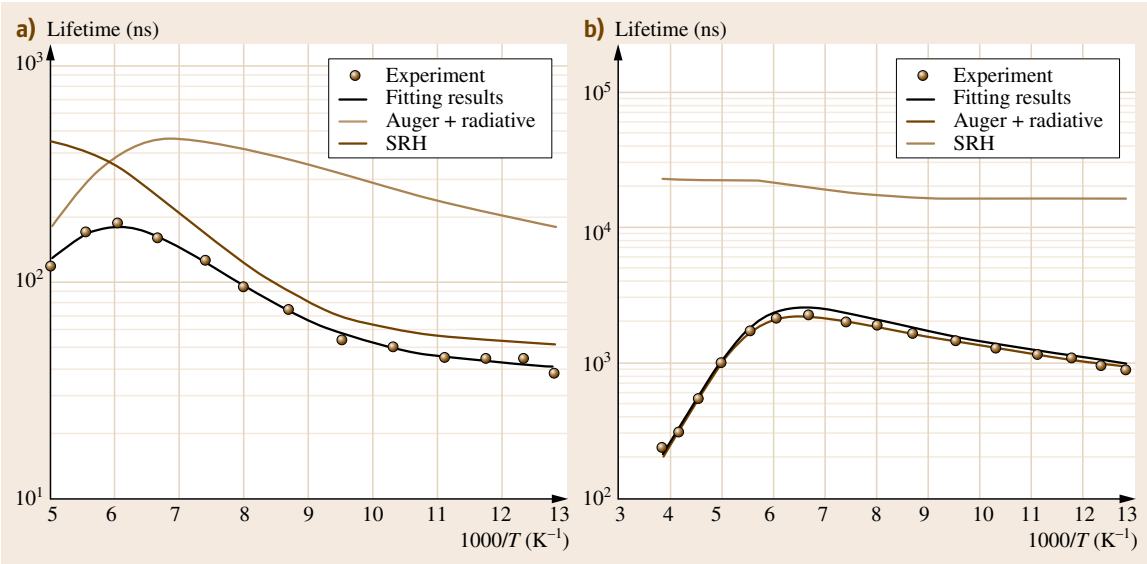


Fig. 32.16a,b Two fits to τ as a function of inverse temperature (T^{-1}) showing the measured values, the fitted intrinsic (*dotted*) and SRH (*solid*) lifetimes, and the fit to the measured values of τ; **(a)** shows a case in which SRH recombination dominates below 150 K, and **(b)** shows a case in which the SRH recombination rate is small and hence less well determined

quality layer is $\tau_{SRH} \gg \tau$. Two fits to τ are shown in Fig. 32.16: (a) with SRH processes dominant below ≈ 150 K and (b) with SRH processes negligible. Note the difference in the dependence of τ on T^{-1} in the two cases. Also, note that in Fig. 32.16b τ_{SRH} is very poorly determined because SRH recombination is so unimportant in that case.

In general, it is found that, for growth on CdZnTe substrates, the strong Auger recombination dominates the determination of τ and thus limits detector performance for p-type layers. However, intrinsic lifetimes are much longer in n-type materials, so it is more difficult to achieve the condition $\tau_{SRH} \gg \tau$. Others [32.293] have found that SRH recombination is dominant for low In doping levels in n-type MWIR HgCdTe samples, but that τ increases with increasing doping, suggesting that In may passivate the SRH defects for MWIR HgCdTe. On the other hand, we have found at MPL that lifetimes $\approx 1\,\mu$s dominated by intrinsic recombination can be reproducibly achieved in n-type LWIR material with doping levels $\approx 10^{15}\,\text{cm}^{-3}$ after annealing under a Hg-rich atmosphere to reduce the concentration of Hg vacancies, which act as SRH recombination centers.

Two other electrical characteristics of a grown HgCdTe epilayer are important: the carrier concentration and the mobility. In the absence of compensation (the presence of only n- or only p-type dopants) both are routinely determined from Hall-effect measurements. However, in the presence of compensation their determination requires Hall-effect measurements as a function of magnetic field as well as temperature [32.293]. For LWIR $Hg_{1-x}Cd_xTe$ layers with $x = 0.23$ grown on Si-based substrates, at 77 K mobilities $\mu \approx 10^5\,\text{cm}^2/\text{Vs}$ are consistently obtainable for n-type layers with $n \approx 10^{15}\,\text{cm}^{-3}$, $\mu \approx 500\,\text{cm}^2/\text{Vs}$ for p-type layers with $p \approx 1.3\times10^{16}\,\text{cm}^{-3}$, and $\mu \approx 200\,\text{cm}^2/\text{Vs}$ for p-type layers with $p \approx 2.0\times10^{16}\,\text{cm}^{-3}$. For layers of the same composition grown on CdZnTe substrates the electron mobilities reported are lower by a factor of ≈ 2–3.

The primary optical characteristic of HgCdTe epilayers is the IR transmittance or absorptivity, which is measured by FTIR spectroscopy. A general review of the IR optical characterization of HgCdTe and of FTIR spectroscopy in general is given elsewhere [32.294]; here we outline how FTIR measurements can be used to assess epilayer quality. Below the energy gap, impurity absorption can be measured and carrier concentrations can be measured by free carrier absorption. By scanning transmission/reflection mapping, the uniformity of composition and thickness across a wafer can be meas-

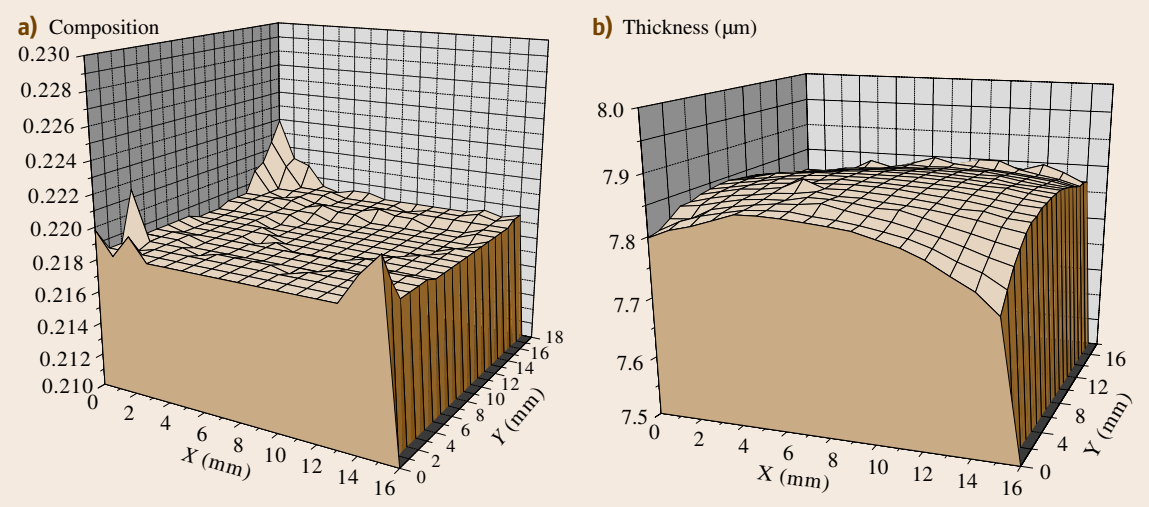


Fig. 32.17a,b Mapping results for a HgCdTe epilayer grown on CdZnTe: **(a)** composition and **(b)** thickness

ured. Typical results obtained at MPL for growth on square 20 mm × 20 mm CdZnTe substrates are shown in Fig. 32.17; the average HgCdTe composition was $x = 0.2182$ with a standard deviation of 0.0006, and the average thickness was $d = 7.84\,\mu\text{m}$ with a standard deviation of 0.03 μm [32.287]. Higher compositions and thinner material was found in about $2 \times 2\,\text{mm}^2$ areas in each corner, presumably due to the low thermal conductivity resulting in a higher growth temperature in the corners and hence a lower Hg sticking coefficient there. This explanation is supported by the existence of a higher density of voids or crater defects near the corners; these are found primarily in Te-rich or higher-T growth. For growth on 3 inch-diameter Cd/Si substrates, there was no evidence of nonuniformity in T, and the nonuniformities in x and d over the central $20 \times 20\,\text{mm}^2$ area were much smaller than for growth on CdZnTe, although over the entire samples somewhat larger nonuniformities were observed, presumably due to flux nonuniformities.

Of greater importance, the Urbach tail energy, which arises primarily from electron–phonon interactions, alloy disorder, and structural disorder, has been shown to be useful in characterizing material quality [32.295]. The absorption coefficient below the energy gap can be expressed as [32.296, 297]

$$\alpha(\omega) = \alpha_0 \exp\left[\frac{\hbar\omega - E_0}{W(T)}\right], \tag{32.9}$$

where $W(T)$ is defined as the Urbach tail energy. It measures the steepness of the band tail and can be obtained by fitting the the absorption data below the energy gap. We obtained values for $W(T)$ by measuring $\alpha(\omega)$ from 4 K to room temperature and fitting the logarithmically linear part of $\alpha(\omega)$ below the bandgap. Those samples with surface microtwin defects arising from low growth temperatures or with high dislocation densities were seen to consistently show relatively high Urbach tail energies having large standard deviations over small areas, suggesting that nonuniformly distributed structural disorder dominates the nonuniformity of device performance figures of merit in HgCdTe IR FPAs. Moreover, Urbach tail energy mapping measurements we have performed correlate strongly with Cd composition mappings and with crystalline defects revealed by x-ray topograph mappings. The contribution of the electron–phonon interaction was calculated and found to be much smaller than the contribution of alloy and structural disorder and to give the observed small temperature dependence of $W(T)$ [32.295]. We also calculated the contribution of alloy disorder, $W_{\text{alloy}} = x(1-x)\text{VBO}^2 m^*_{\text{HH}} L^2/7.2\hbar^2$, where x is the alloy composition, VBO is the valence band offset, and L is the alloy disorder correlation length (estimated to be of the order of one interatomic spacing). It also was found to be much smaller than the contribution of structural disorder. $W(T)$ also was found to be only weakly dependent on T for CdZnTe substrates, for which the alloy contribution also is very small because of the small Zn concentration. Thus, $W(T)$ serves as a quantitative measure of structural disorder for both HgCdTe epilayers and CdZnTe substrates.

32.8.2 Structural Properties

The structural properties of HgCdTe depend greatly on the type of substrate used for its growth. For (211)B growth on CdZnTe, substrate defects are the primary cause of structural defects in the HgCdTe. For (211)B growth on CdTe/ZnTe/Si, the lattice and thermal mismatches at the ZnTe/Si and ZnTe/CdTe interfaces are primary causes of structural defects in the HgCdTe. The lattice mismatches are much greater than any mismatch resulting from thermal changes, but the thermal mismatches can cause strains much greater in magnitude than the residual lattice-mismatch strains remaining after relaxation by dislocation formation during growth. For simplicity consider a CdTe/Si interface. Following *Carmody* et al. [32.110], the equilibrium residual strain in a CdTe film of thickness h is $\varepsilon(h) = (h_c/h)\varepsilon_{misfit} \approx 10^{-4} \times \varepsilon_{misfit}$ for a film of thickness 5 μm, where $h_c \approx 5$ Å is the equilibrium critical thickness and $\varepsilon_{misfit} = -0.193$ is the lattice-mismatch strain, giving a residual stress $\sigma_{resid} \approx 1.4$ MPa. Of course, the CdTe does not come to a complete equilibrium during growth, but even so the stress remaining at the growth temperature at the end of growth is much less than the thermal stress resulting from cooling to room temperature, which is ≈ 45 MPa. Also, the relaxation of the strain induced during cooldown will be much less complete than the relaxation of the lattice-mismatch strain because of the much shorter time at an elevated temperature. Thus, incomplete relaxation upon cooldown can be a greater source of strain than the large lattice mismatch.

The most common measure of crystalline perfection is the XRD FWHM. Under optimal growth conditions the XRD FWHM of HgCdTe grown on CdZnTe substrates is limited only by the substrate FWHM, at least for (112)B, (113)B, and (552)B orientations, and is ≈ 20 arcsec [32.44]. However, the XRD FWHM is substantially larger for HgCdTe grown on CdTe/Si substrates. From 1991 to 1995 the best values obtained were reduced from 210 to 65 arcsec, but little progress has been made in achieving further reduction. In fact, typical values of the FWHM are still above 100 arcsec, and the best consistently obtainable value is ≈ 100 ± 30 arcsec, limited by the FWHM of the underlying CdTe layer [32.96]. One possible source of the 65 arcsec XRD line broadening is incomplete lattice relaxation giving rise to changes in the lattice constant in the growth direction as one moves away from the surface. This possibility is supported by an anticorrelation between the density of threading dislocations and the FWHM; if true, it would explain why no further significant decrease in the FWHM has been observed since 1995. The primary sources of the extra width observed before 1995 were microtwinning and low-angle grain boundaries due to growth at nonoptimal orientations and with nonoptimal buffer layers. The extra width above 65 arcsec still usually seen today probably arises in large part from low-angle grain boundaries.

Another measure is the etch pit density, which measures the density of threading dislocations at the HgCdTe surface. Dislocations adversely affect HgCdTe electrical properties and device performance in many ways [32.298]; the deleterious effects of dislocations were discussed briefly in Sect. 32.3.3. Threading dislocation densities below the mid $10^5\,cm^{-2}$ are not a serious problem unless they form clusters or bend over to form misfit dislocations parallel to the surface under device contacts [32.298]. Growth on CdZnTe yields etch pit densities of order low 10^4 to low $10^6\,cm^{-2}$, depending primarily on the substrate defect density. Growth on CdTe/ZnTe/Si typically yields etch pit densities of order mid 10^6 to mid $10^7\,cm^{-2}$. As discussed in Sect. 32.3.3, this density of threading dislocations to date has prohibited the use of Si-based substrates for the growth of HgCdTe for LWIR and VLWIR detectors and FPAs.

Other structural defects commonly encountered before the last decade, such as two-phase growth, twinning and microtwinning, and low-angle grain boundaries, have been greatly reduced by choice of growth orientation, proper substrate preparation and the use of buffer layers, and improved control over growth conditions. However great care must be exercised to maintain the level of these defects at an acceptable level, especially twinning, which occurs at low growth temperatures T or high Hg flux, and can cause harmful surface defects. Under optimal growth conditions T is high enough that the surface diffusion length of incoming atoms is approximately the terrace width. Then, growth nucleation occurs at kinks or step edges, and step-flow growth occurs. If T is too low or in the presence of defects, growth can also nucleate on the terraces, so that growth from different points must coalesce, forming grain boundaries or laminar microtwins.

32.8.3 Surface Defects

The common surface defects are voids or crater defects, large triangle defects (like a type of void and treated here under the same heading), faceted microvoids or small triangle defects (similar to microvoids and in-

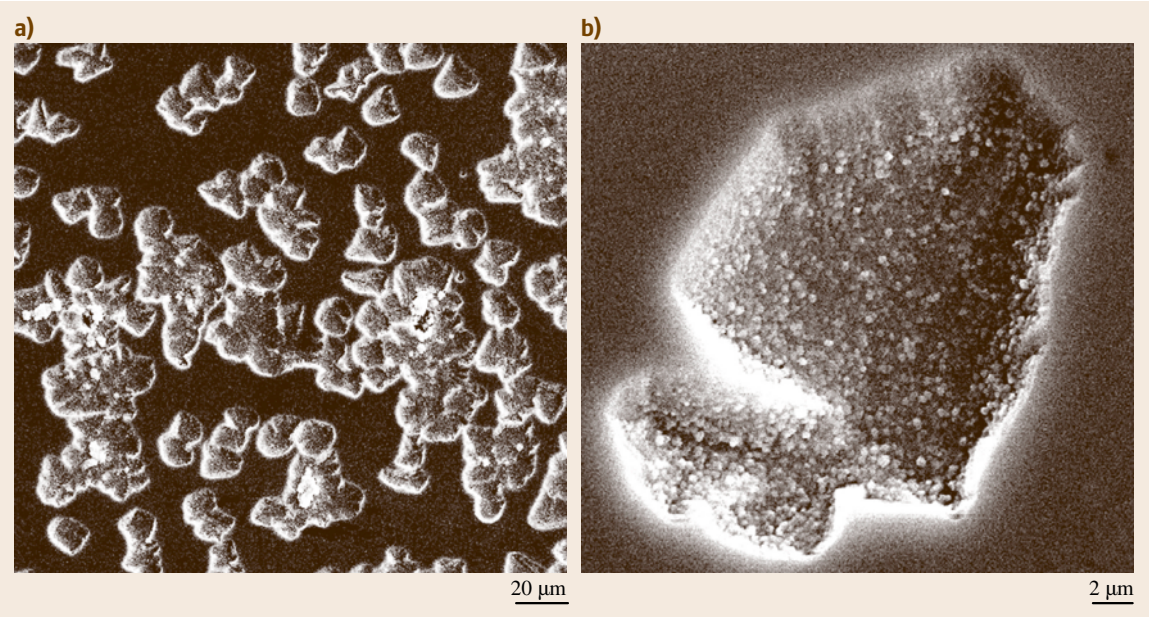


Fig. 32.18a,b SEM micrographs showing **(a)** large crater defects on a HgCdTe/CdZnTe (211)B epilayer grown on a poor substrate and **(b)** the typical morphology of a large crater defect

cluded here under the same heading), flakes, hillocks, needle defects, cross-hatching, and surface roughness. The reduction of the density of surface macroscopic defects (voids, microvoids of size 3–5 µm, and flakes) is one of the most serious challenges in MBE growth of HgCdTe. Voids or crater defects, discussed briefly in Sect. 32.3.2 and shown in Figs. 32.5 and 32.18, are the surface manifestations of threading defects associated with Te precipitates. Small and mid-size voids arise during growth from incomplete Te_2 dissociation (low T) or excess Te under Te-rich growth conditions (high T, low Hg flux or high growth rate) [32.39, 299–301]. For growth on CdZnTe substrates, large voids can arise from dust or other particulate matter on the substrate surface [32.183, 302] or Te precipitates in CdZnTe substrates [32.66, 69]. Detailed descriptions of their origin and morphology are given in [32.300] and [32.301]. They range in size from 2 to 40 µm [32.302] and can bypass p–n junctions and give premature reverse-bias breakdowns or even blind sensor elements [32.303]; hence, they are extremely harmful to FPA performance. They can be minimized by tight control of the growth conditions [32.39, 304] and by the proper preparation of CdZnTe substrates [32.66–70]. Typically, the void or crater density is mid to high $10^2\,cm^{-2}$ for growth on Si-based substrates and low 10^2 to high $10^3\,cm^{-2}$ for growth on CdZnTe substrates [32.38], depending strongly on both the substrate quality and the growth conditions. The minimum reproducible void density is high $10^2\,cm^{-2}$ for growth on Si-based substrates and $\approx 10^3\,cm^{-2}$ for growth on CdZnTe substrates.

Faceted microvoids have a sharp triangular shape and smooth, slightly raised surfaces. For growth on Si-based substrates, they are typically 3–5 µm long, their density is $\leq$ low $10^2\,cm^{-2}$, and they almost always result in electrical shorts at 140 K, blinding sensor elements [32.303]. At EPIR Technologies, 1 µm-long faceted microvoids have been found on the ZnTe and CdTe epilayers in CdTe/ZnTe/Si substrates and have been identified with microtwinning. To the best of the authors' knowledge, there has been no report of faceted microvoids for HgCdTe grown on CdZnTe. It is reasonable to assume that the microvoids observed on the surface of HgCdTe epilayers grown on CdTe/ZnTe/Si arise from microvoids on the substrate surface.

Like faceted microvoids, flakes have been reported only for growth on Si-based substrates. They have been investigated at the US Army Research Laboratory [32.96], the Night Vision Electronic Sensors Directorate [32.96], and EPIR Technologies. A typical flake defect is shown in Fig. 32.19. Flakes on the HgCdTe epilayers are known to result from correspond-

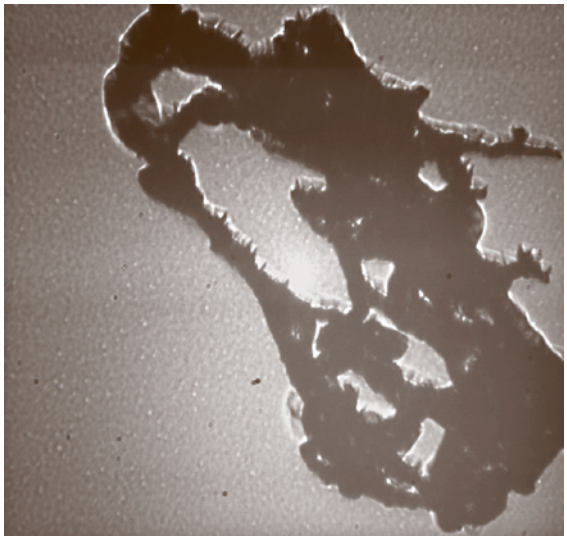


Fig. 32.19 An optical micrograph of a typical flake defect

ing flakes on the top CdTe layers of the substrates; they are completely a substrate problem. Unlike other defects, they have no known intrinsic origin. They arise primarily from sputtering from the CdTe cell during the substrate CdTe growth, but also from spitting from accumulated deposits in the growth chamber. They have irregular shapes, lateral dimensions ranging from several micrometers to $> 1\,\text{mm}$ and vertical dimensions ranging from $< 1\,\mu\text{m}$ to $> 10\,\mu\text{m}$. Their density can be reduced to $\approx 2\,\text{cm}^{-2}$ by reducing the CdTe growth rate to $\leq 2\,\mu\text{m/h}$ and taking care not to overheat the CdTe source, and has been reduced to $\approx 0.2\,\text{cm}^{-2}$ by EPIR Technologies by taking other measures. However, because of their very large sizes and 3-D nature, even one large flake is extremely deleterious to FPA performance. Large flakes must be completely eliminated for satisfactory FPA fabrication. They are the primary factor preventing the commercial success of HgCdTe grown on Si-based substrates, because the high dislocation density in HgCdTe grown on Si-based substrates is not harmful for most SWIR and MWIR applications.

Hillocks arise from structural defects – microtwinning or 3-D growth – and thus have largely been brought under control by the advent of (211)B growth and improved control of the growth conditions [32.305]. Needle defects were considered in Sect. 32.3.2. They occur only for growth on CdZnTe, and their density has been reduced to the range from mid $10^4\,\text{cm}^{-2}$ to mid $10^5\,\text{cm}^{-2}$ for better substrates and mid $10^6\,\text{cm}^{-2}$ for the worst substrates. The densities of needle defects and threading dislocations have been found to be correlated, but not in a cause-and-effect manner. This correlation makes it difficult to see any effect of the needle defects on electrical properties, and no effect has been seen. They appear not to be a problem, at least not below the mid $10^5\,\text{cm}^{-2}$ range. Cross-hatching is a sign of high-quality growth and is not detrimental to device performance. Surface roughness is an indicator of suboptimal growth, but also is not in itself detrimental to device performance.

32.9 HgTe/CdTe Superlattices

HgTe/CdTe SLs were first proposed in 1979 [32.304]. Their first actual growth was performed in 1982 [32.306]. They possess several potential advantages over HgCdTe alloy material for use in IR detectors, especially for very long wavelengths, in particular:

1. Their bands can be engineered to suppress Auger recombination relative to that in comparable bulk detectors, and SRH lifetimes up to $20\,\mu\text{s}$ have been measured [32.307].
2. The band-to-band tunneling currents are lower than those in comparable bulk detectors due to greater effective masses in the growth-axis direction.
3. Their optical cutoff energies E_c or wavelengths λ_c are sharper than those in comparable bulk detectors for well and barrier widths $\leq 90\,\text{Å}$.
4. By doping only in the CdTe barrier layers, reproducible As activation can be achieved at lower temperatures than in MBE-grown HgCdTe alloys.
5. There is no alloy disorder scattering in HgTe wells such as that in HgCdTe alloys.
6. By doping only in the CdTe barrier layers, one can largely eliminate carrier scattering off dopant atoms. Also, their electrons and holes are in the same layers, resulting in strong optical absorption as compared with that in type II SLs.

However, the usefulness of HgTe/CdTe SLs as a narrow-bandgap optical material relies on the stability of the constituent Hg, Cd, and Te atoms to remain in place across the heterointerfaces of the SL structure. Interdiffusion in HgTe/CdTe SLs has been studied

since shortly after their first growth [32.308–315]. Significant interdiffusion would drastically decrease the cutoff wavelength of a SL and greatly reduce the potential advantages that SLs have over conventional bulk HgCdTe devices. The FTIR spectra and E_c values of HgTe/CdTe SLs change rapidly at even moderate annealing temperatures $T_{an} < 300\,°C$ after growth, with the cutoff energies becoming higher. This has furthered serious concerns about the practicality of SLs for use as absorbers in the LWIR (8–12 μm) and VLWIR (> 15 μm) due to the presumed smearing out and further instability of the HgTe/CdTe interfaces during annealing, device processing, and possibly even after long storage times. On the other hand, we have found that the interdiffusion which takes place during growth and initially thereafter results in a sharpening of the initially rough SL interfaces.

The novel electrical and optical properties of these structures have been reviewed by many authors [32.308, 316–320]. The MBE growth of HgTe/CdTe SLs was initiated at many industrial and other research laboratories in the mid to late 1980s – Rockwell in 1983, Honeywell in 1984, North Carolina State University in 1985, Texas Instruments and Bell Labs in 1986, Hughes Research Laboratory in 1987, and McDonnell Douglas in 1988. Research on HgTe/CdTe SLs peaked in the late 1980s and early 1990s, when MBE was still a relatively new technology and the growth of HgTe-based materials was in its infancy. Because even better control over the growth is required to grow a high-quality HgTe/CdTe SL than to grow high-quality HgCdTe alloy material, for practical technological reasons, the great advantages possessed by HgTe/CdTe SLs in theory were not realized in practice despite rather comprehensive research efforts. Among the factors that hindered the growth of high-quality HgTe/CdTe SLs at that time were the poor design of the Hg cells used, the absence of computer controlled shutters on the MBE system cells, and the unavailability of consistently high-quality substrates. Thus, the interest in HgTe/CdTe SLs peaked well ahead of its time.

32.9.1 Theoretical Properties

Either inverted-band or normal HgTe/CdTe SLs can be grown. With 50 Å CdTe barrier thicknesses d_b, a SL has a normal band structure if the HgTe well thickness d_w is less than 90 Å and an inverted band structure if $d_w > 90$ Å. SLs having $d_w < 70$ Å are preferred because SLs with a direct band structure have much longer Auger lifetimes τ_A. For a p-type 60 Å HgTe/50 Å CdTe SL having a 28.8 μm optical cutoff and $p = 5 \times 10^{15}\,cm^{-3}$, $\tau_A = 7$ ns at 40 K, whereas $\tau_A \ll 1$ ns at 40 K for a HgCdTe alloy having the same optical cutoff and doping level. For the same doping level, a 50 Å HgTe/50 Å CdTe SL having a 15 μm cutoff has an intrinsic lifetime of ≈ 1 μs.

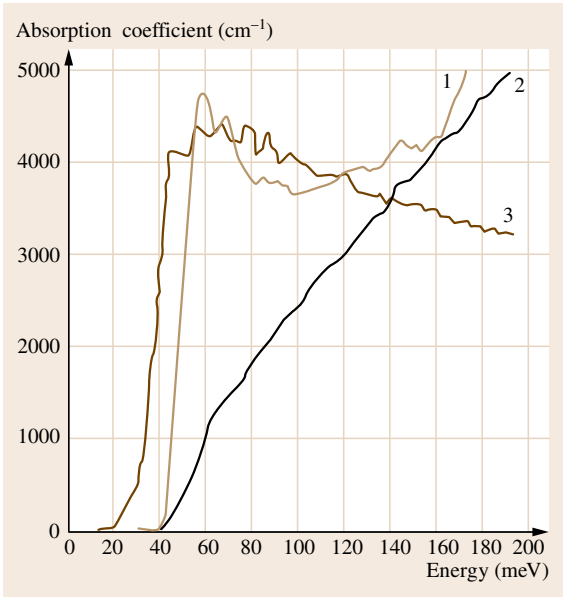


Fig. 32.20 Computed fundamental optical absorption spectra for two SLs (*curve 1* for a normal-band-structure 60 Å HgTe/50 Å $Hg_{0.05}Cd_{0.95}Te$ SL and *curve 3* for an inverted-band-structure 110 Å HgTe/50 Å $Hg_{0.05}Cd_{0.95}Te$ SL) and for bulk HgCdTe alloy (*curve 2*), all with optical bandgaps of approximately 40 meV at 40 K

Another theoretical property of great interest is the sharpness of the optical absorption edge. Here, HgTe/CdTe SLs have a great advantage over HgCdTe alloys. As shown in Fig. 32.20 for a normal SL, an inverted-band SL, and bulk HgCdTe alloy, all having optical cutoffs of ≈ 29 μm (optical bandgaps of ≈ 40 meV), both the normal and the inverted band SLs have much sharper absorption edges than the alloy. Note that interfacial roughness, as long as it is uniform over lateral distances of order 1 μm and occupies no more than ≈ 1/2 of the SL volume, does not significantly affect the sharpness of the SL absorption edges. Even a poorly grown SL will display a sharp absorption edge and other sharp structure from the SL subband structure.

For CdHgTe barrier layers containing ≤ 20% Hg with $d_b \geq 50$ Å, the normal-band SL energy gap and optical cutoff are almost independent of d_b and the Hg

concentration, and depend almost entirely on d_w and on the Cd concentration in the wells, with the dependence on Cd concentration being ≈ 0.9 times as large as that in the HgCdTe alloy with the same cutoff. If one could prevent Cd diffusion into the wells so that they would be pure HgTe, one could control d_w well enough to determine the optical cutoff more accurately in a SL than one could in the alloy by controlling the concentration. However, in reality the combination of interfacial roughness and diffusion make the determination of the optical cutoff less accurate in the SL than in the alloy.

32.9.2 Growth

The growth of HgTe/CdTe offers four challenges not present in the growth of HgCdTe alloy material:

1. The growth of CdTe is optimally performed at $\approx 270\,^{\circ}\mathrm{C}$, $\approx 80\,^{\circ}\mathrm{C}$ higher than the optimal growth temperature of HgTe or Hg-rich HgCdTe
2. The HgTe well widths must be precisely controlled within a fraction of a monolayer
3. The SL interfacial roughness must be minimized and made reproducible to obtain good control of the optical cutoff
4. p-Type doping must be activated as-grown or after only short anneals at temperatures $\leq 300\,^{\circ}\mathrm{C}$.

The substrate preparation and in situ growth monitoring are the same as for the growth of HgCdTe. The growth temperature must be that for HgTe, so the CdTe growth cannot be performed under optimum conditions. Also, the Hg flux must be present at all times, so the barrier layers grown are $Cd_{1-y}Hg_yTe$ with $y \approx 0.05$ rather than $y = 0$. This small amount of Hg in the barriers does not significantly affect the optical cutoff or the band structure.

For (211)B growth the control of the well widths must be achieved by precise control of the growth temperature and the Hg and Te fluxes or by using the virtual interface approximation [32.171,172] and an ellipsometer capable of taking data below the SL optical energy gap. For (211)B growth, which is step-flow growth, one cannot measure the growth of each monolayer by monitoring RHEED oscillations as has been done in the past for (100) growth [32.209]. In step-flow growth the terraces remain unchanged during the growth except for moving laterally; there is no progression from an empty atomic layer to a layer randomly partially occupied to a full layer, and therefore no RHEED oscillations. To have RHEED oscillations one must grow on a surface conducive to the formation of grain boundaries and twinning.

Modulated doping with In in the barrier layers is used for in situ n-type doping, and planar doping with As in the barrier layers, as described in Sect. 32.7.5, is used for in situ p-type doping. The flux shutters are computer controlled and operate on the schedules shown in Fig. 32.21a,b for In and As doping, respectively.

32.9.3 Experimentally Observed Properties

The important observed properties of HgTe/CdTe SLs which do not correspond closely to HgCdTe alloy properties are the macroscopic and microscopic HgTe layer uniformities, the amount of interfacial roughness, and the amount of Cd interdiffused into the HgTe layers during growth. All other observed properties of HgTe/CdTe SLs match expectations based on the observed properties of MBE-grown HgCdTe al-

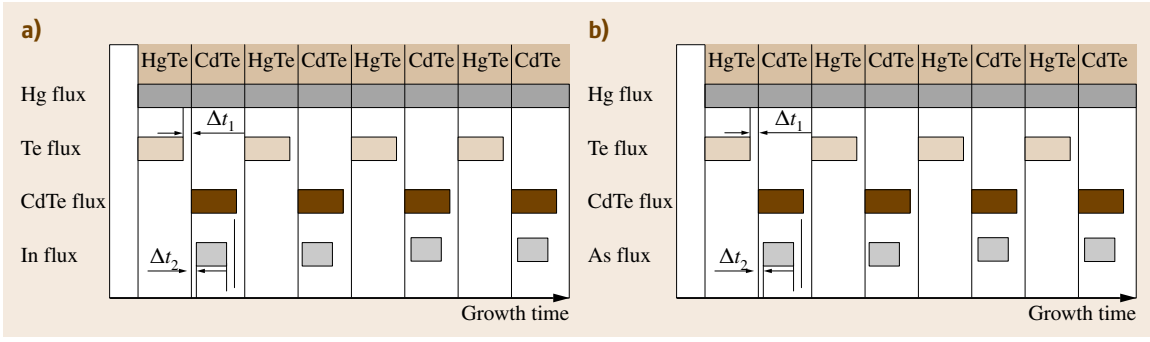


Fig. 32.21a,b MBE shutter sequences for HgTe/CdTe SL growth with: **(a)** In modulation doping and **(b)** As planar doping in the barrier layers

loys. Furthermore, As planar doping of the CdTe layers yields p-activation either as-grown or after only a short low-temperature anneal at $T \leq 300\,^{\circ}\mathrm{C}$.

In situ RHEED and SE measurements indicate high-quality growth, and ex situ XRD measurements yield results indicative of high structural quality. Also, scanning XRD and FTIR measurements have shown the layer uniformities to be excellent. However, both interfacial roughness and Cd interdiffusion during growth pose serious problems. For early (100) growth on $Hg_{0.73}Cd_{0.27}Te/Hg_{0.15}Cd_{0.85}Te$ SLs, it was found by quantitative chemical mapping that the FWHM of the as-grown interfacial roughness is only ≈ 2 monolayers near the surface of an as-grown 0.75 μm-thick SL [32.313]. More recently, we have found by TEM that for (211)B growth the FWHM of the as-grown interfacial roughness can be as large as ≈ 7 monolayers (ML) at the surface of a 10 μm-thick SL and ≈ 5.5 ML at the substrate, with no evidence during growth of poor growth morphology. Much of this observed roughness arose from the step-flow nature of (211)B growth and the existence of (100) steps several monolayers high between successive (111) terraces.

In agreement with earlier calculations [32.311,314], our TEM results show that during growth and during subsequent annealing the CdTe layers become narrower and the HgTe layers wider. Cd atoms diffuse into the HgTe layers from the interfacial roughness region and later, more slowly, from the edges of the CdTe layers, with very little diffusion into the CdTe layers. We have shown that interdiffusion becomes unimportant once the Cd in the alloyed interfacial roughness layers diffuses throughout the HgTe well layers. This is because the diffusion coefficient for Cd in HgTe is much higher than that for Hg in CdTe and because the diffusion coefficient for Cd in HgCdTe increases rapidly with increasing Hg content [32.314]. For a SL of thickness ≈ 10 μm, the growth time is sufficiently long, ≈ 10 h, to induce considerable Cd diffusion far from the surface, but there is little diffusion near the surface, which was exposed to the growth temperature only briefly. Thus a SL of thickness ≈ 10 μm displays a greatly broadened optical absorption edge, because the absorption edge moves to a much higher energy near the substrate than that at the surface, due to Cd diffusion during growth, which increases with increasing time exposed to the growth temperature. In order to avoid this problem, one must either:

1. Grow only thin SL epilayers, say < 3 μm thick
2. Anneal away the as-grown interfacial roughness throughout the epilayer.

It is not a practical option to greatly broaden the HgTe wells or to reduce the height or thickness of the CdTe barriers; either of those options would result in the loss of the primary SL advantages.

The problem with growing only thin SL epilayers is that thin layers give less absorption and thus lower quantum efficiencies. However, this could be overcome by stacking layers on top of one another in a device as is done by DRS in their three-color detector. The problem with annealing away the as-grown interfacial roughness throughout the epilayer is that the resultant Cd concentration in the well layers can easily become too high to obtain an absorption edge in the VLWIR. At 40 K a (211)B 60 Å HgTe/50 Å CdTe SL has a 24.57 μm cutoff. Assuming an average Cd concentration of 0.5 in the roughness layers, to preserve a cutoff ≥ 14 μm one must anneal away ≤ 3 ML of interfacial roughness; to preserve a cutoff ≥ 10 μm one must anneal away < 6 ML of interfacial roughness.

To make SLs practical for use in the VLWIR, one must greatly reduce the amount of as-grown interfacial roughness, possibly by growing on a slightly miscut (100) or (111) plane. On the other hand, going back to (100) or (111) growth would reintroduce all of the problems solved by (211)B growth, such as microtwinning, surface hillocks, and 3-D growth. The achievement of well-controlled SL layer thicknesses with little interfacial roughness is essential to the practical use of SLs for VLWIR detection and possibly even for the LWIR.

32.10 Architectures of Advanced IR Detectors

The mechanisms that convert IR radiation to a device's electrical output and properties (current out, capacitance, etc.) are described by device physics. The reader is referred to [32.321] for a complete description of photon detector behavior. Here, we deal with photon detectors in which IR photons are absorbed by the detector material, creating excess carriers that form an output current proportional to the photon intensity.

An IR detector is a multilayer structure. These layers include contact metal, photon absorbing material, substrates, etc. The architecture describes how the layers are assembled to form the desired structure as well

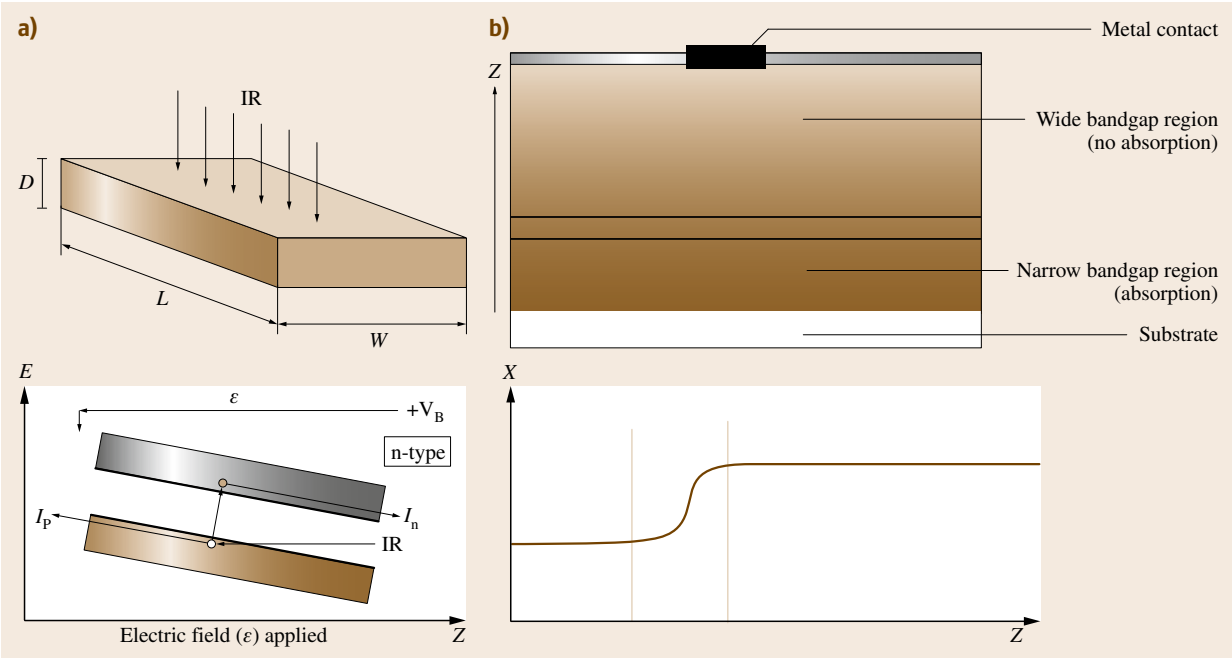


Fig. 32.22a,b Schematic diagrams of **(a)** a photoconductor and **(b)** a graded double-layer heterojunction

as the function of each layer. The two principal IR detectors used today are:

1. The photoconductor (PC), a resistor whose resistance changes under IR illumination, shown in Fig. 32.22a
2. The diode, a graded double-layer heterojunction (DLHJ) shown in Fig. 32.22b, which is a high-impedance minority-carrier device operating under reverse bias.

A detailed description of the performance of these two detector architectures is given in [32.321]. However, these two architectures are not adequate for all IR detector applications and under all external conditions. There are specific applications that require improvement in:

1. The performance at the upper end of the IR spectrum
2. The magnitude of the operating temperature
3. The speed of the detector response
4. The lowering of the detector internal noise.

In summary, there is a need to reduce power consumption, hence lowering cooling costs, to raise the operating temperature while keeping noise at a minimum, to increase the speed of detector response to rapidly changing targets, to increase the number of photoelectrons created per IR photon in and hence to increase the output signal, and/or to increase the detector signal-to-noise ratio. To resolve some of these needs, new detector architectures have been introduced and/or advanced material growth technologies have been exploited.

32.10.1 Reduction of Internal Detector Noise

One of the problems with the graded DLHJ is that optimal performance depends on placing the p–n junction extremely accurately within the graded bandgap region. Optimal p–n junction placement allows the minimization of diode leakage and g-r noise, hence improving the performance. Device physics calculations [32.322] show that the p–n junction must be placed within a few hundred Angstroms of the optimal position; pushing the junction too far into the wider gap region will cause barriers to form, reducing the minority photocurrent. To perform the precise p–n junction placement requires MBE layer growth.

32.10.2 Increasing Detector Response

The response of a detector is determined partly by the quantum efficiency (QE), a measure of the number of

carriers generated per incident photon. For the standard PC or DLHJ, the QE is less than one, which results in an extremely low output current, of the order of nanoamperes. This means that an IR detector requires a preamplifier with a large amplification factor and extremely low noise. This is a tall order, especially if the preamplifier circuit must fit into an area roughly the size of current IR detectors, $\approx 30-50\,\mu m$. It is desirable then to increase the number of current carriers produced per incident IR photon. In order to do so, one must take advantage of a process called impact ionization. Here, an incoming photon creates a free electron with energy well above the conduction-band energy. As the electron moves through the crystal lattice, if the photon-generated free carrier has sufficient energy, when it collides with a bound electron on an atom, that electron becomes a free carrier. At this stage, one photon has created two free carriers. This process will continue until no electron has sufficient energy to free a new electron. To preserve charge neutrality, a free hole is also created. The multiplication process leads to a QE > 1. The probability that a photon-created carrier will create another carrier via impact ionization is called the impact ionization coefficient for that carrier.

The problem with this process is ionization noise. Impact ionization noise is at a maximum if the impact ionization coefficients for electrons and holes are equal. To reduce the noise, one must turn to a more complex device architecture such as the sawtooth variation in concentration shown in Fig. 32.23. Here the teeth created by varying the Cd composition in HgCdTe cause one of the impact ionization coefficients, either for electrons or for holes, to be small. Hence, the ratio of the two coefficients is large. This minimizes the impact ionization noise, resulting in a large multiplication factor (increased response) without increasing the noise, which would defeat the purpose. To create the desired sawtooth structure and maximize multiplication while minimizing impact ionization noise requires precise control of the MBE growth process.

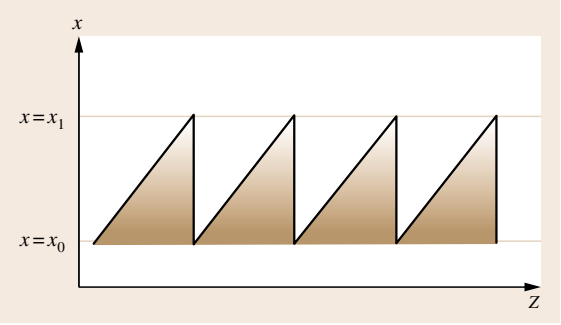


Fig. 32.23 Cd composition in HgCdTe versus distance from substrate for a potential avalanche photodiode (APD) design. This structure was proposed before it could be realized by MBE growth. The sawtooth significantly reduces one carrier's impact ionization coefficient compared with the other, which reduces the impact ionization coefficient

32.10.3 High-Speed IR Detectors

A homojunction detector is a three-layer structure: a p-layer dominated by positively charged hole free carriers; a junction layer in which there is a large built-in electric field, hence the layer is depleted of free carriers; and an n layer dominated by negatively charged electrons. A photon-generated free carrier will diffuse to the edge of the junction and will drift at some rate to the opposite edge. In the absence of an external bias voltage, the junction crossing speed is not high for junction widths of $\approx 0.5\,\mu m$. One can increase the transit speed by applying an external bias whose field is in the same direction as the built-in field. The applied field is a reverse-bias field, which will increase the junction width. So, with the increased width of the region that will support an electric field and with the added external field, the increased total field exerting a force on a carrier will cause the carrier to transit the junction more rapidly.

The transit speed is too low, hence the response time is too long, for many applications such as detecting the wavefront of an electromagnetic wave. To increase the transit speed, one needs to significantly increase the external bias. However, there is a limit to the reverse bias one can apply to the junction before the junction breaks down, avalanches, and behaves like a resistor. To get around this, a three-layer device can be built with the outer layers doped n-type and p-type, respectively. The middle layer is an undoped (intrinsic) layer that contains few free carriers, and hence can support a large electric field. Because the width of the intrinsic layer is much greater than the junction width, say 10–20 times larger, the device can support a very large external voltage, hence a a very large external field without the device breaking down. This type of device is called a p–i–n diode, where i stands for intrinsic. To increase the speed of detector response, the effective width where there are few carriers, and hence which can support a large electric field junction, can be increased by at least a factor of 20 over the p–n junction width. Furthermore, the p- and n-side doping can be made much larger. Hence,

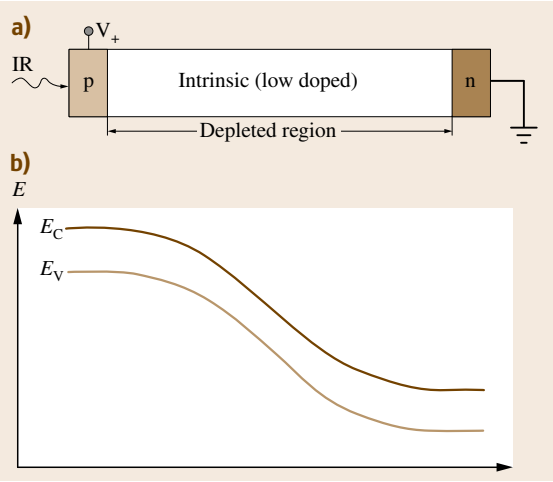


Fig. 32.24a,b Schematic diagram of a rapid-response p–i–n detector. **(a)** Device structure with a wide depleted region with no free carriers present. **(b)** Spatial band structure

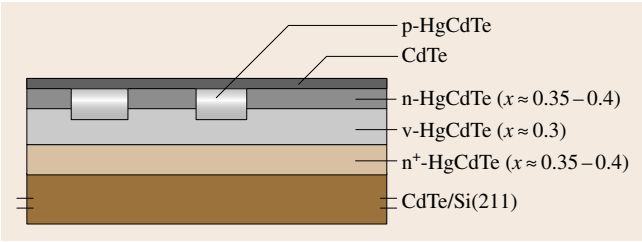


Fig. 32.25 Schematic diagram of a HOT detector architecture used at EPIR Technologies

the electric field in the junction is increased and acts over a longer distance. The carrier when it enters the field of the junction region will feel a very large force that will sweep the carrier to the opposite contact in an extremely short time. As a result, the p–i–n detector could fully respond to IR radiation in a time on the order of a nanosecond or less. The actual structure and the spatial band structure of a p–i–n diode are shown schematically in Fig. 32.24.

32.10.4 High-Operating-Temperature (HOT) IR Detectors

Normally, an IR detector must be cooled so that the near-field IR radiation from its surrounding structure does not interfere with the IR signal coming from a few meters to a few thousand kilometers away. The reason for this is that a body produces IR radiation with a power which increases with increasing temperature T. The near-field radiation is a source of noise because there are mechanisms that generate random numbers of electron–hole pairs, which in turn contribute random fluctuations in the output current and thus noise. Therefore, the higher the operating temperature of the detector, the lower the signal-to-noise ratio (SNR). Depending on the cutoff wavelength, IR photon detectors are operated at anywhere from 4 to 77 K. This lays a heavy burden on the system due to the cooler's weight, power needs, and lifetime.

The ideal, to avoid the cost and weight of cooling, is a detector that yields a good SNR at room temperature, that is, a HOT detector. To achieve this, one must suppress the mechanisms that yield the random (thermal) generation of free carriers. A HOT detector is one that via a specific multilayer architecture suppresses one or more of the principal thermal generation mechanisms. While still in a very early stage of development, these structures promise operation near room temperature or at least at thermoelectrically cooled temperatures. One HOT architecture is shown in Fig. 32.25. Note that the structure requires precise control of layer thickness and doping, and hence is an ideal application of MBE growth technology. Recent references on HOT detectors include [32.323–325].

32.11 IR Focal-Plane Arrays (FPAs)

An IR FPA consists of four layers involving different kinds of materials. The top layer is the detector array, consisting of an array of HgCdTe detectors arranged in a linear or checkerboard pattern. The detector array is grown on a substrate, the second layer, which usually is CdZnTe, GaAs or Si based. Given a sufficiently large substrate and reasonable sized arrays, several detector arrays can be grown on a single substrate. The output of an IR detector is very low, on the order of nA. Therefore, the detector array must be connected to a signal processor, which has a preamplifier for each detector plus various signal processing functions such as multiplexing and analogue-to-digital (A/D) conversion. This readout integrated circuit (ROIC) or multiplexor (MUX) is fabricated from silicon. The difference between the thermal coefficients of expansion of HgCdTe

or CdTe and Si is quite large. Therefore, each detector must be connected to its preamplifier on the ROIC by an In bump. Therefore, a 256×256 detector array requires more than 65 000 separate In bumps. Detector sizes typically are 20 μm on a side to 30 μm on a side, making the In bump cross section extremely small, perhaps $10 \times 10\,\mu m^2$. In is used because it is a good conductor, is ductile, and can accommodate the strain introduced by the mismatched thermal coefficients of expansion. The array of In bump electrical contacts forms the third layer. The Si ROIC is the fourth layer.

IR FPAs designed for astronomical investigations can be as large as 2000×2000. This requires a very precise manufacturing process involving many steps of photolithography. Therefore, one becomes immediately concerned about yield versus cost versus quantity required for a given application.

32.12 Conclusions

The inherent properties of HgCdTe make this alloy the most preferred material for IR detection. During the past several decades HgCdTe alloys have been used as the primary material for IR detectors, which are most important for a variety of military and space applications. Significant advances in the MBE growth of HgCdTe alloys since the early 1980s have offered a new dimension in the ability to fabricate innovative detector structures, control multiple layer heterostructures, and use alternative substrates such as Si for large-format two-dimensional arrays. Today, the MBE of HgCdTe has emerged as a viable technology for research and development and the manufacture of high-performance IR FPAs.

Several factors have contributed to the growth and advancement of HgCdTe MBE technology. The large body of literature published by researchers studying various aspects of HgCdTe deposition experiments and theory has generated much useful information. Various advantages of the MBE technique such as the availability of SE, RHEED, and other in situ analytical characterization techniques to monitor and control the growth have enabled the verification of these experimental data and the related theories. The knowledge base has matured HgCdTe MBE technology in recent years to a point similar to that of III–V compound semiconductor MBE technology in the past. Unlike many III–V semiconductors, it took a while to understand the HgCdTe MBE process because of its stringent thermodynamic constraints. These constraints offer only a narrow window of parameter space for optimum epitaxy. These difficulties are further compounded by temperature changes in the growing HgCdTe layer due to its enhanced IR absorption. In situ tools, particularly SE and RHEED, helped to better understand MBE HgCdTe growth. In the past, and to a certain extent even today, LPE has been the preferred approach for HgCdTe growth. Today, however MBE technology with its various advantages and advancement is becoming more attractive and is required for detectors with advanced architectures.

Abrupt p–n junctions and the control of layer thickness were first demonstrated in GaAs III–V semiconductor MBE technology. Recently, the maturation of the HgCdTe MBE process has also enabled fabrication of device structures with various architectures such as p^+/n, n^+/p, p–i–n, etc. Fabrication of these structures in HgCdTe is now possible because of the thin-film process available in MBE wherein HgCdTe layers can be controlled within submicrometer thicknesses. MBE also has facilitated multiple-layer heterostructure growth with various compositions of stacked HgCdTe layers. Unlike LPE, MBE allows the control of multiple heterostructure growth and thus allows for new device designs and optimization. This very feature also makes it suitable for superlattice structures, and adds the ability to implement bandgap engineering in this alloy. Therefore, MBE has been used to demonstrate both ion-implanted and in situ doped p–n junction devices, and various device designs including superlattice and multicolor detectors have been reported.

The low-temperature and ultrahigh-vacuum nature of MBE adds to its many advantages. Paramount among these is the ability to achieve abrupt material junctions with minimum interdiffusion or interface contamination, and the possibility of depositing HgCdTe on foreign substrates such as Si, GaAs or Al_2O_3. Traditionally, bulk CdZnTe has been preferred as a substrate material for HgCdTe epitaxy. In the past, all three major growth techniques, LPE, MBE, and MOCVD, employed bulk CdZnTe substrates because of their close lattice and chemical match with HgCdTe alloys. During the past decade significant advances have occurred in the MBE of CdTe and HgCdTe on Si and

GaAs substrates. This has been possible due to the low temperature at which MBE occurs, where cross-contamination of the substrate material in the HgCdTe layer does not occur. The primary advantage of growing on Si or GaAs is the opportunity to produce very large-area HgCdTe epitaxial layers. Large-area growth is limited in bulk CdZnTe substrates because of the lack of high-quality large substrates and because of the large difference between the thermal coefficients of expansion of HgCdTe or CdTe and that of Si. The use of Si or GaAs offers more dies of IRFPAs per wafer, large-format IRFPAs, ease of device processing, and a low-cost option for manufacturing.

The challenge of HgCdTe MBE technology has been to produce material structures with structural, optical, and electrical characteristics better than or equivalent to the best LPE-grown materials. MBE technology has developed to the point at which MBE HgCdTe layers grown on bulk CdZnTe substrates have characteristics comparable to those of LPE material. The technology has reached the state at which the quality of MBE HgCdTe layers depends mostly on the characteristics of the substrate material. MBE is sensitive to even minute changes in substrate morphology, and tends to replicate those morphological defects in the growing HgCdTe layer. Various techniques have been developed to mitigate this problem so as to improve the crystalline quality of the HgCdTe layer. The use of buffer layers to block the propagation of defects originating in substrates or at the substrate–epilayer interface has proven useful and is being studied. A thin interfacial CdTe layer has proven effective in reducing substrate surface roughness, but has limitations. The most significant improvement is achieved by growing an interfacial HgTe/CdTe superlattice as a buffer layer. These superlattices are more effective at blocking dislocations and can be grown more nearly lattice matched to the HgCdTe epilayer to be grown.

Today the main challenge of HgCdTe MBE technology is to grow very high-crystalline-quality layers on Si. This epitaxial combination (HgCdTe/Si) creates an extreme 19.2% epitaxial mismatch. Thus growing HgCdTe layers with crystalline quality that is equivalent in characteristics to layers grown on bulk CdZnTe substrates is very difficult. Various techniques have been developed to reduce defect densities in HgCdTe layers grown on Si. Examples of these are variations of II–VI buffer layers such as CdTe and CdSeTe, and other techniques such as reduced-area growth, hydrogen passivation of defects, thermal anneal cycling, etc. Significant progress has been made in the MBE of CdTe and CdSeTe on Si, and extended defect densities as low as $2\times10^5\,\mathrm{cm}^{-2}$ have been achieved. However, progress in reducing defects in the HgCdTe layers grown on buffer/Si substrates is lagging. The best MBE HgCdTe layers grown on buffer/Si substrates achieved thus far exhibit defect densities of $2–5\times10^6\,\mathrm{cm}^{-2}$. Clearly, more research and development is necessary to reduce the defect density by at least an order of magnitude.

Another remaining challenge in the MBE of HgCdTe is to be able to reproducibly grow layers with the desired levels of p-type doping stable against diffusion. In order to form sharp, precisely controlled interfaces and p–n junctions and to obtain well-controlled doping levels, it is essential to grow such junctions by MBE with extrinsic n- and p-type doping. Although stable, well-controlled n-type extrinsic doping of as-grown HgCdTe and HgTe/CdTe SLs has been demonstrated by many authors, p-type extrinsic doping of these materials has remained a problem. Unfortunately, As, which has a low diffusion coefficient, can be reproducibly activated in HgCdTe only at annealing temperatures significantly higher than the growth temperature. Annealing at such high temperatures limits many of the advantages of MBE as a growth technique. Even though high-performance HgCdTe-based photovoltaic photodiodes using various p-doping techniques (in situ and implantation) have been demonstrated, interdiffusion across interfaces, dopant diffusion, and the formation of Hg vacancies all become problems at the high annealing temperatures required for reproducible p-type activation of the preferred group V dopants. The absence of a reproducible p-type doping technique which does not require high-temperature annealing remains a major stumbling block to successfully implementing high-yield fabrication of MBE-grown HgCdTe IRFPAs. Much progress is needed in developing a reproducible in situ extrinsic p-type doping process for MBE technology to advance further.

It is worth noting that, despite the various difficulties faced by the HgCdTe MBE technology, the state of IRFPA technology can still be advanced significantly by further maturation of the MBE process. Considering the current military engagements, advancement of IRFPA technology is needed more than ever. MBE has the potential to emerge as a leading technique to develop innovative IR device architectures and other detector array designs that have not been possible in the past. Already, multicolor IRFPAs have been fabricated from HgCdTe grown by MBE. With more research it is potentially possible to use the HgCdTe MBE process to

develop very large-format IRFPAs. The advent of this technology will undoubtedly lead to important applications required to constantly monitor battlefields and borders. The progress of MBE in the area of alternative substrate technology would lead to more robust and low-cost IRFPAs. This will be most desirable in military applications, where more IR capability must be deployed.

References

32.1 A.Y. Cho: Morphology of epitaxial growth of GaAs by a molecular beam method: The observation of surface structures, J. Appl. Phys. **41**, 782–786 (1970)

32.2 S. Sivananthan, M.D. Lange, G. Monfroy, J.P. Faurie: New achievements in $Hg_{1-x}Cd_xTe$ grown by molecular-beam epitaxy, J. Vac. Sci. Technol. B **6**, 788–793 (1987)

32.3 M.A. Kinch: Fundamental physics of infrared detector materials, J. Electron. Mater. **29**, 809–817 (2000)

32.4 J.P. Faurie, A. Million: Molecular beam epitaxy of II–VI compounds $Cd_xHg_{1-x}Te$, J. Cryst. Growth **54**, 582–585 (1982)

32.5 Y.G. Sidorov, S.A. Dvoretsky, M.V. Yakushev, N.N. Mikhailov, V.S. Varavin, V.I. Liberman: Peculiarities of the MBE growth physics and technology of narrow-gap II–VI compounds, Thin Solid Films **306**, 253–265 (1997)

32.6 J.P. Faurie, A. Million, R. Boch, J.L. Tissot: Latest developments in the growth of $Cd_xHg_{1-x}Te$ and CdTe-HgTe superlattices by molecular beam epitaxy, J. Vac. Sci. Technol. A **1**, 1593–1597 (1983)

32.7 J.P. Faurie, M. Boukerche, J. Reno, S. Sivananthan, C. Hsu: Molecular beam epitaxy of alloys and superlattices involving mercury, J. Vac. Sci. Technol. A **3**, 55–59 (1985)

32.8 L. He, J.R. Yang, S.L. Wang, S.P. Guo, M.F. Yu, X.Q. Chen, W.Z. Fang, Y.M. Qiao, Q.Y. Zhang, R.J. Ding, T.L. Xin: A study of MBE growth and thermal annealing of p-type long wavelength HgCdTe, J. Cryst. Growth **175/176**, 677–681 (1997)

32.9 C.R. Abernethy: Compound semiconductor growth by metallorganic molecular beam epitaxy (MOMBE), Mater. Sci. Eng. R **14**, 203–254 (1995)

32.10 A. Parikh, S.D. Pearson, R.N. Bicknell-Tassius, L.H. Zhang, R. Benz, C.J. Summers: Optimization of the structural properties of $Hg_{1-x}Cd_xTe$ ($x = 0.18 - 0.30$) alloys: Growth and modeling, J. Electron. Mater. **26**, 524–528 (1997)

32.11 J.F. Schetzina: Photoassisted MBE growth of II–VI films and superlattices, Appl. Surf. Sci. **80**, 171–185 (1994)

32.12 R.N. Bicknell, N.C. Giles, J.F. Schetzina, C. Hitzman: Controlled substitutional doping of CdTe thin films grown by photoassisted molecular-beam epitaxy, J. Vac. Sci. Technol. A **5**, 3059–3063 (1987)

32.13 J.R. Arthur: Molecular beam epitaxy, Surf. Sci. **500**, 189–217 (2002)

32.14 J.Y. Tsao: *Materials Fundamentals of Molecular Beam Epitaxy* (Academic, Boston 1993)

32.15 A. Madhukar: Far from equilibrium vapour phase growth of lattice matched III–V compound semiconductor interfaces: some basic concepts and Monte-Carlo computer simulations, Surf. Sci. **132**, 344–374 (1983)

32.16 M.A. Herman: Approaches to understanding MBE growth phenomena, Thin Solid Films **267**, 1–14 (1995)

32.17 L.G. Wang, P. Kratzer, M. Scheffler, N. Moll: Formation and stability of self-assembled coherent islands in highly mismatched heteroepitaxy, Phys. Rev. Lett. **82**, 4042–4045 (1999)

32.18 L.G. Wang, P. Kratzer, N. Moll, M. Scheffler: Size, shape, and stability of InAs quantum dots on the GaAs(001) substrate, Phys. Rev. B **62**, 1897–1904 (2000)

32.19 C. Wagner, W. Schottky: Theory of controlled mixed phases, Z. Phys. Chem. B **11**, 163 (1931)

32.20 C. Wagner, W. Schottky: Theorie der geordneten Mischphasen II (Diffusionsvorgänge), Z. Phys. Chem. Bodenstein-Festband, 177 (1931), in German

32.21 C. Wagner: Theory of ordered mixture phases. III. Appearances of irregularity in polar compounds as a basis for ion conduction and electron conduction, Z. Phys. Chem. B **22**, 181 (1933)

32.22 C. Wagner: Errors in the classifications of crystalized polar compounds as basis for electron and ion conduction, Z. Elektrochem. **39**, 543 (1933)

32.23 W. Schottky: The mechanism of ion movement in solid electrolytes, Z. Phys. Chem. B **29**, 335 (1935)

32.24 W. Schottky: Statistics and thermodynamics of disorder states in crystals, especially in small disarranged states, Z. Elektrochem. **45**, 33 (1939)

32.25 F.A. Kröger, H.J. Vink: *Solid State Physics* Vol. 3, ed. by F. Seitz (Academic, New York 1956) p. 307

32.26 E. Bauer: Phänomenologische Theorie der Kristallabscheidung an Oberflächen I, Z. Kristallogr. **110**, 372–394 (1958), in German

32.27 R. Heckingbottom: Thermodynamic aspects of molecular beam epitaxy: high temperature growth in the $GaAs/Ga_{1-x}Al_xAs$ system, J. Vac. Sci. Technol. B **3**, 572–575 (1985)

32.28 R. Heckingbottom, C.J. Todd, G.J. Davies: The Interplay of thermodynamics and kinetics in molecular beam epitaxy (MBE) of doped gallium arsenide, J. Electrochem. Soc. **127**, 444–450 (1980)

32.29 R. Heckingbottom, G.J. Davies: Germanium doping of gallium arsenide grown by molecular beam epitaxy – Some thermodynamic aspects, J. Cryst. Growth **50**, 644–647 (1980)

32.30 R. Heckingbottom, G.J. Davies, K.A. Prior: Growth and doping of gallium arsenide using molecular beam epitaxy (MBE): Thermodynamic and kinetic aspects, Surf. Sci. **132**, 375–389 (1983)

32.31 H. Seki, A. Koukitu: Thermodynamic analysis of molecular beam epitaxy of III–V semiconductors, J. Cryst. Growth **78**, 342–352 (1986)

32.32 F. Turco, J.C. Guillaume, J. Massies: Thermodynamic analysis of the molecular beam epitaxy of AlInAs alloys, J. Cryst. Growth **88**, 282–290 (1988)

32.33 J.Y. Shen, C. Chatillon: Thermodynamic analysis of molecular beam epitaxy of III–V compounds; application to the $Ga_yIn_{1-y}As$ multilayer epitaxy, J. Cryst. Growth **106**, 553–565 (1990)

32.34 S.V. Ivanov, P.S. Kop'ev, N.N. Ledentsov: Thermodynamic analysis of segregation effects in MBE of $A^{III}–B^{V}$ compounds, J. Cryst. Growth **111**, 151–161 (1991)

32.35 J.P. Gailliard: A thermodynamical model of molecular beam epitaxy, application to the growth of II–VI semiconductors, Rev. Phys. Appl. **22**, 457–463 (1987)

32.36 T. Colin, T. Skauli: Applications of thermodynamical modeling in molecular beam epitaxy of $Cd_xHg_{1-x}Te$, J. Electron. Mater. **26**, 688–696 (1997)

32.37 H.R. Vydyanath, F. Aqariden, P.S. Wijewarnasuriya, S. Sivananthan, G. Chambers, L. Becker: Analysis of the variation in the composition as a function of growth parameters in the MBE growth of indium doped $Hg_{1-x}Cd_xTe$, J. Electron. Mater. **27**, 504–506 (1998)

32.38 E.C. Piquette, M. Zandian, D.D. Edwall, J.M. Arias: MBE growth of HgCdTe epilayers with reduced visible defect densities: kinetics considerations and substrate limitations, J. Electron. Mater. **30**, 627–631 (2001)

32.39 Y. Chang, G. Badano, J. Zhao, C.H. Grein, S. Sivananthan, T. Aoki, D.J. Smith: Formation mechanism of crater defects on HgCdTe/CdZnTe (211)B epilayers grown by molecular beam epitaxy, Appl. Phys. Lett. **83**, 4785–4787 (2003)

32.40 J.W. Garland, C.H. Grein, B. Yang, P.S. Wijewarnasuriya, F. Aqariden: Evidence that arsenic is incorporated as As_4 molecules in the molecular beam epitaxial growth of $Hg_{1-x}Cd_xTe$:As, Appl. Phys. Lett. **74**, 1975–1977 (1999)

32.41 S. Sivananthan, X. Chu, J. Reno, J.P. Faurie: Relation between crystallographic orientation and the condensation coefficients of Hg, Cd and Te during molecular beam epitaxial growth of $Hg_{1-x}Cd_xTe$ and CdTe, J. Appl. Phys. **60**, 1359–1363 (1986)

32.42 S. Sivananthan, X. Chu, J.P. Faurie: Dependence of the condensation coefficient of Hg on the orientation and the stability of the Hg–Te bond for the growth of $Hg_{1-x}M_xTe$ (M = Cd, Mn, Zn), J. Vac. Sci. Technol. B **5**, 694–698 (1987)

32.43 J.P. Faurie: Developments and trends in MBE of II–VI Hg-based compounds, J. Cryst. Growth **81**, 483–488 (1987)

32.44 L.A. Almeida, M. Groenert, J.H. Dinan: Influence of substrate orientation on the growth of HgCdTe by molecular beam epitaxy, J. Electron. Mater. **35**, 1214–1218 (2006)

32.45 J.A. Venables: Kinetic studies of nucleation and growth at surfaces, Thin Solid Films **50**, 357–369 (1978)

32.46 R.J. Koestner, H.F. Schaake: Kinetics of molecular beam epitaxial HgCdTe growth, J. Vac. Sci. Technol. A **6**, 2834–2839 (1988)

32.47 M. Kardar, G. Parisi, Y. Zhang: Dynamic scaling of growing interfaces, Phys. Rev. Lett. **56**, 889–892 (1986)

32.48 P.I. Tamborenea, Z.-W. Lai, S. Das Sarma: Molecular beam epitaxial growth: Simulation and continuum theory, Surf. Sci. **267**, 1–4 (1992)

32.49 S. Das Sarma, Z.-W. Lai, P.I. Tamborenea: Crossover effects in models of kinetic growth with surface diffusion, Surf. Sci. Lett. **268**, L311–L318 (1992)

32.50 P. Kratzer, E. Penev, M. Scheffler: First-principles studies of kinetics in epitaxial growth of III–V semiconductors, Appl. Phys. A **75**, 79–88 (2002)

32.51 R. Triboulet, A. Tromson-Carli, D. Lorans, T. Nguyen Duy: Substrate issues for the growth of mercury cadmium telluride, J. Electron. Mater. **22**, 827–834 (1993)

32.52 R. Sporken, Y.P. Chen, S. Sivananthan, M.D. Lange, J.P. Faurie: Current status of direct growth of CdTe and HgCdTe on silicon by molecular-beam epitaxy, J. Vac. Sci. Technol. B **10**, 1405–1409 (1992)

32.53 R. Korenstein, P. Madison, J.P. Hallock: Growth of (111)CdTe on GaAs/Si and Si substrates for HgCdTe epitaxy, J. Vac. Sci. Technol. B **10**, 1370–1375 (1992)

32.54 Y.P. Chen, S. Sivananthan, J.P. Faurie: Structure of CdTe(111)B grown by MBE on misoriented Si(001), J. Electron. Mater. **22**, 951–957 (1993)

32.55 Y.P. Chen, J.P. Faurie, S. Sivananthan, G.C. Hua, N. Otsuka: Suppression of twin formation in CdTe(111)B epilayers grown by molecular beam epitaxy on misoriented Si(001), J. Electron. Mater. **24**, 475–481 (1995)

32.56 L.A. Almeida, Y.P. Chen, J.P. Faurie, S. Sivananthan: D.J. Smith, S.C.Y. Tsen: Growth of high quality CdTe on Si substrates by molecular beam epitaxy, J. Electron. Mater. **25**, 1402–1405 (1996)

32.57 M. Kawano, A. Ajisawa, N. Oda, M. Nagashima, H. Wada: HgCdTe and CdTe($\bar{1}\bar{1}\bar{3}$)B growth on Si(112)5° off by molecular beam epitaxy, Appl. Phys. Lett. **69**, 2876–2879 (1996)

32.58 M.D. Lange, R. Sporken, K.K. Mahavadi, J.P. Faurie, Y. Nakamura, N. Otsuka: Molecular beam epi-

taxy and characterization of CdTe(211) and CdTe(133) films on GaAs(211)B substrates, Appl. Phys. Lett. **58**, 1988–1990 (1991)

32.59 J.P. Faurie, R. Sporken, Y.P. Chen, M.D. Lange, S. Sivananthan: Heteroepitaxy of CdTe on GaAs and silicon substrates, Mater. Sci. Eng. B **16**, 51–56 (1993)

32.60 V.S. Varavin, S.A. Dvoretsky, V.I. Liberman, N.N. Mikhailov, Y.G. Sidorov: The controlled growth of high-quality mercury cadmium telluride, Thin Solid Films **267**, 121–125 (1995)

32.61 T.J. de Lyon, D. Rajaval, S.M. Johnson, C.A. Cockrum: Molecular-beam epitaxial growth of CdTe(112) on Si(112) substrates, Appl. Phys. Lett. **66**, 2119–2121 (1995)

32.62 T. Colin, D. Minsas, S. Gjoen, R. Sizmann, S. Lovold: Influence of surface step density on the growth of mercury cadmium telluride by molecular beam epitaxy, Mater. Res. Soc. Symp. Proc. **340**, 575 (1994)

32.63 F. Aqariden, H.D. Shih, A.M. Turner, D. Chandra, P.K. Liao: Molecular beam epitaxial growth of HgCdTe on CdZnTe(311)B, J. Electron. Mater. **29**, 727–728 (2000)

32.64 G.A. Carini, C. Arnone, A.E. Bolotnikov, G.S. Camarda, R. de Wames, J.H. Dinan, J.K. Markunias, B. Raghothamachar, S. Sivananthan, R. Smith, J. Zhao, Z. Zhong, R.B. James: Material quality characterization of substrates for HgCdTe epitaxy, J. Electron. Mater. **35**, 1495–1502 (2006)

32.65 H. Abad, J. Zhao, G. Badano, Y. Chang, S. Sivananthan: Correlation of pre-growth surface morphology of substrates with the quality of HgCdTe epilayers, Mil. Sens. Symp. (Tucson 2004)

32.66 H.R. Vydyanath, J. Ellsworth, J.J. Kennedy, B. Dean, C.J. Johnson, G.T. Neugebauer, J. Sepich, P.K. Liao: Recipe to minimize Te precipitation in CdTe and (Cd,Zn)Te crystals, J. Vac. Sci. Technol. B **10**, 1476–1484 (1992)

32.67 B. Li, J. Zhu, X. Zhang, J. Chu: Effect of annealing on near-stoichiometric and non-stoichiometric wafers, J. Cryst. Growth **181**, 204–209 (1997)

32.68 S.J.C. Irvine, A. Stafford, M.U. Ahmed: Substrate/layer relationships in II–VIs, J. Cryst. Growth **197**, 616–625 (1999)

32.69 E. Weiss, O. Klin, E. Benory, E. Kedar, Y. Juravel: Substrate quality impact on the carrier concentration of undoped annealed HgCdTe LPE layers, J. Electron. Mater. **30**, 756–761 (2001)

32.70 S. Sen, C.S. Liang, D.R. Rhiger, J.E. Stannard, H.F. Arlinghaus: Reduction of defects and relation to epitaxial HgCdTe quality, J. Electron. Mater. **25**, 1188–1195 (1996)

32.71 J.P. Faurie, S. Sivananthan, P.S. Wijewarnasuriya: Current status of the growth of mercury cadmium telluride by molecular beam epitaxy on (211)B HgCdTe substrates, SPIE Proc. **1735**, 141–150 (1992)

32.72 R. Korenstein, R.J. Olsen, D. Lee, P.K. Liao, C.A. Castro: Copper outdiffusion from substrates and its effect on the properties of metalorganic chemical-vapor deposition-grown HgCdTe, J. Electron. Mater. **24**, 511–514 (1995)

32.73 J.P. Tower, S.P. Tobin, M. Kestigian, P.W. Norton, A.B. Bollong, H.F. Shaake, C.K. Ard: Substrate impurities and their effects on LPE HgCdTe, J. Electron. Mater. **24**, 497–504 (1995)

32.74 J.P. Tower, S.P. Tobin, P.W. Norton, A.B. Bollong, A. Socha, J.H. Tregilgas, C.K. Ard, H.F. Arlinghaus: Trace copper measurements and electrical effects in LPE HgCdTe, J. Electron. Mater. **25**, 1183–1187 (1996)

32.75 S. Sen, D.R. Rhiger, C.R. Curtis, P.R. Norton: Extraction of mobile impurities from CdZnTe, J. Electron. Mater. **29**, 775–780 (2000)

32.76 D.R. Rhiger, J.M. Peterson, R.M. Emerson, E.E. Gordon, S. Sen, Y. Chen, M. Dudley: Investigation of the cross-hatch pattern and localized defects in epitaxial HgCdTe, J. Electron. Mater. **27**, 615–623 (1998)

32.77 M. Martinka, L.A. Almeida, J.D. Benson, J.H. Dinan: Suppression of strain-induced cross-hatch on molecular beam epitaxy (211)B HgCdTe, J. Electron. Mater. **31**, 732–737 (2001)

32.78 J. Zhao, Y. Chang, G. Badano, S. Sivananthan, J. Markunas, S. Lewis, J.H. Dinan, P.S. Wijewarnasuriya, Y. Chen, G. Brill, N. Dhar: Correlation of (211)B substrate surface morphology and HgCdTe(211)B epilayer defects, J. Electron. Mater. **33**, 881–885 (2005)

32.79 J.N. Johnson, L.A. Almeida, M. Martinka, J.D. Benson, J.H. Dinan: Use of electron cyclotron resonance plasmas to prepare (211)B substrates for HgCdTe molecular beam epitaxy, J. Electron. Mater. **28**, 817–820 (1999)

32.80 R. Singh, S. Velicu, J. Crocco, Y. Chang, J. Zhao, L.A. Almeida, J. Markunas, A. Kaleczyc, J.H. Dinan: Molecular beam epitaxy growth of high-quality HgCdTe LWIR layers on polished and repolished substrates, J. Electron. Mater. **34**, 885–890 (2005)

32.81 P. Moravec, P. Höschl, J. Franc, E. Belas, R. Fesh, R. Grill, P. Horodyský, P. Praus: Chemical polishing of substrates fabricated from crystals grown by the vertical-gradient freezing method, J. Electron. Mater. **35**, 1206–1218 (2006)

32.82 Y.S. Wu, C.R. Becker, A. Waag, R.N. Bicknell-Tassius, G. Landwehr: Thermal effects on (100) substrates as studied by x-ray photoelectron spectroscopy and reflection high energy electron diffraction, Appl. Phys. Lett. **60**, 1878–1880 (1992)

32.83 R.H. Sewell, C.A. Musca, J.M. Dell, L. Faraone, B.F. Usher, T. Dieing: High-resolution x-ray diffraction studies of molecular beam epitaxy-grown HgCdTe heterostructures and substrates, J. Electron. Mater. **34**, 795–803 (2005)

32.84 R. Triboulet, A. Durand, P. Gall, J. Bonnafé, J.P. Fillard, S.K. Krawszyk: Qualification by optical means of CdTe substrates, J. Cryst. Growth **117**, 227–232 (1992)

32.85 W.J. Everson, C.K. Ard, J.L. Sepich, B.E. Dean, G.T. Neugebauer, H.F. Shaake: Etch pit characterization of CdTe and CdZnTe substrates for use in mercury cadmium telluride epitaxy, J. Electron. Mater. **24**, 505–510 (1995)

32.86 K. Nakagawa, K. Maeda, S. Takeuchi: Observation of dislocations in cadmium telluride by cathodoluminescence microscopy, Appl. Phys. Lett. **34**, 574–575 (1979)

32.87 M. Kestigian, A.B. Bollong, J.J. Derby, H.L. Glass, K. Harris, H.L. Hettich, P.K. Liao, P. Mitra, P.W. Norton, H. Wadley: Cadmium zinc telluride growth, characterization, and evaluation, J. Electron. Mater. **28**, 726–731 (1999)

32.88 P. Capper, E.S. O'Keefe, C. Maxey, D. Dutton, P. Mackett, C. Butler, I. Gale: Matrix and impurity element distributions in CdHgTe (CMT) and (Cd,Zn)(Te,Se) compounds by chemical analysis, J. Cryst. Growth **161**, 104–118 (1996)

32.89 B. Li, J. Zhu, X. Zhang, J. Chu: Effect of annealing on near-stoichiometric and non-stoichiometric wafers, J. Cryst. Growth **181**, 204–209 (1997)

32.90 S.L. Price, H.L. Hettich, S. Sen, M.C. Currie, D.R. Rhiger, E.O. McLean: Progress in substrate producibility and critical drivers of IRFPA yield originating with substrates, J. Electron. Mater. **27**, 564–572 (1998)

32.91 Y. Chang, J. Zhao, H. Abad, C.H. Grein, S. Sivanathan, T. Aoki, D.J. Smith: Performance and reproducibility enhancement of HgCdTe molecular beam epitaxy growth on substrates using interfacial HgTe/CdTe superlattice layers, Appl. Phys. Lett. **86**, 131924 (2005)

32.92 Y. Chang, C.H. Grein, J. Zhao, S. Sivananthan, C.Z. Wang, T. Aoki, D.J. Smith, P.S. Wijewarnasuriya, V. Nathan: Improve molecular beam epitaxy growth of HgCdTe on (211)B substrates using interfacial layers of HgTe/CdTe superlattices, J. Appl. Phys. **100**, 114316–1–114316–6 (2006)

32.93 R. Sporken, S. Sivananthan, K.K. Mohavadi, G. Monfroy, M. Boukerche, J.P. Faurie: Molecular beam epitaxial growth of CdTe and HgCdTe on Si(100), Appl. Phys. Lett. **55**, 1879–1881 (1989)

32.94 S. Sivananthan, Y.P. Chen, P.S. Wijewarnasuriya, J.P. Faurie, F.T. Smith, P.W. Norton: Properties of $Hg_{1-x}Cd_xTe$ grown on CdZnTe and Si substrates, Inst. Phys. Conf. Ser. **144**, 239–244 (1995)

32.95 S. Velicu, T.S. Lee, C.H. Grein, P. Boieriu, Y.P. Chen, N.K. Dhar, J. Dinan, D. Lianos: Monolithically integrated HgCdTe focal plane arrays, J. Electron. Mater. **34**, 820–831 (2005)

32.96 L.A. Almeida, L. Hirsch, M. Martinka, P.R. Boyd, J.H. Dinan: Improved morphology and crystalline quality of MBE CdZnTe/Si, J. Electron. Mater. **30**, 608–610 (2001)

32.97 Y.P. Chen, G. Brill, N.K. Dhar: MBE growth of CdSeTe/Si composite substrate for long-wavelength IR HgCdTe applications, J. Cryst. Growth **252**, 270–274 (2004)

32.98 Y.P. Chen, G. Brill, E.M. Campo, T. Hierl, J.C.M. Hwang, N.K. Dhar: Molecular beam epitaxial growth of $Cd_{1-y}Zn_ySe_xTe_{1-x}$ on Si(211), J. Electron. Mater. **33**, 498–502 (2004)

32.99 N.K. Dhar, C.E.C. Wood, A. Gray, H.-Y. Wei, L. Salamanca, J.H. Dinan: Heteroepitaxy of CdTe on {211} Si using crystallized amorphous ZnTe templates, J. Vac. Sci. Technol. B **14**, 2366–2370 (1996)

32.100 S. Rujirawat, L.A. Almeida, Y.P. Chen, S. Sivanathan, D.J. Smith: High quality large-area CdTe(211)B on Si(211) grown by molecular beam epitaxy, Appl. Phys. Lett. **71**, 1810–1812 (1998)

32.101 J.M. Peterson, J.A. Franklin, M. Reddy, S.M. Johnson, E. Smith, W.A. Radford, I. Kasai: High-quality large-area MBE HgCdTe/Si, J. Electron. Mater. **35**, 1283–1286 (2006)

32.102 M.F. Vilela, A.A. Buell, M.D. Newton, G.M. Venzor, A.C. Childs, J.M. Peterson, J.J. Franklin, R.E. Bornfreund, W.A. Radford, S.M. Johnson: Growth and control of middle wave infrared (MWIR) $Hg_{(1-x)}Cd_xTe$ on Si by molecular beam epitaxy, J. Electron. Mater. **34**, 898–904 (2005)

32.103 M. Carmody, J.G. Pasko, D. Edwall, R. Bailey, J. Arias, S. Cabelli, J. Bajaj, L.A. Almeida, J.H. Dinan, M. Groenert, A.J. Stoltz, Y. Chen, G. Brill, N.K. Dhar: Molecular beam epitaxy grown long wavelength infrared HgCdTe on Si detector performance, J. Electron. Mater. **34**, 832–838 (2005)

32.104 M. Carmody, J.G. Pasko, D. Edwall, M. Daraselia, L.A. Almeida, J. Molstad, J.H. Dinan, J.K. Markunas, Y. Chen, G. Brill, N.K. Dhar: Long wavelength infrared, molecular beam epitaxy, HgCdTe-on-Si diode performance, J. Electron. Mater. **33**, 531–537 (2004)

32.105 K. Jówikowski, A. Rogalski: Effect of dislocations on performance of LWIR HgCdTe photodiodes, J. Electron. Mater. **29**, 736–741 (2000)

32.106 S.M. Johnson, D.R. Rhiger, J.P. Rosbeck, J.M. Peterson, S.M. Taylor, M.E. Boyd: Effect of dislocations on the electrical and optical properties of long-wavelength infrared HgCdTe photovoltaic detectors, J. Vac. Sci. Technol. B **10**, 1499–1506 (1992)

32.107 T.J. de Lyon, R.D. Rajavel, J.A. Vigil, J.E. Jensen, O.K. Wu, C.A. Cockrum, S.M. Johnson, G.M. Venzor, S.L. Bailey, I. Kasai, W.L. Ahlgren, M.S. Smith: Molecular beam epitaxial growth of HgCdTe infrared focal-plane arrays on Si substrates for midwave infrared applications, J. Electron. Mater. **27**, 550–555 (1998)

32.108 T.J. de Lyon, J.E. Jensen, I. Kasai, G.M. Venzor, K. Kosai, J.B. de Bruin, W.L. Ahlgren: Molecular-beam epitaxial growth and high-temperature

performance of HgCdTe midwave infrared detectors, J. Electron. Mater. **31**, 220–226 (2002)
32.109 S.M. Johnson, A.A. Buell, M.F. Vilela, J.M. Peterson, J.B. Varesi, M.D. Newton, G.M. Venzor, R.E. Bornfreund, W.A. Radford, E.P.G. Smith, J.P. Rosbeck, T.J. de Lyon, J.E. Jensen, V. Nathan: HgCdTe/Si materials for long wavelength infrared detectors, J. Electron. Mater. **33**, 526–530 (2004)
32.110 M. Carmody, J.G. Pasko, D. Edwall, R. Bailey, J. Arias, M. Groenert, L.A. Almeida, J.H. Dinan, Y. Chen, G. Brill, N.K. Dhar: LWIR HgCdTe on Si detector performance and analysis, J. Electron. Mater. **35**, 1417–1422 (2006)
32.111 E.M. Campo, S. Nakahara, T. Hierl, J.C.M. Hwang, Y. Chen, G. Brill, N.K. Dhar, V. Vaithyanathan, D.G. Schlom, X.-M. Fang, J.M. Fastenau: Epitaxial growth of CdTe on Si through perovskite oxide buffers, J. Electron. Mater. **35**, 1219–1223 (2006)
32.112 Y. Liang, H. Li, J. Finder, C. Overgaard, J. Kulik, D. McCready, S. Shutthanandan: Mater. Res. Soc. Symp. (2004) p. 218
32.113 T.D. Golding, O.W. Holland, M.J. Kim, J.H. Dinan, L.A. Almeida, J.M. Arias, J. Bajaj, H.D. Shih, W.P. Kirk: HgCdTe on Si: Present status and novel buffer layer concepts, J. Electron. Mater. **32**, 882–889 (2003)
32.114 X. Zhou, S. Jiang, W.P. Kirk: Molecular beam epitaxy of BeTe on vicinal Si(100) surfaces, J. Cryst. Growth **175/176**, 624–631 (1997)
32.115 S.H. Shin, J.M. Arias, D.D. Edwall, M. Zandian, J.G. Pasko, R.E. DeWames: Dislocation reduction in HgCdTe on GaAs and Si, J. Vac. Sci. Technol. B **10**, 1492–1498 (1992)
32.116 Y. Lo: New approach to grow pseudomorphic structures over the critical thickness, Appl. Phys. Lett. **59**, 2311–2313 (1991)
32.117 Z. Yang, J. Alperin, W.I. Wang, S.S. Iyer, T.S. Kuan, F. Semendy: In situ relaxed $Si_{1-x}Ge_x$ epitaxial layers with low threading dislocation densities grown on compliant Si-on-insulator substrates, J. Vac. Sci. Technol. B **16**, 1489–1491 (1998)
32.118 P.D. Moran, D.M. Hansen, R.J. Matyi, L.J. Mawst, T.F. Kuech: Experimental test for elastic compliance during growth on glass-bonded compliant substrates, Appl. Phys. Lett. **76**, 2541–2543 (2000)
32.119 S.D. Hersee, D. Zubia, R. Bommena, X. Sun, M. Fairchild, S. Zhang, D. Burckel, A. Frauenglass, S.R.J. Brueck: Nanoheteroepitaxy for the integration of highly mismatched semiconductor materials, IEEE J. Quantum Electron. **QE-38**, 1017–1028 (2002)
32.120 S. Nakamura, M. Senoh, S.-I. Nagahama, N. Iwasa, T. Yamada, T. Matsushita, H. Kiyoku, Y. Sugimoto, T. Kozaki, H. Umemoto, M. Sano, K. Chocho: InGaN/GaN/AlGaN-based laser diodes with modulation-doped strained-layer superlattices grown on an epitaxially laterally overgrown GaN substrate, Appl. Phys. Lett. **72**, 211–213 (1998)
32.121 H. Marchand, X.H. Wu, J.P. Ibbetson, P.T. Fini, P. Kozodoy, S. Keller, J.S. Speck, S.P. DenBaars, U.K. Mishra: Microstructure of GaN laterally overgrown by metalorganic chemical vapor deposition, Appl. Phys. Lett. **73**, 747–749 (1998)
32.122 R. Zhang, I. Bhat: Selective growth of CdTe on Si and GaAs substrates using metalorganic vapor phase epitaxy, J. Electron. Mater. **29**, 765–769 (2000)
32.123 I. Bhat, R. Zhang: Anisotropy in selective metalorganic vapor phase epitaxy of CdTe on GaAs and Si substrates, J. Electron. Mater. **35**, 1293–1298 (2006)
32.124 S.C. Lee, K.J. Malloy, L.R. Dawson: Selected growth and associated faceting and lateral overgrowth of GaAs on a nanoscale limited area bounded by a SiO_2 mask in molecular beam epitaxy, J. Appl. Phys. **92**, 6567–6571 (2002)
32.125 R. Bommena, C. Fulk, J. Zhao, T.S. Lee, S. Sivananthan, S.R.J. Brueck, S.D. Hersee: Cadmium telluride growth onpatterned substrates for mercury cadmium telluride infrared detectors, J. Electron. Mater. **34**, 704–709 (2005)
32.126 X.G. Zhang, P. Li, G. Zhao, D.W. Parent, F.C. Jain, J.E. Ayers: Removal of threading dislocations from patterned heteroepitaxial semiconductors by glide to sidewalls, J. Electron. Mater. **27**, 1248–1253 (1998)
32.127 X.G. Zhang, A. Rodriguez, P. Li, F.C. Jain, J.E. Ayers: Patterned heteroepitaxial processing applied to ZnSe and $ZnS_{0.02}Se_{0.98}$ on GaAs(001), J. Appl. Phys. **91**, 3912–3917 (2002)
32.128 Y. Dong, R.M. Feenstra, D.W. Greve, J.C. Moore, M.D. Sievert, A.A. Baski: Effects of hydrogen on the morphology and electrical properties of GaN grown by plasma-assisted molecular-beam epitaxy, Appl. Phys. Lett. **86**, 121914 (2005)
32.129 Y.F. Chen, C.S. Tsai, Y.H. Chang, Y.M. Chang, T.K. Chen, Y.M. Pang: Hydrogen passivation in $Cd_{1-x}Zn_xTe$ studied by photoluminescence, Appl. Phys. Lett. **58**, 493–495 (1991)
32.130 A.P. Jacobs, Q.X. Zhao, M. Willander, T. Baron, N. Magnea: Hydrogen passivation of nitrogen acceptors confined in CdZnTe quantum well structures, J. Appl. Phys. **90**, 2329–2332 (2001)
32.131 H.Y. Lee, T.W. Kang, T.W. Kim: Temperature dependence of the optical properties in p-$Cd_{0.96}Zn_{0.04}Te$ single crystals, J. Mater. Res. **16**, 2196–2199 (2001)
32.132 A.I. Evstigneev, V.F. Kuleshov, G.A. Lubochkova, M.V. Pashkovskii, E.B. Yakimov, N.A. Yarkin: Influence of hydrogen on the concentration of deep-level centers in $Cd_xHg_{1-x}Te$ crystals, Sov. Phys. Semicond. **19**, 562 (1985)
32.133 S.P. Komissarchuk, L.N. Limarenko, E.P. Lopatinskaya: *Narrow Gap Semiconductors and Semimetals* (LVOV, Moscow 1983) p. 126
32.134 H. Jung, H. Lee, C. Kim: Enhancement of the steady state minority carrier lifetime in HgCdTe photodiode using ECR plasma hydrogenation, J. Electron. Mater. **25**, 1266–1269 (1996)

32.135 Y. Kim, T. Kim, D. Redfern, C. Musca, H. Lee, C. Kim: Characteristics of gradually doped LWIR diodes by hydrogenation, J. Electron. Mater. **29**, 859–864 (2000)
32.136 P. Boieriu, C.H. Grein, S. Velicu, J. Garland, C. Fulk, A. Stoltz, W. Mason, L. Bubulac, R. DeWames, J.H. Dinan: Effects of hydrogen on majority carrier transport and minority carrier lifetimes in LWIR HgCdTe on Si, Appl. Phys. Lett. **88**, 62106 (2006)
32.137 Y. Xin, S. Rujirawat, N.D. Browning, R. Sporken, S. Sivananthan, S.J. Pennycook, N.K. Dhar: The effect of As passivation on the molecular beam epitaxial growth of high-quality single-domain CdTe(111)B on Si(111) substrates, Appl. Phys. Lett. **75**, 349–351 (1999)
32.138 P. Sen, S. Ciraci, I.P. Batra, C.H. Grein, S. Sivananthan: Finite temperature studies of Te adsorption on Si(001), Surf. Sci. **519**, 79–89 (2002)
32.139 P. Sen, I.P. Batra, S. Sivananthan, C.H. Grein: Nabir Dhar, S. Ciraci, Electronic structure of Te- and As-covered Si(211), Phys. Rev. B **68**, 045314 (2003)
32.140 M. Jaime-Vazquez, M. Martinka, R.N. Jacobs, M. Groenert: In-situ spectroscopic study of the As and Te on the Si(112) surface for high-quality epitaxial layers, J. Electron. Mater. **35**, 1455–1460 (2006)
32.141 S.D. Chen, L. Lin, X.Z. He, M.J. Ying, R.Q. Wu: High quality HgCdTe epilayers grown on (211)B GaAs by molecular beam epitaxy, J. Cryst. Growth **152**, 261–265 (1995)
32.142 P. Ballet, F. Noël, F. Pottier, S. Plissard, J.P. Zanatta, J. Baylet, O. Gravrand, E. De Borniol, S. Martin, P. Castelain, J.P. Chamonal, A. Million, G. Destefanis: Dual-band infrared detectors made on high-quality HgCdTe epilayers grown by molecular beam epitaxy on or CdTe/Ge substrates, J. Electron. Mater. **33**, 667–672 (2004)
32.143 J.P. Zanatta, G. Badano, P. Ballet, C. Largeron, J. Baylet, O. Gravrand, J. Rothman, P. Castelain, J.P. Chamonal, A. Million, G. Destefanis, S. Mibord, E. Brochier, P. Costa: Molecular beam epitaxy growth of HgCdTe on Ge for third-generation infrared detectors, J. Electron. Mater. **35**, 1231–1236 (2006)
32.144 A.I. D'Souza, J. Bajaj, R.E. de Wames, D.D. Edwall, P.S. Wijewarnasuriya, N. Nayar: MWIR DLPH photodiode performance dependence on substrate material, J. Electron. Mater. **27**, 727–732 (1998)
32.145 A.J. Norieka, R.F.C. Farrow, F.A. Shirland, W.J. Takai, J. Greggi Jr., S. Wood, W.J. Choyke: Characterization of molecular beam epitaxially grown HgCdTe on CdTe and InSb buffer layers, J. Vac. Sci. Technol. A **4**, 2081–2085 (1986)
32.146 G. Brill, S. Velicu, Y. Chen, N.K. Dhar, T.S. Lee, Y. Selamet, S. Sivananthan: MBE growth and device processing of MWIR HgCdTe on large area Si substrates, J. Electron. Mater. **30**, 717–722 (2001)
32.147 S. Sivananthan: Experimental study on the properties of HgCdTe grown by molecular beam epitxy. Ph.D. Thesis (Department of Physics. Univ. Illinois, Chicago 1987) p. 160
32.148 B.V. Shanabrook, J.R. Waterman, J.L. Davis, R.J. Wagner: Large temperature changes induced by molecular-beam epitaxial growth on radiatively heated substrates, Appl. Phys. Lett. **61**, 2338–2340 (1992)
32.149 B.V. Shanabrook, J.R. Waterman, J.L. Davis, R.J. Wagner, D.S. Katzer: Variations in substrate temperature induced by molecular-beam epitaxial growth on radiatively heated substrates, J. Vac. Sci. Technol. B **11**, 994–997 (1993)
32.150 P. Thompson, Y. Li, J.J. Zhou, D.L. Sato, L. Flanders, H.P. Lee: Diffuse reflectance spectroscopy measurement of substrate temperature and temperature transient during molecular beam epitaxy and implications for low-temperature III–V epitaxy, Appl. Phys. Lett. **70**, 1605–1607 (1997)
32.151 L.A. Almeida, N.K. Dhar, M. Martinka, J.H. Dinan: HgCdTe heteroepitaxy on three-inch (112) CdZnTe/Si: Ellipsometric control of substrate temperature, J. Electron. Mater. **29**, 754–759 (2000)
32.152 M. Daraselia, C.H. Grein, R. Rujirawat, B. Yang, S. Sivananthan, F. Aqariden, S. Shih: In-situ monitoring of temperature and alloy composition of $Hg_{1-x}Cd_xTe$ using FTIR spectroscopic techniques, J. Electron. Mater. **28**, 743–748 (1999)
32.153 T.P. Pearsall, S.R. Saban, J. Booth, B.T. Beard Jr., S.R. Johnson: Precision of noninvasive temperature measurement by diffuse reflectance spectroscopy, Rev. Sci. Instrum. **66**, 4977–4980 (1995)
32.154 T.J. de Lyon, J.A. Roth, D.H. Chow: Substrate temperature measurement by absorption-edge spectroscopy during molecular beam epitaxy of narrow-band gap semiconductor films, J. Vac. Sci. Technol. B **15**, 329–336 (1997)
32.155 F.G. Johnson, G.W. Wicks, R.E. Viturro, R. LaForce: Molecular-beam epitaxial growth of arsenide/phosphide heterostructures using valved, solid group V sources, J. Vac. Sci. Technol. B **11**, 823–825 (1993)
32.156 D.D. Edwall, D.B. Young, A.C. Chen, M. Zandian, J.M. Arias, B. Dlugosch, S. Priddy: Initial Evaluation of a valved Te source for MBE growth of HgCdTe, J. Electron. Mater. **28**, 740–742 (1999)
32.157 W.V. McLevige, J.M. Arias, D.D. Edwall, S.L. Johnson: Ellipsometric profiling of HgCdTe heterostructures, J. Vac. Sci. Technol. B **9**, 2483–2486 (1991)
32.158 D.R. Rhiger: Use of ellipsometry to characterize the surface of HgCdTe, J. Electron. Mater. **22**, 887–898 (1993)
32.159 K.K. Svitashev, S.A. Dvoretsky, Y.G. Sidorov, V.A. Shvets, A.S. Mardezhov, I.E. Nis, V.S. Varavin, V. Liberman, V.G. Remesnik: The growth of high-quality MCT films by MBE using in-situ ellipsometry, Cryst. Res. Technol. **29**, 931–937 (1994)

32.160 S.D. Murthy, I.B. Bhat, B. Johs, S. Pittal, P. He: Application of spectroscopic ellipsometry for real-time control of CdTe and HgCdTe growth in an OMCVD system, J. Electron. Mater. **24**, 445–449 (1995)

32.161 S. Pittal, B. Johs, P. He, J.A. Woollam, S.D. Murthy, I.B. Bath: In-situ monitoring and control of MOCVD growth using multiwavelength ellipsometry, Compound Semiconductors, Inst. Phys. Conf. Ser. **141**, 41–44 (1994)

32.162 J.D. Benson, A.B. Cornfeld, M. Martinka, K.M. Singley, Z. Derzko, P.J. Shorten, J.H. Dinan: In-situ spectroscopic ellipsometry of HgCdTe, J. Electron. Mater. **25**, 1406–1410 (1996)

32.163 B. Johs, C. Herzinger, J. Dinan, A. Cornfeld, J.D. Benson, D. Doctor, G. Olson, I. Ferguson, M. Pelcynski, P. Chow, C.H. Kuo, S. Johnson: Real-time monitoring and control of epitaxial semiconductor growth in a production environment by in situ spectroscopic ellipsometry, Thin Solid Films **313**, 490–495 (1998)

32.164 R.M.A. Azzam, N.M. Bashara: *Ellipsometry and Polarized Light* (Elsevier North-Holland, New York 1977)

32.165 J.W. Garland, C. Kim, H. Abad, P.M. Raccah: Determination of accurate critical-point energies, linewidths and line shapes from spectroscopic ellipsometry data, Thin Solid Films **233**, 148–152 (1993)

32.166 D.E. Aspnes, G.P. Shwartz, G.J. Gualtieri, A.A. Studna: Optical properties of gallium arsenide and its electrochemically grown anodic oxide from 1.5 to 6.0 eV, J. Electrochem. Soc. **128**, 590–597 (1981)

32.167 R.W. Collins: Automatic rotating element ellipsometers: calibration, operation, and real-time applications, Rev. Sci. Instrum. **61**, 2029–2062 (1990)

32.168 B. Johs: Regression calibration method for rotating element ellipsometers, Thin Solid Films **234**, 395–398 (1993)

32.169 G.E. Jellison, F.A. Modine: Two-modulator generalized ellipsometry: Theory, Appl. Opt. **36**, 8190–8198 (1997)

32.170 G.E. Jellison: Windows in ellipsometry measurements, Appl. Opt. **38**, 4784–4789 (1999)

32.171 D.E. Aspnes: Minimal-data approaches for determining outer-layer dielectric responses of films from kinetic reflectometric and ellipsometric measurements, J. Opt. Soc. Am. **10**, 974–983 (1993)

32.172 D.E. Aspnes: Optical approaches to determine near-surface compositions during epitaxy, J. Vac. Sci. Technol. A **14**, 960–966 (1996)

32.173 S.F. Nee: Polarization of specular reflection and near-specular scattering by a rough surface, Appl. Opt. **35**, 3570–3582 (1996)

32.174 D.A. Aspnes, J.B. Theeten, F. Hottier: Investigation of effective-medium models of microscopic surface roughness by spectroscopic ellipsometry, Phys. Rev. B **20**, 3292–3302 (1979)

32.175 H. Fujiwara, J. Koh, P.I. Rovira, R.W. Collins: Assessment of effective-medium theories in the analysis of nucleation and microscopic surface roughness evolution for semiconductor thin films, Phys. Rev. B **61**, 10832–10844 (2000)

32.176 J.D. Jackson: *Classical Electrodynamics* (Wiley, New York 1962), Sect. 6

32.177 S.F. Nee: Ellipsometric analysis for surface roughness and texture, Appl. Opt. **27**, 2819–2831 (1988)

32.178 D.E. Aspnes: *Optical Properties of Solids: New Developments*, ed. by O. Seraphin (North-Holland, Amsterdam 1976) p. 799

32.179 D.E. Aspnes, A.A. Studna: Optical detection and minimization of surface overlayers on semiconductors using spectroscopic ellipsometry, SPIE Proc. **276**, 227–232 (1981)

32.180 D.E. Aspnes, A.A. Studna: Chemical etching and cleaning procedures for silicon, germanium and some III–V compound semiconductors, Appl. Phys. Lett. **39**, 316–318 (1981)

32.181 B. Johs, C.M. Herzinger, J.H. Dinan, A. Cornfeld, J.D. Benson: Development of a parametric optical constant model for HgCdTe for control of composition by spectroscopic ellipsometry during MBE growth, Thin Solid Films **313/314**, 137–142 (1998)

32.182 J.D. Phillips, D.D. Edwall, D.L. Lee: Control of very-long-wavelength infrared HgCdTe detector-cutoff wavelength, J. Electron. Mater. **31**, 664–668 (2002)

32.183 L.A. Almeida, M. Thomas, W. Larsen, K. Spariosu, D.D. Edwall, J.D. Benson, W. Mason, A.J. Stolt, J.H. Dinan: Development and fabrication of two-color mid- and short-wavelength infrared simultaneous unipolar multispectral integrated technology focal-plane arrays, J. Electron. Mater. **31**, 669–676 (2002)

32.184 T.J. DeLyon, G.L. Olson, J.A. Roth, J.E. Jensen, A.T. Hunter, M.D. Jack, S.L. Bailey: HgCdTe composition determination using spectroscopic ellipsometry during molecular beam epitaxy growth of near-infrared avalanche photodiode device structures, J. Electron. Mater. **31**, 688–693 (2002)

32.185 J. Phillips, D. Edwall, D. Lee, J. Arias: Growth of HgCdTe for long-wavelength infrared detectors using automated control from spectroscopic ellipsometry measurements, J. Vac. Sci. Technol. B **19**, 1580–1584 (2001)

32.186 D. Edwall, J. Phillips, D. Lee, J. Arias: Composition control of long wavelength MBE HgCdTe using in-situ spectroscopic ellipsometry, J. Electron. Mater. **30**, 643–646 (2001)

32.187 L.A. Almeida, J.H. Dinan: In situ compositional control of advanced HgCdTe-based IR detectors, J. Cryst. Growth **202**, 22–25 (1999)

32.188 L.A. Almeida, J.N. Johnson, J.D. Benson, J.H. Dinan, B. Johs: Automated compositional control of

$Hg_{1-x}Cd_xTe$ during MBE using in situ spectroscopic ellipsometry, J. Electron. Mater. **27**, 500–503 (1998)
32.189 G. Badano, J.W. Garland, S. Sivananthan: Accuracy of the in situ determination of the temperature by ellipsometry before the growth of HgCdTe by MBE, J. Cryst. Growth **251**, 571–575 (2003)
32.190 C.C. Kim, J.W. Garland, H. Abad, P.M. Raccah: Modeling the optical dielectric function of semiconductors – Extension of the critical-point parabolic-band approximation, Phys. Rev. B **45**, 11749–11767 (1992)
32.191 H. Ehrenreich, M.H. Cohen: Self-consistent field approach to the many-electron problem, Phys. Rev. **115**, 786–790 (1959)
32.192 J.W. Garland, H. Abad, M. Vaccaro, P.M. Raccah: Line shape of the optical dielectric function, Appl. Phys. Lett. **52**, 1176–1178 (1988)
32.193 C.C. Kim, S. Sivananthan: Modeling the optical dielectric function of II–VI compound CdTe, J. Appl. Phys. **78**, 4003–4010 (1995)
32.194 C.C. Kim, S. Sivananthan: Optical properties of ZnSe and its modeling, Phys. Rev. B **53**, 1475–1483 (1996)
32.195 C.C. Kim, M. Daraselia, J.W. Garland, S. Sivananthan: Temperature dependence of the optical properties of CdTe, Phys. Rev. B **56**, 4786–4797 (1997)
32.196 C.C. Kim, S. Sivananthan: Temperature dependence of the optical properties of $Hg_{1-x}Cd_xTe$, J. Electron. Mater. **26**, 561–566 (1997)
32.197 C.C. Kim, J.W. Garland, P.M. Raccah: Modeling the optical dielectric function of the aluminum gallium arsenide alloy system $Al_xGa_{1-x}As$, Phys. Rev. B **47**, 1876–1888 (1993)
32.198 O. Castaing, J.T. Benhlal, R. Granger: An attempt to model the dielectric function in II–VI ternary compounds $Hg_{1-x}Zn_xTe$ and $Cd_{1-x}Zn_xTe$, Eur. Phys. J. B **7**, 563–572 (1999)
32.199 G. Badano, P. Ballet, J.P. Zanatta, X. Baudry, A. Million: Ellipsometry of rough CdTe(211)B-Ge(211) surfaces grown by molecular beam epitaxy, J. Opt. Soc. Am. B **23**, 2089–2096 (2006)
32.200 G. Badano, P. Ballet, X. Baudry, J.P. Zanatta, A. Million, J.W. Garland: Molecular beam epitaxy of BeTe on vicinal Si(100) surfaces, J. Cryst. Growth **296**, 129–134 (2006)
32.201 G. Badano, Y. Chang, J.W. Garland, S. Sivananthan: In-situ ellipsometry studies of adsorption of Hg on CdTe(211)B/Si(211) and molecular beam epitaxy growth of HgCdTe(211)B, J. Electron. Mater. **33**, 583–589 (2004)
32.202 G. Badano, Y. Chang, J.W. Garland, S. Sivananthan: Temperature-dependent adsorption of Hg on CdTe(211)B studied by spectroscopic ellipsometry, Appl. Phys. Lett. **83**, 2324–2326 (2003)
32.203 D.G.M. Anderson, R. Barakat: Necessary and sufficient conditions for a Mueller matrix to be derivable from a Jones matrix, J. Opt. Soc. Am. A **11**, 2305–2319 (1994)
32.204 S.F. Nee: Polarization of specular reflection and near-specular scattering by a rough surface, Appl. Opt. **35**, 3570–3582 (1996)
32.205 W. Braun: *Applied RHEED, Reflection High-Energy Electron Diffraction During Crystal Growth*, Springer Tracts in Modern Physics, Vol. 154 (Springer, Berlin 1999)
32.206 J.J. Harris, B.A. Joyce, P.J. Dobson: Oscillations in the surface structure of Sn-doped GaAs during growth by MBE, Surf. Sci. **103**, L90–L96 (1981)
32.207 C.E.C. Wood: RED intensity oscillations during MBE of GaAs, Surf. Sci. **108**, L441–443 (1981)
32.208 J.M. van Hove, C.S. Lent, P.R. Pukite, P.I. Cohen: Damped oscillations in reflection high energy electron diffraction during GaAs MBE, J. Vac. Sci. Technol. B **1**, 741–746 (1983)
32.209 J. Arias, J. Singh: Use of cation-stabilized conditions to improve compatibility of CdTe and HgTe molecular beam epitaxy, Appl. Phys. Lett. **55**, 1561–1563 (1989)
32.210 W. Braun, L. Däweritz, K.H. Ploog: Origin of electron diffraction oscillations during crystal growth, Phys. Rev. Lett. **80**, 4935–4938 (1998)
32.211 M. Itoh, T. Ohno: Probing the submonolayer morphology change in epitaxial growth: A simulation study, Appl. Phys. Lett. **90**, 073111 (2007)
32.212 T.J. de Lyon, S.M. Johnson, C.A. Cockrum, O.K. Wu, W.J. Hamilton, G.S. Kamath: CdZnTe on Si(001) and Si(112): Direct MBE growth for large-area HgCdTe infrared focal-plane array applications, J. Electrochem. Soc. **141**, 2888–2893 (1994)
32.213 W.-S. Wang, I. Bhat: Growth of high quality CdTe and ZnTe on Si substrates using organometallic vapor phase epitaxy, J. Electron. Mater. **24**, 451–455 (1995)
32.214 P.S. Wijewarnasuriya, M. Zandian, D.D. Edwall, W.V. McLevige, C.A. Chen, J.G. Pasko, G. Hildebrandt, A.C. Chen, J.M. Arias, A.I. D'Souza, S. Rujirawat, S. Sivananthan: MBE p-on-n $Hg_{1-x}Cd_xTe$ heterostructure detectors on silicon substrates, J. Electron. Mater. **27**, 546–549 (1998)
32.215 P. Boieriu, G. Brill, Y. Chen, S. Velicu, N.K. Dhar: $Hg_{1-x}Cd_xTe$(112) nucleation on silicon composite substrates, SPIE Proc. **4454**, 60–70 (2001)
32.216 W. Kern, D.A. Puotinen: Cleaning solutions based on hydrogen peroxide for use in silicon semiconductor technology, RCA Review **31**, 187 (1970)
32.217 A. Ishikawa, Y. Shiraki: Low temperature surface cleaning of silicon and its application to silicon MBE, J. Electrochem. Soc. **133**, 666–671 (1986)
32.218 P.J. Taylor, W.A. Jesser, M. Martinka, K.M. Singley, J.H. Dinan, R.T. Lareau, M.C. Wood, W.W. Clark III: Reduced carbon contaminant, low-temperature silicon substrate preparation for *defect-free* homoepitaxy, J. Vac. Sci. Technol. A **17**, 1153–1159 (1999)
32.219 N.K. Dhar, P.R. Boyd, M. Martinka, J.H. Dinan, L.A. Almeida, N. Goldsman: Heteroepitaxy on 3″

(112)Si: Interface, surface, and layer characteristics, J. Electron. Mater. **29**, 748–753 (2000)
32.220 Y. Xin, S. Rujirawat, N.D. Browning, R. Sporken, S. Sivananthan, S.J. Pennycook, N.K. Dhar: The effect of As passivation on the molecular beam epitaxial growth of high-quality single-domain CdTe(111)B on Si(111) substrates, Appl. Phys. Lett. **75**, 349–351 (1999)
32.221 M. Jaime-Vasquez, M. Martinka, R.N. Jacobs, M. Groenert: In situ spectroscopic study of the As and Te on the Si(112) surface for high-quality epitaxial layers, J. Electron. Mater. **35**, 1455–1460 (2006)
32.222 P. Sen, I.P. Batra, S. Sivananthan, C.H. Grein, N.K. Dhar, S. Ciraci: Electronic structure of Te- and As-covered Si(211), Phys. Rev. B **68**, 045314 (2003)
32.223 Y.P. Chen, G. Brill, N.K. Dhar: MBE growth of CdSeTe/Si composite substrate for long-wavelength IR HgCdTe applications, J. Cryst. Growth **252**, 270–274 (2003)
32.224 M. Boukerche, P.S. Wijewarnasuriya, S. Sivananthan, I.K. Sou, Y.J. Kim, K.K. Mahavadi, J.P. Faurie: The doping of mercury cadmium telluride grown by molecular-beam epitaxy, J. Vac. Sci. Technol. A **6**, 2830–2833 (1988)
32.225 P.S. Wijewarnasuriya, M.D. Lange, S. Sivananthan, J.P. Faurie: Carrier recombination in indium-doped HgCdTe(211)B epitaxial layers grown by molecular beam epitaxy, J. Appl. Phys. **75**, 1005–1009 (1994)
32.226 M. Boukerche, P.S. Wijewarnasuriya, J. Reno, I.K. Sou, J.P. Faurie: Electrical properties of molecular beam epitaxy produced mercury cadmium telluride layers doped during growth, J. Vac. Sci. Technol. A **4**, 2072–2076 (1986)
32.227 M.L. Wroge, D.J. Peterman, B.J. Feldman, B.J. Morris, D.J. Leopold, J.G. Broerman: Impurity doping of mercury telluride-cadmium telluride superlattices during growth by molecular-beam epitaxy, J. Vac. Sci. Technol. A **7**, 435–439 (1989)
32.228 C.A. Hoffman, J.R. Meyer, F.J. Bartoli, Y. Lansari, J.W. Cook Jr., J.F. Schetzina: Electron mobilities and quantum Hall effect in modulation-doped mercury telluride/cadmium telluride superlattices, Phys. Rev. B **44**, 8376–8379 (1991)
32.229 K.A. Harris, T.H. Myers, R.W. Yanka, L.M. Mohnkern, N. Otsuka: A high quantum efficiency in situ doped midwavelength infrared p-on-n homojunction superlattice detector grown by photoassisted molecular-beam epitaxy, J. Vac. Sci. Technol. B **9**, 1752–1758 (1991)
32.230 A. Sher, M.A. Berding, M. Van Schilfgaarde, A.B. Chen: HgCdTe status review with emphasis on correlations, native defects and diffusion, Semicond. Sci. Technol. **6**, C59–C70 (1991)
32.231 P. Boieriu, C.H. Grein, S. Velicu, J. Garland, C. Fulk, S. Sivananthan, A. Stoltz, L. Bubulac, J.H. Dinan: Effects of hydrogen on majority carrier transport and minority carrier lifetimes in long wavelength infrared HgCdTe on Si, Appl. Phys. Lett. **88**, 062106 (2006)
32.232 M.A. Berding, A. Sher, A.B. Chen: Mercury cadmium telluride, defect structure overview, Mater. Res. Soc. Symp. Proc. **216**, 3–10 (1991)
32.233 M.A. Berding, A. Sher, M. Van Schilfgaarde: Behavior of p-type dopants in (Hg,Cd)Te, J. Electron. Mater. **26**, 625–628 (1997)
32.234 R. Balcerak: Infrared material requirements for the next generation of systems, Semicond. Sci. Technol. **6**, C1–C5 (1991)
32.235 P.A. Bakhitin, S.A. Dvoretskii, V.S. Varavin, A.P. Korabkin, N.N. Mikhailor, I.V. Sabanina, Y.G. Sidorov: Effect of low-temperature annealing on electrical properties of n-HgCdTe, Semiconductors **38**, 1172–1175 (2004)
32.236 P.S. Wijewarnasuriya, M.D. Lange, S. Sivananthan, J.P. Faurie: Minority carrier lifetime in Indium-doped HgCdTe(211)B epitaxial layers grown by molecular beam epitaxy, J. Electron. Mater. **24**, 545–549 (1995)
32.237 D.L. Polla, R.L. Aggarwal, D.A. Nelson, J.F. Shanley, M.B. Reine: Mercury vacancy related lifetime in mercury cadmium telluride ($Hg_{0.68}Cd_{0.32}Te$) by optical modulation spectroscopy, Appl. Phys. Lett. **43**, 941–943 (1983)
32.238 M.E. d'Souza, M. Boukerche, J.P. Faurie: Minority-carrier lifetime in p-type (111)B mercury cadmium telluride grown by molecular-beam epitaxy, J. Appl. Phys. **68**, 5195–5199 (1990)
32.239 R. Fastow, Y. Nemirovsky: The excess carrier lifetime in vacancy-doped and impurity-doped HgCdTe, J. Vac. Sci. Technol. A **8**, 1245–1250 (1990)
32.240 M. Boukerche, S. Sivananthan, P.S. Wijewarnasuriya, I.K. Sou, J.P. Faurie: Electrical properties of intrinsic p-type shallow levels in HgCdTe grown by molecular-beam epitaxy in the (111)B orientation, J. Vac. Sci. Technol. A **7**, 311–313 (1989)
32.241 M. Zandian, A.C. Chen, D.D. Edwall, J.G. Pasko, J.M. Arias: p-type arsenic doping of $Hg_{1-x}Cd_xTe$ by molecular beam epitaxy, Appl. Phys. Lett. **71**, 2815–2817 (1997)
32.242 M. Zandian, E. Goo: TEM investigation of defects in arsenic doped layers grown in-situ by MBE, J. Electron. Mater. **30**, 623–626 (2001)
32.243 M.A. Berding, A. Sher, M. van Schilfgaarde, A.C. Chen: Model for As activation, J. Electron. Mater. **27**, 605 (1998)
32.244 M.A. Berding, A. Sher: Amphoteric behavior of arsenic in HgCdTe, Appl. Phys. Lett. **74**, 685–687 (1999)
32.245 M.A. Berding, A. Sher: Arsenic incorporation during MBE growth of HgCdTe, J. Electron. Mater. **28**, 799–803 (1999)

32.246 H.R. Vydynath: Amphoteric behavior of group V mass action constants for lattice site transfers, Semiconductors **5**, S231 (1990)
32.247 S. Sivananthan, P.S. Wijewarnasuriya, F. Aqariden, H.R. Vydynath, M. Zandian, D.D. Edawll, J.M. Arias: Mode of arsenic incorporation in HgCdTe grown by MBE, J. Electron. Mater. **26**, 621–624 (1997)
32.248 P. Boieriu, C.H. Grein, H.S. Jung, J.W. Garland, V. Nathan: Arsenic activation in molecular beam epitaxy grown, in situ doped HgCdTe(211), Appl. Phys. Lett. **86**, 212106 (2005)
32.249 P.S. Wijewarnasuriya, S. Sivananthan: Arsenic incorporation in HgCdTe grown by molecular beam epitaxy, Appl. Phys. Lett. **72**, 1694–1696 (1998)
32.250 A.C. Chen, M. Zandian, D.D. Edwall, J.M. Arias, P.S. Wijewarnasuriya, S. Sivananthan, M.A. Berding, A. Sher: MBE Growth and characterization of in situ arsenic doped HgCdTe, J. Electron. Mater. **27**, 595–599 (1998)
32.251 J. Wu, F.F. Xu, Y. Wu, L. Chen, Y.Z. Wang, M.F. Yu, Y.M. Qiao, L. He: As-doping HgCdTe by MBE, SPIE Proc. **5640**, 637–646 (2005)
32.252 L.O. Bubulac, W.E. Tennant, D.S. Lo, D.D. Edwall, J.C. Robinson, J.S. Chen, G. Bostrup: Ion implanted junction formation in mercury cadmium telluride ($Hg_{1-x}Cd_xTe$), J. Vac. Sci. Technol. A **5**, 3166–3170 (1987)
32.253 L.O. Bubulac, D.D. Edwall, C.R. Viswanathan: Dynamics of arsenic diffusion in metalorganic chemical vapor deposited mercury cadmium telluride on gallium arsenide/silicon substrates, AIP Conf. Proc. **235**, 1695–1704 (1991)
32.254 L.O. Bubulac, D.D. Edwall, S.J.C. Irvine, E.R. Gertner, S.H. Shin: P-type doping of double layer mercury cadmium telluride for junction formation, J. Electron. Mater. **24**, 617–624 (1995)
32.255 P.S. Wijewarnasuriya, S.S. Yoo, J.P. Faurie, S. Sivananthan: P-doping with arsenic in (211)B HgCdTe grown by MBE, J. Electron Mater. **25**, 1300–1305 (1996)
32.256 M.C. Chen, L. Colombo, J.A. Dodge, J.H. Tregilgas: The minority-carrier lifetime in doped and undoped p-type $Hg_{0.78}Cd_{0.22}Te$ liquid-phase epitaxy films, J. Electron. Mater. **24**, 539–544 (1995)
32.257 S.H. Shin, J.M. Arias, M. Zandian, J.G. Pasko, L.O. Bubulac, R.E. DeWames: Annealing effect on the p-type carrier concentration in low-temperature processed arsenic-doped HgCdTe, J. Electron. Mater. **22**, 1039–1047 (1993)
32.258 C.A. Merilainen, C.E. Jones: Deep centers in gold-doped mercury cadmium telluride, J. Vac. Sci. Technol. A **1**, 1637–1640 (1983)
32.259 M. Brown, A.F.W. Willoughby: Diffusion of gold and mercury self-diffusion in n-type Bridgman-grown mercury cadmium telluride ($Hg_{1-x}Cd_xTe$) ($x = 0.2$), J. Vac. Sci. Technol. A **1**, 1641–1645 (1983)
32.260 L.O. Bubulac, W.E. Tennant, R.A. Riedel, J. Bajaj, D.D. Edwall: Some aspects of lithium behavior in ion-implanted mercury cadmium telluride, J. Vac. Sci. Technol. A **1**, 1646–1650 (1983)
32.261 P.S. Wijewarnasuriya, I.K. Sou, Y.J. Kim, K.K. Mahavadi, S. Sivananthan, M. Boukerche, J.P. Faurie: Electrical properties of lithium-doped mercury cadmium telluride ($Hg_{1-x}Cd_xTe$)(100) by molecular beam, Appl. Phys. Lett. **51**, 2025–2027 (1987)
32.262 D.J. Peterman, M.L. Wroge, B.J. Morris, D.J. Leopold, J.G. Broerman: p-on-n heterojunctions of mercury cadmium telluride by molecular-beam epitaxy, controlled silver doping and compositional grading, J. Appl. Phys. **63**, 1951–1954 (1988)
32.263 A. Uedono, K. Ozaki, H. Ebe, T. Moriya, S. Tanigawa, K. Yamamoto, Y. Miyamoto: A study of native defects in Ag-doped HgCdTe by positron annihilation, Jpn. J. Appl. Phys. **36**, 6661–6667 (1997)
32.264 N. Tanaka, K. Ozaki, H. Nishino, H. Ebe, Y. Miyamoto: Electrical properties of HgCdTe epilayers doped with silver using an $AgNO_3$ solution, J. Electron. Mater. **27**, 579–582 (1998)
32.265 M. Chu, S. Terterian, P.C. Wang, S. Mesropian, H.K. Gurgenian, D.-S. Pan: Au-doped HgCdTe for infrared detectors and focal plane arrays, SPIE Proc. **4454**, 116–122 (2001)
32.266 Y. Selamet, A. Ciani, C.H. Grein, S. Sivananthan: Extrinsic p-type doping and analysis of HgCdTe grown by molecular beam epitaxy, SPIE Proc. **4795**, 8–16 (2002)
32.267 Y. Selamet, R. Singh, J. Zhao, Y.D. Zhou, S. Sivananthan, N.K. Dhar: Gold diffusion in mercury cadmium telluride grown by molecular beam epitaxy, SPIE Proc. **5209**, 67–74 (2003)
32.268 A.I. D'Souza, M.G. Stapelbroek, E.R. Bryan, J.D. Beck, M.A. Kinch, J.E. Robinson: Au- and Cu-doped HgCdTe HDVIP detectors, SPIE Proc. **5406**, 205–213 (2004)
32.269 A.J. Ciani, S. Ogut, I.P. Batra, S. Sivananthan: Diffusion of gold and native defects in mercury cadmium telluride, J. Electron. Mater. **34**, 868–872 (2005)
32.270 C.H. Grein, J.W. Garland, S. Sivanantuan, P.S. Wijewarnasuriya, M. Fuchs: Arsenic Incorporation in MBE Grown $Hg_{1-x}Cd_xTe$, J. Electron. Mater. **28**, 789–792 (1999)
32.271 H.F. Schaake: Kinetics of activation of group V impurities in $Hg_{1-x}Cd_xTe$ alloys, J. Appl. Phys. **88**, 1765–1770 (2000)
32.272 D. Chandra, H.F. Schaake, M.A. Kinch, F. Aqariden, C.F. Wan, D.F. Weirauch, H.D. Shih: Activation of arsenic as an acceptor in $Hg_{1-x}Cd_xTe$ under equilibrium conditions, J. Electron. Mater. **31**, 715–719 (2002)
32.273 T.S. Lee, J.W. Garland, C.H. Grein, M. Sumstine, A. Jandeska, Y. Selamet, L.S. Sirem: Correlation of arsenic incorporation and its electrical activation in MBE HgCdTe, J. Electron. Mater. **29**, 869 (2000)
32.274 H.R. Vydyanath, L.S. Lichtman, S. Sivananthan, P.S. Wijewarnasuriya, J.P. Faurie: Annealing ex-

perimenta in heavily arsenic-doped (Hg,Cd)Te, J. Electron. Mater. **24**, 625–634 (1995)
32.275 L.O. Bubulac, D.D. Edwall, C.R. Viswanathan: Dynamics of arsenic diffusion in metalorganic chemical vapor deposition HgCdTe on GaAs/Si substrates, J. Vac. Sci. Technol. B **9**, 1695–1704 (1991)
32.276 P. Boieriu, Y. Chen, V. Nathan: Low-temperature activation of As in $Hg_{1-x}Cd_xTe$(211) grown on Si by molecular beam epitaxy, J. Electron. Mater. **31**, 694–698 (2002)
32.277 J.W. Han, S. Hwang, Y. Lansari, R.L. Harper, Z. Yang, N.C. Giles, J.W. Cook, J.F. Schetzina, S. Sen: p-type modulation-doped HgCdTe, Appl. Phys. Lett. **54**, 63–65 (1989)
32.278 R.L. Harper, S. Hwang, N.C. Giles, J.F. Schetzina, D.L. Dreifus, T.H. Myers: Arsenic-doped CdTe epilayers grown by photoassisted molecular beam epitaxy, Appl. Phys. Lett. **54**, 170 (1989)
32.279 J.M. Arias, S.H. Shin, D.E. Cooper, M. Zandian, J.G. Pasko, E.R. Gertner, R.E. DeWames, J. Singh: p-type arsenic doping of cadmium telluride and mercury telluride/cadmium telluride superlattices grown by photoassisted and conventional molecular-beam epitaxy, J. Vac. Sci. Technol. A **8**, 1025–1033 (1990)
32.280 J. Arias, M. Zandian, J.G. Pasko, S.H. Shin, L.O. Bubulac, R.E. DeWames, W.E. Tennant: Molecular-beam epitaxy growth and in situ arsenic doping of p-on-n HgCdTe heterojunctions, J. Appl. Phys. **69**, 2143–2148 (1991)
32.281 O.K. Wu, G.S. Kamath, W.A. Radford, P.R. Bratt, E.A. Patten: Chemical doping of HgCdTe by molecular-beam epitaxy, J. Vac. Sci. Technol. A **8**, 1034–1038 (1990)
32.282 O.K. Wu, D.N. Jamba, G.S. Kamath: Growth and properties of In- and As-doped HgCdTe by MBE, J. Cryst. Growth **127**, 365–370 (1993)
32.283 S. Sivananthan, P.S. Wijewarnasuriya, J.P. Faurie: Recent progress in the doping of MBE HgCdTe, SPIE Proc. **2554**, 55–68 (1995)
32.284 P.S. Wijewarnasuriya, F. Aqariden, C.H. Grein, J.P. Faurie, S. Sivananthan: p-type doping with arsenic in (211)B HgCdTe grown by MBE, J. Cryst. Growth **175**, 647–652 (1997)
32.285 P.S. Wijewarnasuriya, S. Sivananthan: Arsenic incorporation in HgCdTe grown by molecular beam epitaxy, Appl. Phys. Lett. **72**, 1694–1696 (1998)
32.286 F. Aqariden, P.S. Wijewarnasuriya, S. Sivananthan: Arsenic incorporation in HgCdTe grown by molecular beam epitaxy, J. Vac. Sci. Technol. B **16**, 1309–1311 (1998)
32.287 Y. Chang, G. Badano, E. Jiang, J.W. Garland, J. Zhao, C.H. Grein, S. Sivananthan: Composition and thickness distribution of HgCdTe molecular beam epitaxy wafers by infrared microscope mapping, J. Cryst. Growth **277**, 78–84 (2005)
32.288 A.R. Beattie, P.T. Landsberg: Auger effect in semiconductors, Proc. R. Soc. Lond. Ser. A **249**, 16–29 (1959)
32.289 J.S. Blakemore: *Semiconductor Statistics* (Pergamon, New York 1962), Chap. 6
32.290 V.C. Lopes, A.J. Syllaios, M.C. Chen: Minority carrier lifetime in mercury cadmium telluride, Semicond. Sci. Technol. **8**, 824–842 (1993), and ref. cit.
32.291 W. Shockley, W.T. Read: Statistics of the recombinations of holes and electrons, Phys. Rev. **87**, 835–842 (1952)
32.292 R.N. Hall: Electron-hole recombination in germanium, Phys. Rev. **87**, 387 (1952)
32.293 C.H. Swartz, S. Chandril, R.P. Tompkins, N.C. Giles, T.H. Myers, D.D. Edwall, E.C. Piquette, C.S. Kim, I. Vurgaftman, J.R. Meyer: Accurate measurement of composition, carrier concentration, and photoconductive lifetime in $Hg_{1-x}Cd_xTe$ grown by molecular beam epitaxy, J. Electron. Mater. **35**, 1360–1368 (2006)
32.294 Y. Chang, J.W. Garland, S. Sivananthan: Infrared optical characterization of the narrow gap semiconductor HgCdTe. In: *Advanced Materials in Electronics*, ed. by Q. Guo (Research Signpost, Trivandrum 2004) pp. 249–264
32.295 Y. Chang, G. Badano, J. Zhao, Y.D. Zhou, R. Ashokan, C.H. Grein, V. Nathan: Near-bandgap infrared absorption properties of HgCdTe, J. Electron. Mater. **33**, 709–713 (2004)
32.296 F. Urbach: The long-wavelength edge of photographic sensitivity and of the electronic absorption of solids, Phys. Rev. **92**, 1324 (1953)
32.297 C.H. Grein, S. John: Temperature dependence of the Urbach optical absorption edge: A theory of multiple phonon absorption and emission sidebands, Phys. Rev. B **39**, 1140–1151 (1989)
32.298 M. Carmody, D. Lee, M. Zandian, J. Phillips, J. Arias: Threading and misfit dislocation motion in molecular-beam epitaxy-grown HgCdTe epilayers, J. Electron. Mater. **32**, 710–716 (2003)
32.299 T. Aoki, D.J. Smith, Y. Y, J. Zhao, G. Badano, C.H. Grein, S. Sivananthan: Mercury cadmium telluride/tellurium intergrowths in HgCdTe epilayers grown by molecular beam epitaxy, Appl. Phys. Lett. **82**, 2275–2277 (2003)
32.300 I.V. Sabinina, A.K. Gutakovsky, Y.G. Sidorov, A.V. Latyshev: Nature of V-shaped defects in HgCdTe epilayers grown by molecular beam epitaxy, J. Cryst. Growth **274**, 339–346 (2005)
32.301 T. Aoki, Y. Chang, G. Badano, J. Zhao, C.H. Grein, S. Sivananthan, D.J. Smith: Electron microscopy of surface-crater defects on HgCdTe/CdZnTe(211)B epilayers grown by molecular beam epitaxy, J. Electron. Mater. **32**, 703–709 (2003)
32.302 M. Zandian, J.M. Arias, J. Bajaj, J.G. Pasko, L.O. Bubulac, R.E. DeWames: Origin of void defects in $Hg_{1-x}Cd_xTe$ grown by molecular beam epitaxy, J. Electron. Mater. **24**, 1207–1210 (1995)

32.303 J.B. Varesi, A.A. Buell, J.M. Peterson, R.E. Bornfreund, M.F. Vilela, W.A. Radford, S.M. Johnson: Performance of molecular-beam epitaxy-grown midwave infrared HgCdTe detectors on four-inch Si substrates and the impact of defects, J. Electron. Mater. **32**, 661–666 (2003)
32.304 J.N. Schulman, T.C. McGill: The CdTe/HgTe superlattice: Proposal for a new infrared material, Appl. Phys. Lett. **34**, 663–665 (1979)
32.305 D. Chandra, F. Aquariden, J. Frazier, S. Gutzler, T. Orent, H.D. Shih: Isolation and control of voids and void-hillocks during molecular beam epitaxial growth of HgCdTe, J. Electron. Mater. **29**, 887–892 (2000)
32.306 J.P. Faurie, A. Million, J. Piaguet: CdTe-HgTe multilayers grown by molecular beam epitaxy, Appl. Phys. Lett. **41**, 713–715 (1982)
32.307 K.A. Harris, R.W. Yanka, L.M. Mohnkern, A.R. Reisinger, T.H. Myers, Z. Yang, Z. Yu, S. Hwang, J.F. Schetzina: Properties of (211)B HgTe–CdTe superlattices grown by photon assisted molecular-beam epitaxy, J. Vac. Sci. Technol. B **10**, 1574–1581 (1992)
32.308 J.P. Faurie: Growth and properties of HgTe-CdTe and other Hg-based superlattices, IEEE J. Quantum Electron. **QE-22**, 1656–1665 (1986)
32.309 J.P. Faurie, J. Reno, M. Boukerche: II–VI semiconductor compounds: New superlattice systems for the future?, J. Cryst. Growth **72**, 111–116 (1985)
32.310 D.K. Arch, J.L. Staudenmann, J.P. Faurie: Layer intermixing in HgTe-CdTe superlattices, Appl. Phys. Lett. **48**, 1588–1590 (1986)
32.311 K. Zanio: The effect of interdiffusion on the shape of HgTe/CdTe superlattices, J. Vac. Sci. Technol. A **4**, 2106–2109 (1986)
32.312 D.J. Leopold, J.G. Broerman, D.J. Peterman, M.L. Wroge: Effect of annealing on the optical properties of HgTe-CdTe superlattices, Appl. Phys. Lett. **52**, 969–971 (1988)
32.313 Y. Kim, A. Ourmazd, M. Bode, R.D. Feldman: Nonlinear diffusion in multilayered semiconductor systems, Phys. Rev. Lett. **63**, 636–639 (1989)
32.314 S. Holander-Gleixner, H.G. Robinson, C.R. Helms: Simulation of HgTe/CdTe interdiffusion using fundamental point diffusion mechanisms, J. Electron. Mater. **27**, 672–679 (1998)
32.315 Y. Selamet, Y.D. Zhou, J. Zhao, Y. Chang, C.R. Becker, R. Ashokan, C.H. Grein, S. Sivananthan: HgTe/HgCdTe superlattices grown on CdTe/Si by molecular beam epitaxy for infrared detection, J. Electron. Mater. **33**, 503–508 (2004)
32.316 J.P. Faurie, S. Sivananthan, J. Reno: Present status of molecular beam epitaxial growth and properties of HgTe-CdTe superlattices, J. Vac. Sci. Technol. A **4**, 2096–2100 (1986)
32.317 T.C. McGill, G.Y. Wu, S.R. Hetzler: Superlattices: Progress and prospects, J. Vac. Sci. Technol. A **4**, 2091–2095 (1986)
32.318 C.A. Hoffman, J.R. Meyer, F.J. Bartoli, Y. Lansari, J.W. Cook Jr., J.F. Schetzina: Electron mobilities and quantum Hall effect in modulation-doped HgTe-CdTe superlattices, Phys. Rev. B **44**, 8376–8379 (1991)
32.319 J.M. Arias, S.H. Shin, D.E. Cooper, M. Zandian, J.G. Pasko, E.R. Gertner, R.E. DeWames: p-type arsenic doping of CdTe and HgTe/CdTe superlattices grown by photoassisted and conventional molecular-beam epitaxy, J. Vac. Sci. Technol. A **8**, 1025–1033 (1990)
32.320 C.R. Becker, L. He, M.M. Regnet, M.M. Kraus, Y.S. Wu, G. Landwehr, X.F. Zhang, H. Zhang: The growth and structure of short period (001) $Hg_{1-x}Cd_xTe$-HgTe superlattices, J. Appl. Phys. **74**, 2486–2493 (1993)
32.321 A. Rogalski, M. Kimata, V.F. Kocherov, J. Piotrowski, F.F. Sizov, I.I. Taubkin, N. Tubouchi, N.B. Zaletaev: *Infrared Photon Detectors* (SPIE Optical Engineering Press, Bellingham 1995)
32.322 P.R. Bratt, T.N. Casselman: Potential barriers in HgCdTe heterojunctions, J. Vac. Sci. Technol. A **3**, 238–245 (1985)
32.323 M.A. Kinch, F. Aqariden, D. Chandra, P.K. Liao, H.F. Schaake, H.D. Shih: Minority carrier lifetime in p-HgCdTe, J. Electron. Mater. **34**, 880–884 (2005)
32.324 S. Velicu, R. Ashokan, C.H. Grein, S. Sivananthan, P. Boieriu, D. Rafol: High-temperature HgCdTe/CdTe/Si infrared photon detectors by MBE, SPIE Proc. **4454**, 180–187 (2001)
32.325 J.F. Piotrowski, A. Rogalski: *High-Operating-Temperature Infrared Photodetectors* (SPIE Optical Engineering, Bellingham 2007)

33. Metalorganic Vapor-Phase Epitaxy of Diluted Nitrides and Arsenide Quantum Dots

Udo W. Pohl

Metalorganic vapor-phase epitaxy offers the ability for controlled layer deposition down to the monolayer range. Versatile application in a wide range of materials and its upscaling ability has established this growth technique in industrial mass production, particularly in the field of semiconductor devices. A topic of current research is the extension of the well-developed GaAs-based technology to the near-infrared spectral range for optoelectronic applications. The complementary approaches of either employing dilute nitrides quantum wells or quantum dots have recently achieved significant advances in the field of laser diodes. This chapter introduces the basics of metalorganic vapor-phase epitaxy and illustrates current issues in the growth of InGaAsN/GaAs quantum wells and InAs/GaAs quantum dots. Section 33.1 gives a brief introduction to the growth technique, exemplified by the classical GaAs epitaxy. Sections 33.2 and 33.3 address two current topics of GaAs-related MOVPE, which are intensely studied for, e.g., datacom laser applications: Epitaxy of dilute nitrides and InGaAs quantum dots.

33.1 **Principle of MOVPE** 1133
33.1.1 MOVPE Precursors 1133
33.1.2 Growth Process 1135
33.2 **Diluted Nitride InGaAsN Quantum Wells** .. 1137
33.2.1 Nitrogen Precursors 1138
33.2.2 Structural and Electronic Properties of InGaAsN 1139
33.2.3 Dilute Nitride Quantum Well Lasers. 1141
33.3 **InAs/GaAs Quantum Dots** 1142
33.3.1 The Stranski–Krastanow 2-D–3-D Transition..... 1142
33.3.2 MOVPE of InAs Quantum Dots 1144
33.3.3 Quantum Dot Lasers..... 1147
33.4 **Concluding Remarks** 1148
References 1148

33.1 Principle of MOVPE

Metalorganic vapor-phase epitaxy (MOVPE), also termed metalorganic chemical vapor deposition (MOCVD; sometimes O and M in the acronyms are exchanged), is the most frequently applied CVD technique for semiconductor device fabrication. Industrial large-scale reactors presently have the capacity for simultaneous deposition on 50 2 inch wafers, and the majority of advanced semiconductor devices are produced using this technique. Applications of MOVPE are not restricted to semiconductors, but also include oxides, metals, and organic materials. The technique emerged in the 1960s [33.1–4], when epitaxy was dominated by liquid-phase epitaxy and chloride vapor-phase epitaxy, and molecular-beam epitaxy (MBE) did not exist in its present form. Complex sample structures with abrupt interfaces and excellent uniformity may today be fabricated using either MOVPE or MBE, though application of MOVPE is advantageous in realizing graded layers or in As-P alloys and nitride materials.

33.1.1 MOVPE Precursors

A common feature of chemical vapor-phase techniques is the transport of the constituent elements in the gas phase to the vapor–solid interface in the form of volatile molecules. In MOVPE these species consist of metalorganic compounds, and the transport is made by a carrier gas such as hydrogen at typically 100 mbar total pres-

sure. The gaseous species dissociate thermally at the growing surface of the heated substrate, thereby releasing the elements for layer growth. The dissociation at the surface is generally assisted by chemical reactions.

The net reaction for the MOVPE of GaAs using the standard source compounds trimethylgallium and arsine reads

$$\mathrm{Ga(CH_3)_3 + AsH_3 \rightarrow GaAs + 3CH_4 \uparrow} \ . \qquad (33.1)$$

The reaction is actually much more complicated and comprises many successive steps and species in the chemistry of deposition [33.6] such as, e.g., some steps of precursor decomposition

$$\mathrm{Ga(CH_3)_3 \rightarrow Ga(CH_3)_2 + CH_3}$$
$$\mathrm{\rightarrow GaCH_3 + 2CH_3 \rightarrow Ga + 3CH_3} \ .$$

The source compounds employed for MOVPE must meet some basic requirements. Their stability is low to allow decomposition in the process, but still sufficient for long-term storage. Furthermore their volatility should be high, and a liquid state is favorable to provide a steady-state source flow. Most source molecules have the form MR_n, where M denotes the element used for MOVPE, and R are alkyls such as methyl $\mathrm{CH_3}$. By choosing a suitable organic ligand, the bond strength to a given element M can be selected to comply with the requirements of MOVPE for the solid to be grown. The metal–carbon bond strength depends on the electronegativity of the metal M and the size and configuration of the ligand R [33.7]. As a rule of thumb the bond strength decreases as the number of carbons bonded to the central carbon in the alkyl is increased. This trend is also reflected in the dissociation energy of the first carbon–hydrogen bond, given in Table 33.1 [33.8].

The organic radicals R most frequently used for MOVPE precursors are depicted in Fig. 33.1.

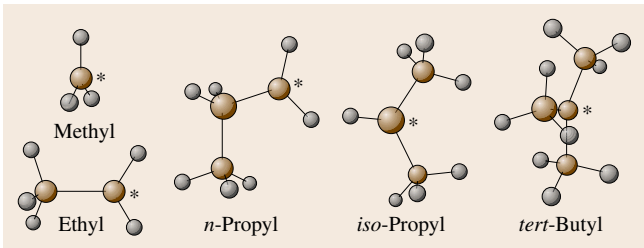


Fig. 33.1 Alkyl radicals used as organic ligands in MOVPE source molecules. *Brown* and *gray spheres* represent carbon and hydrogen atoms, respectively, and the location of a bond to an element M used for epitaxy is indicated by an *asterisk*

Table 33.1 Dissociation energy of the carbon–hydrogen bond for radicals R used in MOVPE source molecules

R	E (kJ/mol)	R	E (kJ/mol)
Methyl	435	*iso*-Propyl	398
Ethyl	410	*tert*-Butyl	381
n-Propyl	410	Allyl	368

Besides metalorganic sources also hydrides such as arsine are employed as precursors. Their use is interesting since they release hydrogen radicals under decomposition that can assist removal of carbon-containing radicals from the surface. A major obstacle is their high toxicity and their very high vapor pressure, requiring extensive safety precautions. To reduce the hazardous potential, hydrides are increasingly replaced by metalorganic alternatives, e.g., arsine by tertiarybutylarsine, where one of the three hydrogen radicals is replaced by a tertiarybutyl radical. Thereby the vapor pressure is strongly reduced, yielding usually liquids at ambient conditions. In addition the toxicity decreases significantly.

Partial pressures for some standard precursors used in the MOVPE of As-related III–V semiconductors are given in Table 33.2. The values are expressed in terms of the parameters a and b to account for the exponential temperature dependence of the vapor pressure according to

$$\log(P_{\mathrm{eq\,MO}}) = a - b/T \ , \qquad (33.2)$$

$P_{\mathrm{eq\,MO}}$ and T given in Torr and K, respectively. Hydrides $\mathrm{AsH_3}$ and $\mathrm{PH_3}$ are stored at 20 °C as liquids under pressures of 11 250 and 26 250 Torr, respectively, and introduced as gases to the MOVPE setup.

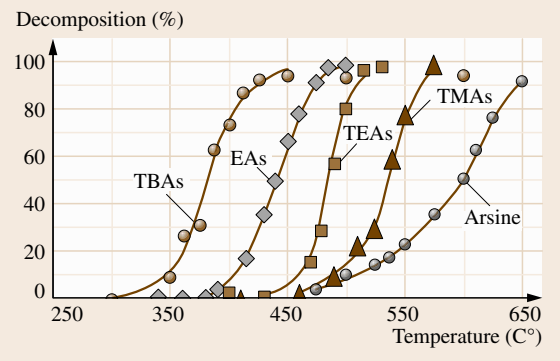


Fig. 33.2 Decomposition of As precursors; the labels TBAs, EAs, TEAs, and TMAs denote *tert*-butyl-As, ethyl-As, triethyl-As, and trimethyl-As, respectively [33.5]

Table 33.2 Equilibrium vapor pressure data of some metalorganic compounds used for III–V MOVPE. (Vapor pressure data taken from data sheets of several precursor suppliers.)

Element	Precursor	Vapor pressure		
		a	b (K)	$P_{\text{eq MO}}$ (Torr) at 20 °C
Al	Trimethylaluminum	8.224	2135	8.7
Ga	Trimethylgallium	8.07	1703	182
	Triethylgallium	8.083	2162	5.1
In	Trimethylindium	10.520	3014	1.7
P	Tertiarybutylphosphine	7.586	1539	142 (10 °C)
As	Tertiarybutylarsine	7.243	1509	81 (10 °C)
Sb	Trimethylantimony	7.708	1697	83
	Triethylantimony	7.904	2183	2.9
N	Dimethylhydrazine	8.646	1921	123

Note: 1 Torr = 1.333 mbar

Precursor molecules may decompose by a number of pyrolytic mechanisms, the most simple being free-radical homolysis, i. e., simple bond cleavage. Since the M–H bond is generally stronger than the M–C bond, metalorganic alternatives of the stable hydrides decompose at lower temperatures – a further incentive for their use. Results of pyrolysis studies for various As precursors, performed in an isothermal flow tube, are given in Fig. 33.2. The bond strength rule of thumb noted above is well reflected in these curves.

33.1.2 Growth Process

Most metalorganic sources are liquids, which are stored in bubblers. For transport to the reactor a carrier gas with a flow Q_{MO} is introduced by a dip tube ending near the bottom. At a fixed temperature the metalorganic liquid forms an equilibrium vapor pressure $P_{\text{eq MO}}$ given by (33.2), and the bubbles saturate with precursor molecules. At the outlet port of the bubbler a pressure controller is installed, which acts like a pressure relief valve and allows to define a fixed pressure P_{bub} ($> P_{\text{eq MO}}$) in the bubbler, thereby decoupling the bubbler pressure from the total pressure P_{tot} in the reactor. The partial pressure of a metalorganic source in the reactor P_{MO} results from the mentioned parameters by

$$P_{\text{MO}} = \frac{Q_{\text{MO}}}{Q_{\text{tot}}} \times \frac{P_{\text{tot}}}{P_{\text{bub}}} \times P_{\text{eq MO}} \,. \tag{33.3}$$

The two fractions in (33.3) are employed to control the partial pressure of the source in the reactor. For sources used as dopants or compounds with very high vapor pressures an additional dilution by mixing with a controlled flux of carrier gas is applied. The gaseous hydrides are directly controlled by their flux Q_{Hyd}, and (33.3) simplifies to

$$P_{\text{Hyd}} = \frac{Q_{\text{Hyd}}}{Q_{\text{tot}}} \times P_{\text{tot}} \,. \tag{33.4}$$

The total flux in the reactor Q_{tot} results from the sum of all component fluxes and the flux of the carrier gas, which is additionally introduced into the reactor by a separate mass flow controller. This flux is generally much higher than that of all sources, and the sum of all source partial pressures P_{MO} and P_{Hyd} is consequently much smaller than the total pressure in the reactor P_{tot}. The reactor pressure P_{tot} is controlled as an independent parameter by a control valve attached to an exhaust pump behind the reactor.

The complete treatment of the MOVPE growth process involves numerous gas-phase and surface reactions, in addition to hydrodynamic aspects. Such complex studies require a numerical approach, and solutions were developed for specific processes such as MOVPE of GaAs from trimethylgallium and arsine [33.6, 9]. We will draw a more general picture of the growth process and outline some relations of growth parameters.

Growth represents a nonequilibrium process. The driving force is given by a drop in the chemical potential μ from the input phase to the solid. For the discussion of the MOVPE process a description by consecutive steps as depicted in Fig. 33.3 is convenient. The reactants in the carrier gas represent the source. Near the solid surface a vertical diffusive transport component originates from reactions of source molecules and incorporation into the growing layer. All processes from adsorption at the surface to the incorporation are sum-

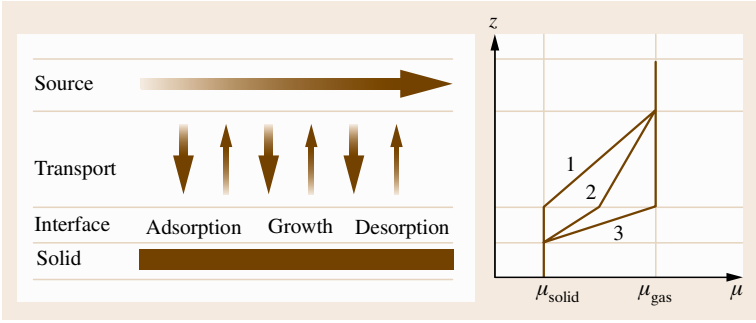


Fig. 33.3 Scheme of the chemical potential μ near the surface of the growing solid during MOVPE. *Path 1* signifies growth controlled by mass transport, *paths 3* and *2* denote growth being limited by interface reactions and the general case, respectively

marized by interface reactions. Finally excess reaction products desorb from the interface by diffusion.

The slowest process of the successive steps limits the growth rate. Without considering mechanisms of growth in detail, processes limited by either transport or kinetics can be well distinguished. Figure 33.4 shows on a logarithmic scale the dependence of the GaAs growth rate on the reciprocal substrate temperature. At low temperature, experiment and simulation show an exponential relation, indicating that thermally activated processes limit the growth rate. Precursor decomposition and interface growth reactions lead to a pronounced temperature dependence, the slope $\propto -\Delta E/(k_B T)$ yields an activation energy ΔE near

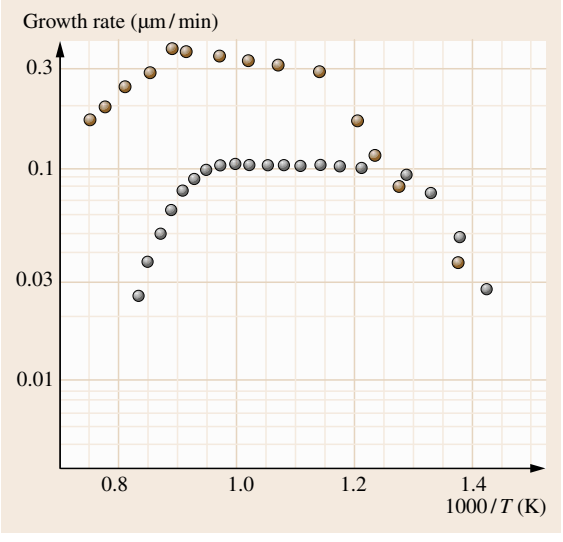


Fig. 33.4 Growth rate of (001)-oriented GaAs as a function of reciprocal temperature. Trimethylgallium and arsine are used as precursors. *Brown* and *gray circles* represent measured data [33.10] and model predictions [33.6], respectively

19 kcal/mole for the given process. This regime is referred to as *kinetically limited growth*. The gas phase supplies precursors to the surface at a rate well exceeding the rate of growth reactions. As the temperature is increased, the growth rate becomes nearly independent of temperature. In this range precursor decomposition and surface reactions are much faster than mass transport from the source to the interface of the growing solid. Since diffusion in the gas phase depends only weakly on temperature, this process is called *transport-limited growth*. Mass transport in this regime depends on the geometry of the reactor, because flow field and temperature profile above the substrate affect cracking and arrival of precursors at the interface; this accounts for the difference in the maximum growth rates in Fig. 33.4. In the high temperature range growth rates decrease due to enhanced desorption and parasitic deposition at the reactor walls, inducing a depletion of the gas phase.

MOVPE is usually performed in the mid-temperature range of transport-limited growth, where variations of the substrate temperature have only a minor effect on growth rate, composition of alloys, and doping. For III–V semiconductors the range is typically 500–800 °C.

Mass transport of the reactants from the gas source to the interface of the growing solid are essentially controlled by diffusion: The mass flow $\boldsymbol{j}_i$ of component i is given by the direct flow component $\boldsymbol{v}$ normal to the interface, the diffusion along the partial pressure gradient $\partial P_i/\partial \boldsymbol{r}$, and the thermodiffusion, according to [33.11]

$$\boldsymbol{j}_i = \frac{P_i \boldsymbol{v}}{k_B T} - \frac{D_i}{k_B T}\left(\frac{\partial}{\partial \boldsymbol{r}} P_i + \frac{\alpha_i}{T} P_i \frac{\partial}{\partial \boldsymbol{r}} T\right), \qquad (33.5)$$

where D_i and α_i are the diffusion constant and the thermodiffusion factor, respectively. The direct flow component normal to the interface should be negligible in a laminar gas flow. Also thermodiffusion is

generally assumed to make no sizeable contribution. Equation (33.5) is hence reduced to the diffusion term.

A simplified one-dimensional model assumes that the partial pressures P_i drop over a so-called diffusion boundary layer of thickness d from their values in the source to values $P_i^{\text{interface}}$ at the interface to the solid [33.7]. Equation (33.5) then reduces to

$$j_i = \frac{D_i}{k_B T\, d}\left(P_i - P_i^{\text{interface}}\right) . \tag{33.6}$$

The factor $D_i/(k_B T\, d)$ may be considered as an effective coefficient of mass transport for component i.

Due to the supersaturation set to induce growth, the partial pressures of the components at the inlet of the reactor P_i are much higher than the near-equilibrium values at the interface to the solid $P_i^{\text{interface}}$. For III–V compounds such as GaAs this means $P_{\text{III}}\, P_{\text{V}} \gg P_{\text{III}}^{\text{interface}}\, P_{\text{V}}^{\text{interface}}$. Furthermore, the group V precursors are far more volatile than the group III species (except for Sb sources); III–V semiconductors are hence usually grown with a large excess of group V species, i. e., $P_{\text{V}}/P_{\text{III}} \gg 1$. These conditions and the requirement of stoichiometric growth lead to the relations of the partial pressures at the interface and the reactor inlet $P_{\text{III}}^{\text{interface}} \ll P_{\text{III}}$, and $P_{\text{V}}^{\text{interface}} \approx P_{\text{V}}$. This means that the growth rate is limited by the flow of group III species, and all group III species arriving at the interface are incorporated into the solid. Equation (33.6) then further reduces to

$$j_{\text{III}} = \frac{D_{\text{III}} P_{\text{III}}}{k_B T\, d} . \tag{33.7}$$

The transport properties expressed by (33.7) may be related to other growth parameters by applying the *boundary-layer model*. Though being oversimplified, this model provides a reasonable description of basic relations. The model assumes a horizontal flow reactor and considers, that the velocity component of the gas flow parallel to the substrate must be zero at the interface due to friction. In a distance δ above the substrate the velocity flow arrives at a constant value v. The range of δ was interpreted in terms of a stagnant layer for the mass transport to the interface δ may be written [33.7]

$$\delta \cong 5\sqrt{\frac{D}{v}} . \tag{33.8}$$

Substituting d in (33.7) by δ and bearing in mind that the diffusion constant D in the gas phase is inversely proportional to the total pressure P_{tot}, yields for the growth rate

$$r = \text{const}\; P_{\text{III}} \sqrt{\frac{v}{P_{\text{tot}}}} . \tag{33.9}$$

The growth rate shows the proportional dependence on the partial pressure of the group III species already expressed by (33.7). According to (33.3), r is hence proportional to the flow of carrier gas through the bubbler Q_{MO} of the group III source. Equation (33.9) predicts a growth rate that is independent of the total reactor pressure, because both P_{III} (cf., (33.3)) and v (inversely proportional to P_{tot}) are functions of P_{tot}. Moreover, the model yields a square-root decrease of the growth rate as the total flow in the reactor Q_{tot} is increased. Deviations from these reasonable predictions are observed at low pressures and low flow velocities, where the boundary-layer thickness is in the range of the reactor height.

The outline of metalorganic vapor-phase epitaxy given above is intended to provide an insight into the basics of the growth technique. Albeit being established in many fields of materials fabrication, new areas of application are steadily developed. Even the classical GaAs-related materials are widely studied to further extend applications, and two current topics are considered in the following sections.

33.2 Diluted Nitride InGaAsN Quantum Wells

The quaternary dilute nitride alloy InGaAsN (or GINA, GaInNAs), which can be grown lattice-matched to GaAs substrates, has recently gained considerable attention as a promising material for laser diodes in the datacom wavelength range. The constituent binaries GaAs and GaN have large differences in electronegativities and lattice constants, leading to an extraordinary large bowing parameter in GaAsN alloys and a strong bandgap decrease for even small N compositions. The same applies for the In-related binaries, cf. Fig. 33.5; in this connection it must be noted that the InN bandgap energy is still controversial, but serious indications exist for a gap near or below 1 eV [33.12–14]. Large band offsets between InGaAsN and GaAs provide the opportunity for good carrier confinement in quantum structures. The attractive quaternary compound is, however, metastable, and introduction of nitrogen is difficult due to a large miscibility gap. Furthermore, intrinsic

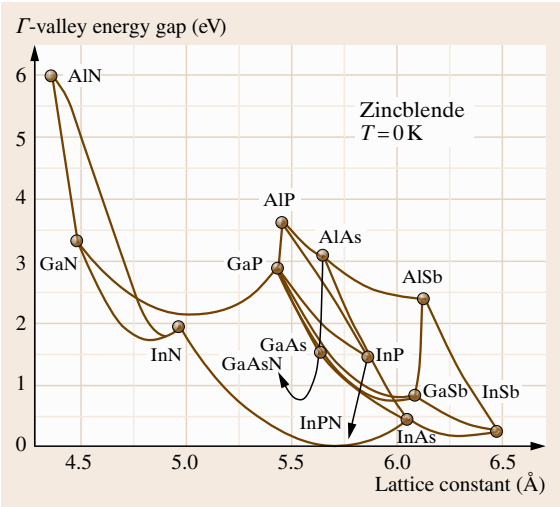


Fig. 33.5 Direct energy gap as a function of lattice constant for zincblende III–V compounds and their alloys; *arrows* indicate limits for predictions with any accuracy [33.15]. For InN more recently a value of 0.78 eV was given [33.13]

regions of high strain and strong localization result from the large differences in lattice parameters and the large energy range induced by statistical composition fluctuations. However, significant advances have been achieved by applying epitaxy at rather low temperatures. This sections addresses the issue of suitable precursors for MOVPE of InGaAsN, the materials properties of epitaxial layers on GaAs substrate, and eventually the current state of laser device applications.

33.2.1 Nitrogen Precursors

The small covalent radius of nitrogen leads to very small solubilities in the conventional III–V compound semiconductors GaAs and InAs, predicted to be only $10^{14}\,\mathrm{cm}^{-3}$ and $10^{17}\,\mathrm{cm}^{-3}$, respectively, at typical growth temperatures [33.17]. To increase the N concentration well above the equilibrium solubility limit, the temperature is lowered to achieve metastable nonequilibrium growth in the range between the concentrations of the spinodal and the stable binodal limit. The latter was determined to be below 2% at 800 K for GaN in GaAs [33.18].

The standard nitrogen source for GaN-related group III nitride semiconductors is ammonia (NH_3). Since dissociation of ammonia in MOVPE requires very high growth temperatures, alternative precursors

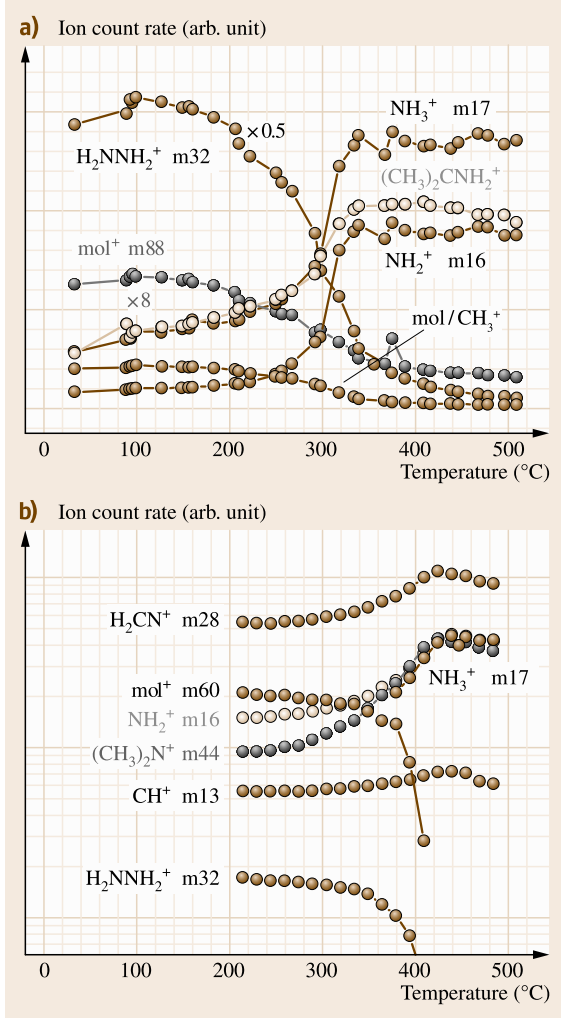


Fig. 33.6a,b Thermal decomposition of the nitrogen sources *tert*-butylhydrazine ((**a**), mass 88) and dimethylhydrazine ((**b**), mass 60), measured by mass spectrometry in an isothermal reaction tube [33.16]

are required for epitaxy of dilute nitride alloys. Hydrazine (H_2NNH_2) is much less stable than ammonia, has a convenient vapor pressure, and provides reactive nitrogen by N–N bond cleavage at temperatures as low as 400 °C. Though low-temperature growth of GaN was proved [33.19], this source is not used due to its toxicity and very high reactivity. Instead, derivatives with one [33.20] or two methyl [33.21, 22] or with tertiarybutyl [33.23, 24] ligands were introduced. Comparative pyrolysis studies of dimethylhydrazine $(CH_3)_2NNH_2$ and tertiarybutylhydrazine $(C_4H_9)(H)NNH_2$ demon-

strated that both sources decompose below 400 °C by N–N bond cleavage, enabling low-temperature nitride growth [33.16]. As shown in Fig. 33.6, decomposition of both precursors produce reactive NH_2 radicals (mass 16) required for the growth process.

Dimethylhydrazine, more precisely (1,1)-dimethylhydrazine with both methyl radicals attached to one nitrogen (also referred to as unsymmetric UDMHy), has developed as the standard source for InGaAsN epitaxy. Pressure data are included in Table 33.2. Tertiarybutylarsine is generally employed as As source due to its favorable pyrolysis properties at low temperature, shown in Fig. 33.2. For the same reason triethylgallium is often used as Ga source instead of trimethylgallium employed for standard GaAs MOVPE, while the generally applied trimethylindium precursor for In supply is sufficiently unstable to be also used for InGaAsN growth.

Lattice-match of InGaAsN to GaAs requires an In/N ratio of about 3. Such a ratio cannot be maintained for a quaternary-alloy QW of 1.3 µm devices due to the low miscibility for N alloying. Instead *strained* QWs with ≈ 30% In and up to 4% N are used. Even then, introduction of N is difficult, since In alloying strongly counteracts N incorporation into the solid. The dependence of the nitrogen content on the In concentration is shown in Fig. 33.7 for two growth temperatures, illustrating the benefit of low-temperature growth. For MOVPE of QWs deposition temperatures as low as 525 °C are used, while thicker layers are grown at slightly increased temperatures (e.g., 550 °C) to achieve a higher growth rate.

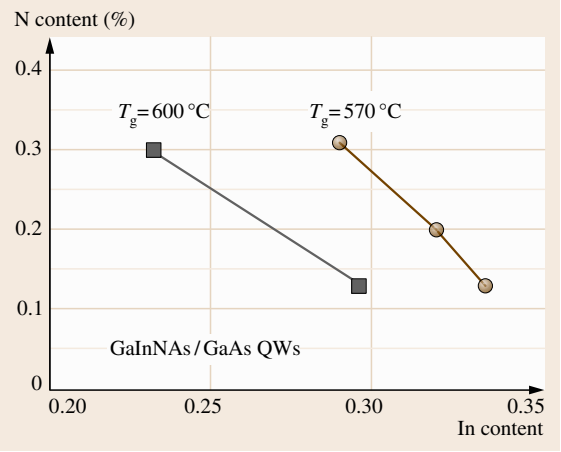


Fig. 33.7 Nitrogen content in InGaAsN/GaAs quantum wells as a function of In content for two deposition temperatures T_g [33.25]

Composition of the quaternary alloy also depends sensitively on other growth parameters. A high tertiarybutylarsine flow was found to favor N incorporation up to a maximum value near 4% [33.26], and higher In concentration was obtained for lower III/V ratios [33.25]. A particular issue of MOVPE is the incorporation of carbon and hydrogen. p-Type character of as-grown material was assigned to C incorporation, and a type conversion upon annealing was attributed to an N–H complex [33.27]. High C levels were ascribed to the strong N–C bond, and growth using various precursors indicated triethylgallium as a possible source [33.28]. Conditions to supply sufficient other radicals to the surface to prevent bonding of N radicals to C-containing groups are suggested to lower C contamination.

33.2.2 Structural and Electronic Properties of InGaAsN

The lattice parameter of GaAsN calculated from total energy minimum yields excellent agreement with Vegard's law despite the large lattice mismatch of more than 20% [33.18]. Due to the good miscibility with In such linear dependence on the composition also applies for the quaternary InGaAsN. Deviations were observed for high N concentrations exceeding 2.9% and were assigned to interstitial, i. e., nonsubstitutional incorporation [33.29]. In molecular-beam epitaxy Sb was used as a surfactant, significantly improving structural and optical properties of highly strained material [33.30–32].

In and also N were found to be randomly distributed on a large scale. While this applies for In also on a microscopic scale, a slightly enhanced number of nitrogen pairs oriented along the [001] growth direction was found in scanning tunneling micrographs of GaAsN [33.33] and InGaAsN layers [33.34, 35]. Calculations based on the pseudopotential method [33.36] and the Keating valence force field model [33.37] show that such N pairs reduce the strain compared with two isolated N atoms. Moreover, adding of further N atoms to a [001]-oriented chain is found to be energetically favorable. The ordering of nitrogen in the quaternary alloy leads to local nanometer-sized strain fields [33.37, 38]. Figure 33.8 shows a transmission electron microscopy (TEM) image taken from an as-grown QW under strain-sensitive imaging conditions. The apparent columnar structure along the [001] growth direction is also found in thick layers. It is not visible if chemically sensitive imaging conditions are used. The

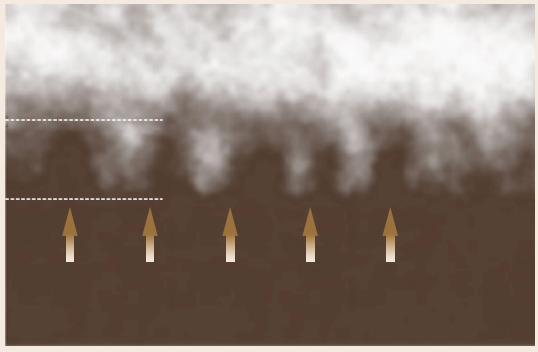


Fig. 33.8 Cross-sectional transmission electron micrograph of an as-grown InGaAsN/GaAs quantum well with 30% In and 2.7% N. *Dashed lines* depict QW boundaries; *arrows* indicate column-like strain fields [33.37]

density of the columns was shown to increase with the N content.

As-grown InGaAsN material with high N content generally shows poor luminescence efficiency, which strongly improves upon annealing. The strain fields visible in TEM images dissolve by annealing. Studies of the nearest-neighbor bonds using x-ray absorption fine structure show that the short-range ordering of the nitrogen chains largely disappears, accompanied by an increase of In–N bonds at the expense of Ga–N bonds [33.39–41]. While N chains build a stable configuration at the growth surface, the highly In-coordinated state of N was calculated to have a lower total energy in the bulk [33.37,42,43]. Strain in the layers does not significantly affect the nearest-neighbor bonding [33.40], but the optimum annealing temperature with respect to photoluminescence (PL) efficiency was found to decrease with strain [33.44]. The origin of the improved luminescence efficiency after anneal was assigned to the removal of nonsubstitutional nitrogen, which builds a complex weakly bonded to Ga and acts as a trap for photoexcited charge carriers [33.29].

Thermal annealing of as-grown InGaAsN induces a blue-shift of the bandgap due to the changed bonding [33.40, 45]. This finding agrees with the described reduction of N pairing and predictions of total energy calculations, which yield a reduced bandgap in case of clustering and also a reduced momentum matrix element (i. e., PL intensity) [33.36]. It was pointed out that the band structure of diluted nitride alloys shows an alloying mechanism different from that of conventional semiconductors [33.46]. Addition of nitrogen induces localized cluster states below the conduction band edge, which are gradually overtaken by states composed of both localized and delocalized states. These perturbed host states form the alloy conduction band edge and lead to the strong, composition-dependent bowing parameter. The effect of a few percent of nitrogen on the valence band is small and neglected in the generally accepted parameterized band anticrossing (BAC) model [33.47, 48]. The splitting into a lower-lying E_- and an E_+ band can be considered as the interaction between the spatially localized N level and the delocalized conduction band states of the host semiconductor, and the respective dependence on the nitrogen fraction x reads

$$E_\pm = \frac{1}{2}\left((E_\mathrm{C}+E_\mathrm{N}) \pm \sqrt{(E_\mathrm{C}-E_\mathrm{N})^2+4V^2x}\right), \tag{33.10}$$

with E_C being the conduction band edge of the unperturbed host semiconductor, E_N the position of the N impurity level in this host, and V the interaction potential between the two bands. Respective model parameters for diluted nitride alloys are given in Table 33.3.

The offsets of valence and conduction bands are basic parameters for fabricating heterostructure devices. Data of InGaAsN/GaAs QW samples were particularly obtained from photoreflectance spectra by evaluating the splitting between the heavy-hole and light-hole ground state [33.49], and from photovoltage spectra by directly measuring the transition energy between confined and free states [33.50]. A type I band alignment is found for electron and heavy hole, with values for $Q_\mathrm{C} = \Delta E_\mathrm{C}/(\Delta E_\mathrm{C}+\Delta E_\mathrm{V})$ of 55–70%. The light-hole valence band offset of a tensile-strained GaAsN QW

Table 33.3 Band anticrossing model parameters for dilute nitride semiconductors; the energy of the isoelectronic nitrogen impurity E_N is given with respect to the valence band maximum [33.15]

Alloy	E_N (eV)	V (eV)
GaAsN	1.65	2.7
InAsN	1.44	2.0
$In_xGa_{1-x}AsN$	$1.65(1-x)+1.44x-0.38x(1-x)$	$2.7(1-x)+2.0x-3.5x(1-x)$

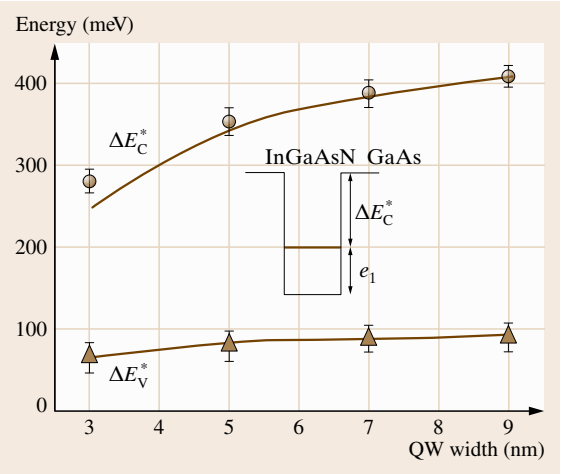


Fig. 33.9 Energy differences ΔE^* between confined states and band edges of $In_{0.3}Ga_{0.7}As_{0.98}N_{0.02}$ quantum wells with various widths, measured using photovoltage spectra; the *continuous lines* represent simulated values [33.50]

with 2% N was reported to be type II, and the counteracting strain introduced by 15% In was expected not to convert the type of alignment [33.49]. Introduction of Sb leads essentially to an increase of the hole QW depth [33.51]. Figure 33.9 shows measured energy differences between confined states and band edges for QWs with 2% N and 30% In. These values are of practical relevance and related to the band offsets by $\Delta E_C = \Delta E_C^* + e_1$, and $\Delta E_V = \Delta E_V^* + h_1$, with e_1 and h_1 being the energies of the first quantized states of electrons and holes, respectively.

33.2.3 Dilute Nitride Quantum Well Lasers

The ability to grow InGaAsN quantum wells lattice matched on GaAs substrates and the formation of a type I band alignment to confine both electrons and holes makes this dilute nitride alloy promising for device applications. Most work focused on fabrication of injection lasers emitting in the 1.3 and 1.55 μm wavelength windows of common datacom fibres, stimulated by the great commercial demand and the compatibility with the well-developed GaAs/AlAs technology. In fact, InP-based device structures used in optical networks suffer from poor thermal conductivity of the substrate material and low refractive index variation of alloys. InGaAsN also gained interest as a 1 eV subcell added to commercially available $InGaP_2$/GaAs/Ge multijunction solar cells. Inserted between GaAs and Ge, conversion efficiencies exceeding 40% are expected [33.52]. In the following some aspects of the work related to edge-emitting lasers are discussed.

InGaAsN QWs for long-wavelength lasers were introduced in the mid 1990s using both gas-source MBE (GSMBE, also denoted metalorganic MBE, MOMBE) and MOVPE [33.53]. The first edge-emitting ridge waveguide lasers used single QWs and achieved emission wavelengths near 1.2 μm at 300 K with high threshold current densities exceeding 1 kA/cm^2. Lasers grown using MOVPE have focused since then on the 1.3 μm range, while considerable progress in the 1.55 μm range was achieved with MBE by additionally introducing Sb as a surfactant during growth of the active quantum wells in lasers [33.30]. Most lasers have a single QW with a thickness of 6–8 nm. During the first years of research on diluted nitride lasers a general trend of increasing threshold current densities was found when the emission wavelength was increased to longer wavelengths, irrespective of the growth technique employed, MOVPE or MBE [33.28, 54–56]. Wavelength tuning beyond 1.2 μm requires a sizeable incorporation of nitrogen into InGaAs. This can be achieved by a decreased growth temperature, an increased growth rate, and an increased N/(N+As) ratio [33.54]. The latter is more easily achieved using a low As/(Ga+In) ratio rather than a high flow of the N source, another advantage of employing tertiarybutylarsine over AsH_3 as arsenic source [33.57].

Generally such wavelength tuning was accompanied by material deterioration. A direct correlation of the threshold current density to the concentration of carbon in the QWs was observed [33.28]. For lasers grown using MOVPE the finding could be related to C from the gallium source. The high levels were claimed to be a consequence of the strong C–N bond. A similar trend found in MBE-grown lasers indicated a respective issue also for this growth technique.

A high PL efficiency of the active QW is crucial to obtain a low lasing threshold. As noted in Sect. 33.2.2, the PL, which drops for high N incorporation, largely recovers by annealing. A major improvement of InGaAsN QW lasers was therefore obtained by adding an annealing step to the fabrication procedure [33.58]. This step can also be accomplished in situ during growth of the upper cladding layer at increased temperature [33.57]. Further improvements were achieved by engineering the confinement and strain in the quantum well. Cladding of the compressively strained QW by a strain-compensating tensile

GaAsP barrier layer led to a threshold current density as low as 211 A/cm^2 for lasing at 1295 nm [33.57]. Using tensile-strained GaAsN instead of GaAsP proved to be even more advantageous [33.59, 60]. The conduction band discontinuity from the quaternary QW to the GaAs waveguide is segmented into two steps, thereby reducing the quantum confinement of the electrons and hence inducing a red-shift of the emission. In addition, stronger hole confinement is achieved due to a larger valence band step from the QW to GaAsN with respect to GaAs, resulting in improved high-temperature characteristics of the device. The characteristic temperature T_0 was shown to increase by 15 K for lasers emitting near 1315 nm with thresholds of 210–270 A/cm^2 [33.59]. Using this approach, lasing at the longest wavelength reported to date of 1410 nm was achieved, with a threshold of 1.4 kA/cm^2 [33.60].

Shift of the lasing wavelength to 1.55 μm will certainly be difficult to achieve with InGaAsN QWs. The results obtained using MOVPE are largely comparable to those achieved using MBE of lasers with quaternary QWs [33.61–63]. The most striking difference is a substantially lower temperature during growth of the active QW, ranging between 350 °C and 330 °C in MBE compared with about 530 °C in MOVPE. The longest wavelength reached to date, by applying a carefully adjusted low-temperature MBE, is 1510 nm, with a still very low threshold of 780 A/cm^2 [33.56]. Substantial progress beyond 1.5 μm was recently achieved with quinternary quantum wells in MBE [33.64]. Using a single $In_{0.38}Ga_{0.62}As_{0.943}N_{0.03}Sb_{0.027}$ QW with cladding by strain-compensating $GaAs_{0.96}N_{0.04}$ barriers, lasing was observed at 1.55 μm with a low threshold of 579 A/cm^2.

33.3 InAs/GaAs Quantum Dots

Semiconductor quantum dots (QDs) are nanometer-sized objects in which charge carriers are confined in all three spatial dimensions. Their size is in the range of the de Broglie matter wavelength of the confined particles, and the quantum size effect leads to a delta-function-like electronic density of states. The properties of such zero-dimensional structures resemble those of atoms. The unique electronic properties of QDs, the discovery of self-organization processes which control their coherent (i. e., defect-free) growth, and the ease of incorporating QDs into a semiconductor device led to rapid development in this field. This section focuses on InAs QDs in GaAs matrix material and provides an insight into the basics of their formation, structural, and electronic properties. Finally, results on recent applications in edge-emitting lasers with gain media comprising QDs are reported.

33.3.1 The Stranski–Krastanow 2-D–3-D Transition

Efforts to realize quantum dots in semiconductor heterostructures were initially devoted mainly to lithographic patterning of quantum wells, e.g., [33.65]. Such structures generally suffered from residual damage introduced by nanopatterning that affected the electronic and optical properties. In the 1990s the concept of exploiting the fundamental growth mode named after Stranski and Krastanow was introduced to form large arrays of defect-free quantum dots in a self-organized (also called self-assembled or self-ordered) way. The original paper describes the formation of islands on a flat substrate surface in heteroepitaxy of lattice-matched ionic crystals with different charges [33.66]. Stranski–Krastanow growth can also be induced if a two-dimensional layer is deposited on a crystalline substrate which has a sufficiently large *mismatch* of the lateral lattice constant (typically $\Delta a/a > 2\%$). The layer adopts the lateral atomic spacing of the substrate and hence accumulates strain with increasing thickness. Above a critical layer thickness the strain is relaxed by introduction of misfit dislocations. Below this critical thickness, a considerable part of the strain can be relaxed *elastically*, i. e., without introduction of dislocations, by forming facetted surface structures. A precondition for this 2-D–3-D transition is a total energy gain of the heteroepitaxial system, expressed by the actual surface area A and the change of the areal energy density $\Delta\gamma$. This quantity is given by

$$\Delta\gamma = \gamma_{\text{surface}} + \gamma_{\text{interface}} - \gamma_{\text{substrate}} , \qquad (33.11)$$

the three summands being the surface energy of the epitaxial layer, the energy of the interface between layer and substrate, and the surface energy of the substrate, respectively. If $\Delta\gamma < 0$, the layer wets the substrate surface and grows two-dimensionally in the Frank–van-der-Merve layer-by-layer growth mode. If $\Delta\gamma > 0$, the layer tries to leave the substrate uncovered and

grows three-dimensionally in the Volmer–Weber island growth mode. In the intermediate case of Stranski–Krastanow growth, $\Delta\gamma < 0$ applies for the first layer(s), which grow two-dimensionally, while $\Delta\gamma > 0$ for subsequently grown layers, e.g., due to accumulated strain. The driving force of the 2-D–3-D transition is therefore minimization of the total strain energy [33.67]. The total energy gain per volume of a single QD can be expressed by [33.68]

$$\frac{E_{\text{total}}}{V} = \varepsilon_{\text{QD}}^{\text{elast}} - \varepsilon_{\text{layer}}^{\text{elast}} + \frac{A\gamma_{\text{facet}} - L^2\gamma_{\text{layer}}(d_0)}{V} + \left(\frac{1}{\rho} - L^2\right)\frac{\gamma_{\text{layer}}(d) - \gamma_{\text{layer}}(d_0)}{V}\,, \tag{33.12}$$

$\varepsilon_{\text{QD}}^{\text{elast}}$ and $\varepsilon_{\text{layer}}^{\text{elast}}$ being the elastic energy densities of the QD and the uniformly strained layer. The third term describes the change in surface energy due to the QD, with γ_{facet} being the surface energy of the island facets, A their area and L the base length of the QD, which is assumed to have a pyramidal shape. The fourth term accounts for that part of the layer which converts to the QD; ρ is the area density of QDs, $\gamma_{\text{layer}}(d_0)$ and $\gamma_{\text{layer}}(d)$ are the formation energies of the layer as a function of its thickness d. The sum of these contributions is the total energy density, which has an energy minimum for a particular dot size as shown in Fig. 33.10. This minimum causes preferential formation of dots around this size and a stability of the dot ensemble against Ostwald ripening in thermodynamic equilibrium. The ripening still observed in experiments indicates the presence of kinetic barriers under the usually applied growth conditions. It should be noted that, even under equilibrium conditions, the entire material of the two-dimensional layer does not entirely reorganize to QDs; rather, a thin 2-D layer, the wetting layer, remains.

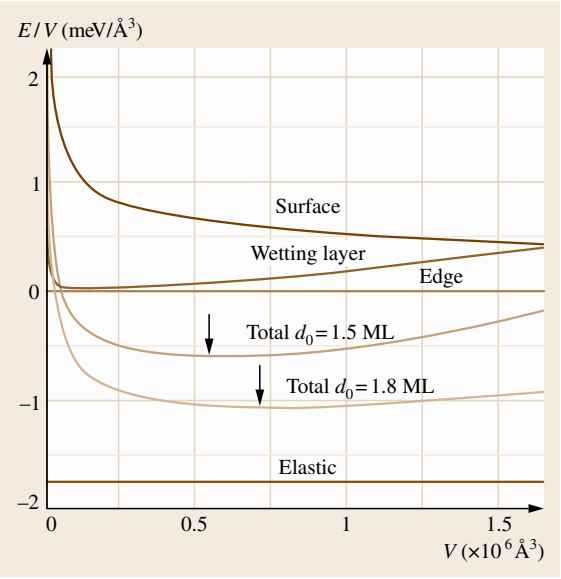


Fig. 33.10 Total energy gain by QD formation from $d_0 = 1.8$ monolayers (ML) and 1.5 monolayers InAs films deposited on GaAs. A QD area density of $\rho = 10^{10}$ cm^{-2} is assumed; notations on the curves denote contributions from the surface, the wetting layer, the edges of the assumed pyramid shape of the QDs, and the elastic relaxation energy [33.68]. *Arrows* mark the minima of the curves

Self-organized growth of InAs QDs on GaAs was first realized using MBE [33.70]. Since then InAs and $In_{1-x}Ga_xAs$ [33.71] dots in GaAs matrix have become a model system for self-organized QD growth. The InAs lattice constant exceeds that of GaAs by $\approx 7\%$. Details of QD formation sensitively depend on growth parameters, particularly on temperature, growth rate, and arsenic partial pressure. The 2-D–3-D transition of InAs layers occurs on (001)-oriented GaAs between 1.5 and 1.8 monolayer thickness for typical deposition temperatures of 450–520 °C. Nucleation of 3-D dots starts on top of a 2-D InAs layer which exceeds some critical thickness d_c of coverage. The density ρ of the emerging dots was measured from atomic force microscopy (AFM) images of InAs layers with varied thickness, deposited on (001) GaAs at 530 °C using MBE, cf. Fig. 33.11.

The dot density ρ depicted in Fig. 33.11 follows a dependence on the InAs coverage similar to that of

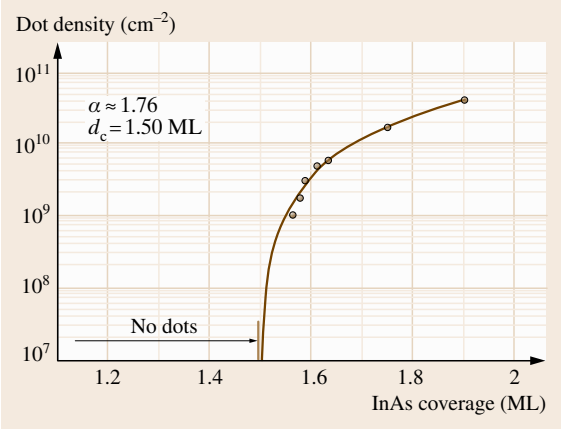


Fig. 33.11 Areal density of self-organized InAs dots on GaAs as a function of deposition thickness d [33.69]

a first-order phase transition [33.69]

$$\rho = \rho_0 (d - d_c)^{\alpha} \,. \tag{33.13}$$

The least-squares fit represented by the solid line yields $\rho_0 = 2 \times 10^{11}\,\mathrm{cm}^{-2}$, $\alpha = 1.76$, and $d_c = 1.50\,\mathrm{ML}$.

33.3.2 MOVPE of InAs Quantum Dots

Most studies on the MOVPE of quantum dots in the InGaAs system comprise ternary $In_xGa_{1-x}As$ dots with In content x of 30–70% [33.73–75]. The reduced strain in the layer with respect to InAs retards the onset of dot nucleation, facilitating deposition control, and the tendency to form large, dislocated clusters is smaller [33.74]. The growth procedure is often varied in many respects, complicating a comparison of results: The dot material may be deposited on a thin InGaAs buffer layer with lower In content to obtain a high dot density, or overgrown by a respective thin InGaAs layer to induce a red-shift of the emission wavelength [33.76]. In addition, seeding or stacking of dot layers is applied to modify optical properties or increase the dot density in the matrix [33.72, 77], and also atomic layer epitaxy [33.78] or use of a surfactant [33.79] has been reported. Structural and electronic properties of InGaAs QDs were found to differ significantly from those of InAs dots. For InGaAs depositions a depletion of the effective In content in the wetting layer was found [33.80], and an inhomogeneous In distribution in the InGaAs dot was observed [33.81]. To treat a more concise situation, we focus on the MOVPE of binary InAs dots. The precursors used for QD growth are standard sources for III–V epitaxy, namely trimethylindium, trimethylgallium, and either arsine or tertiarybutylarsine.

The general procedure of self-organized QD fabrication starts with the preparation of the surface. The low deposition temperature of typically about 500 °C, which is required for Stranski–Krastanow growth of InAs dots, is much lower than the temperature needed to grow GaAs matrix material with good quality at reasonable growth rate. Therefore first a GaAs buffer layer is grown at a suitable high temperature (e.g., 650 °C), and then the surface is stabilized with arsenic partial pressure while setting up the lowered temperature and eventually adjusting the V/III ratio. Then InAs is deposited, and subsequently a growth interruption is applied. Finally a GaAs cap layer is grown on top. Each of these steps comprises some crucial issues to obtain defect-free structures.

The deposition of InAs must occur with some minimum rate to prevent the formation of large incoherent clusters, in contrast to MBE, where processes can largely be retarded by using very low temperatures. The QD properties depend sensitively on the InAs thickness and the As partial pressure during InAs deposition and the growth interruption. Figure 33.12 shows PL spectra of InAs/GaAs QD samples prepared at 485 °C and 8 s growth interruption with different depositions thicknesses and V/III ratios [33.72]. The spectra were nonresonantly excited into the matrix and show the near-band-edge emission of GaAs above 1.4 eV, emission from the 2-D wetting layer (WL), and the luminescence of the QD ensemble. For a given InAs deposition thickness d_0, Fig. 33.12a shows the emission from a 2-D wetting layer clad by GaAs (dark line

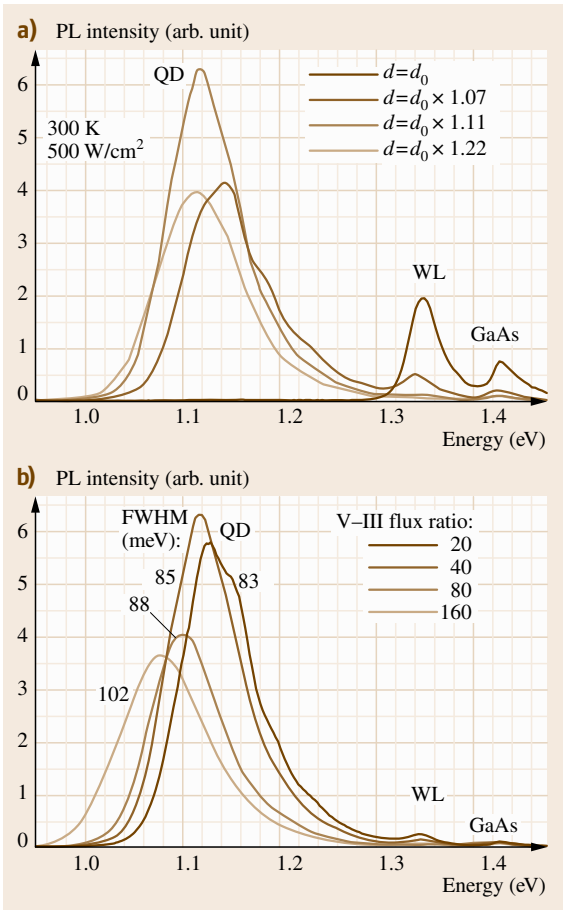


Fig. 33.12a,b PL spectra of InAs/GaAs quantum dot samples with different layer thicknesses **(a)** and different V/III ratios **(b)** applied during InAs deposition [33.72]

$d = d_0$); from the emission energy a thickness of 1.5 ML was evaluated. QDs are formed when this critical value is exceeded, and an optimum at 1.11 d_0 was found for the given growth conditions. The decrease of PL intensity for thicker InAs depositions indicates the onset of plastic strain relaxation. Keeping the optimum thickness d_0, the V/III ratio was varied in the series given in Fig. 33.12b by gradually changing the arsine partial pressure from 0.09 to 0.72 mbar. The PL shows a redshift, accompanied by a broadening of the full width at half maximum (FWHM) of the emission. Maximum efficiency and minimum FWHM are obtained for lowest V/III ratios. It must be noted that the effective V/III ratio at the growing surface is much lower than the quoted gas-phase ratios due to a quite incomplete decomposition of the stable arsine at 485 °C, cf. Fig. 33.2. A high V/III ratio apparently degrades the QDs. This applies also for the use of tertiarybutylarsine instead of arsine. Due to the high cracking efficiency, growth temperature and V/III ratio then become largely independent parameters [33.82], and a V/III ratio near unity is usually applied [33.79, 83].

The growth interruption conditions also strongly influence QD properties. In the sample series characterized in Fig. 33.13 the arsine flow was varied during a constant growth interruption (GRI) duration of 14 s [33.84]. Maximum PL efficiency is obtained when arsine is switched off during the first 12 s. In contrast, if the surface is stabilized under an As partial pressure during the entire growth interruption, the PL intensity drops by two orders of magnitude. AFM images of such samples show a high density of large, dislocated clusters. The only slight drop in PL intensity for the intermediate case of 3 s without arsine supply (factor 2) shows that the first stage after InAs deposition is most critical. The presence of arsine during the growth interruption of QDs prepared using MOVPE may obviously affect QD properties. Such behavior is not reported for InAs dots grown using MBE. An enhanced In mobility in MOVPE was suggested to account for this finding [33.84].

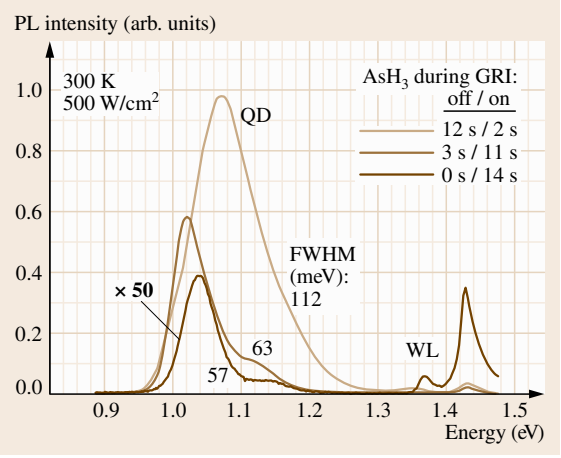


Fig. 33.13 PL spectra of InAs/GaAs QD samples with different arsine fluxes supplied during growth interruption after InAs deposition [33.84]

A cap layer is deposited after the growth interruption, allowing for integration of QDs into a device structure. The temperature is kept unchanged at least during growth of the first few nanometers due to the instable state of uncovered QDs. GaAs growth at such low temperatures favors the formation of defects, which affect the optical properties of QDs. Since buried InAs QDs are stable below 600 °C [33.85], an overgrowth procedure was established which maintains a high luminescence efficiency [33.86]. Here, the dots are buried by a thin (2–3 nm) GaAs cap at QD growth temperature, and further GaAs overgrowth is accomplished during a temperature increase to 600 °C. Besides an essential improvement of optical properties, a flat growth front is re-established at the otherwise corrugated surface above QDs. This is a precondition for continued QD growth in a stack, made to achieve an increased optical confinement factor and to overcome gain saturation in lasers with QD gain media. Furthermore, the flattening reduces interface roughness for subsequently deposited upper waveguide cladding, and thereby optical cavity losses in edge-emitting lasers.

Detailed insight into the dynamics of formation and evolution of InAs QDs was gained by studying ensembles with a multimodal size distribution [33.87, 88]. The particular feature of this kind of QDs is decomposition of the ensemble luminescence into individual lines as shown in Fig. 33.14. The QDs have the shape of a truncated InAs pyramid with a flat top facet, and the individual lines were shown to originate from subensembles which differ in height by an integer number of InAs monolayers. The clear detection of size quantization allows to trace redistribution of material among the QDs during the growth interruption.

InAs QDs with a multimodal size distribution form after a comparatively fast deposition of a rough, two-dimensional layer with a thickness close to the critical value for the 2-D–3-D transition [33.89]. If no growth interruption is applied, a quantum *well* is formed as shown in the inset of Fig. 33.14. The long exponential low-energy tail of the PL indicates significant thickness

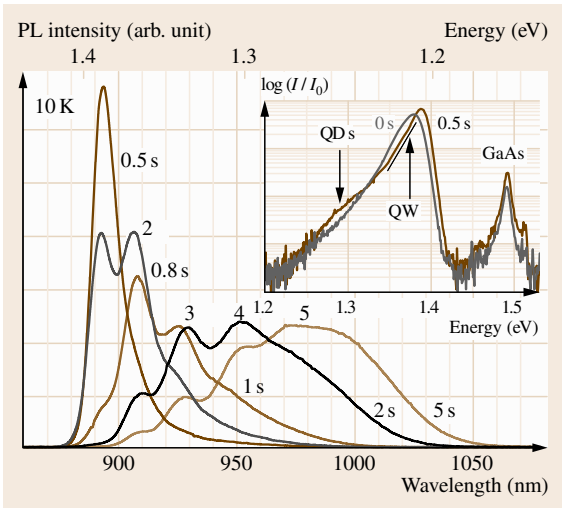


Fig. 33.14 PL spectra of QD samples grown with various growth interruptions (indicated in seconds on the curves); *numbers* denote QD heights in units of monolayers. *Inset*: spectra of samples grown without and with 0.5 s growth interruption [33.91]

fluctuations. Introduction of a short interruption leads to a blue-shift of the QW PL-maximum, i. e., a thinning of the QW, and the appearance of locally thicker regions (QDs). For longer interruptions (≥ 0.8 s) the emission of a multimodal QD ensemble with individual subensemble peaks evolves from shallow localizations in the density-of-states (DOS) tail of the rough QW. The material hence partially concentrates at some QD precursors. It must be noted that the thickness and composition of the wetting layer were found to remain constant during the entire duration of the growth interruption [33.87, 88]. The PL of the QD ensemble experiences a red-shift for longer interruptions due to an increase of the average QD volume, a behavior generally reported for self-organized QDs, cf. Fig. 33.15a. The persistent shift also occurring for very long interruptions indicates the action of kinetic barriers which prevent the achievement of thermal equilibrium on the studied time scale. The drop of PL intensity at the longest growth interruption indicates the onset of plastic relaxation, as confirmed by TEM images. Interestingly, the subensemble peaks shift to the *blue* during QD evolution. This and the PL intensity decrease of a given subensemble peak with prolonged interruption time indicate that the ripening of the QDs if fed by dissolution of smaller QDs. The development of the integral subensemble PL intensities was considered a reasonable measure for the number of QDs with a specific corresponding height, showing directly the formation of subensembles with larger dots by dissolution of smaller QDs [33.90], cf. Fig. 33.15b. The subensemble with 5 ML QD height, e.g., grows during the initial 5 s growth interruption at the expense of those with 2–4 ML height. Then it starts to dissolve, feeding the 6 ML-high subensemble, which after 10 s interruption also starts to dissolve.

The optical study noted above is confirmed by structural data of the InAs QDs. Using strain and chemically sensitive {200} reflections, InAs dots appear as bright central areas in TEM images with a low aspect ratio and flat top and bottom interfaces, cf. Fig. 33.16. Side facets are steep, but less well resolved due to strain perturbations. Analysis of such images show that both the base and the top layer of the flat dots apparently consist of plain, continuous InAs layers [33.87, 88]. This yields QDs with a shape of a truncated pyramid, which

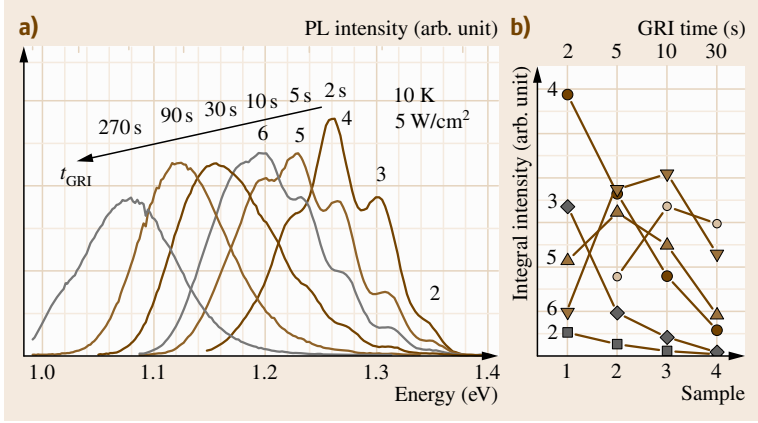


Fig. 33.15 **(a)** PL spectra of QD samples grown with long growth interruptions t_{GRI}. **(b)** Integral PL intensity of subensemble peaks as a function of growth interruption time. *Numbers* refer to QD heights as indicated in (a)

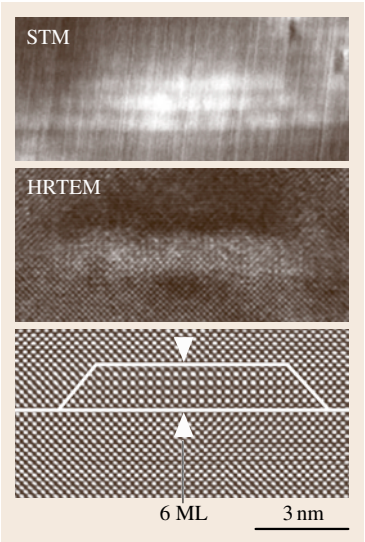


Fig. 33.16 Cross-sectional scanning tunneling micrograph (*top*, [33.92]) and high-resolution transmission electron micrograph [33.87] of InAs/GaAs QDs with 6 ML height; *bottom*: Fourier-filtered image of the transmission electron micrograph

differ in height by integer numbers of InAs monolayers. The same conclusions were drawn from recent scanning tunneling micrographs [33.92].

33.3.3 Quantum Dot Lasers

Work on InGaAs quantum dot lasers is a competing development to InGaAsN quantum well lasers to use GaAs-based laser technology for the long-wavelength datacom range. Zero-dimensional charge-carrier localization in the active region of a semiconductor laser was predicted two decades ago to lead to superior device performance, e.g., with respect to decreased threshold current and high-temperature stability [33.93, 94]. Furthermore, the reduction of lateral charge-carrier diffusion suppresses beam filamentation and strongly enhances robustness against facet degradation during high-power operation. The first realization of a QD injection laser using a single layer of MBE-grown self-organized $In_{0.5}Ga_{0.5}As$ QDs proved the fundamental validity of theoretical predictions [33.95]. A low lasing threshold of 120 A/cm^2 and a characteristic temperature T_0 as high as 350 K were achieved at liquid-nitrogen temperature. At room temperature, lasing occurred at a high threshold of 950 A/cm^2 from an excited state due to ground-state gain saturation. Generally the lasing threshold is reached if the modal gain g_{mod} just equilibrates the internal losses α_{int} and those of the cavity mirrors α_{mirr},

$$g_{mod} = \Gamma g_{mat} = \alpha_{int} + \alpha_{mirr} \,. \qquad (33.14)$$

The material gain of QD gain media g_{mat} exceeds that of QWs by far [33.94, 96]. The optical confinement factor Γ of a QD layer is, however, much smaller than that of a QW, leading to a largely comparable modal gain of QD and QW lasers. Since the number of states in a QD layer is quite small compared with a QW, low lasing threshold and also a low gain saturation level are easily reached in single active layers. To increase both confinement factor and gain saturation level, actual QD lasers comprise a *stack* of QD layers. The spacers between the QD layers should preferably be thin to position the QDs near the optical field maximum. Further improvements of QD lasers aim to confine charge carriers efficiently in the QDs at energies well below that of the wetting layer.

The first MOVPE-grown QD lasers, reported in 1997, already used a threefold InAs QD stack [33.97] and a tenfold $In_{0.5}Ga_{0.5}As$ QD stack [33.98]. Both approaches achieved ground-state lasing at room temperature, and a T_0 of 385 K was demonstrated for the tenfold stack laser up to 50 °C. The stacks were grown at the same low temperature for QD layers and GaAs spacers. Substantial progress in laser performance has been made since then, particularly by introducing temperature ramping to improve spacer quality and smoothen the interfaces [33.86]. For threefold stacked dot layers emitting at 1.16 μm, threshold and transparency current densities of 110 and 18 A/cm^2, respectively, with internal quantum efficiency exceeding 90% were achieved [33.99]. Also the internal loss below 1.5 cm^{-1} substantiates the benefit of temperature ramping. All these devices were grown using arsine as group V precursor. Good data were likewise reported for lasers grown using tertiarybutylarsine. Devices emitting at 1.1 μm showed a transparency current below 30 A/cm^2, 91% internal quantum efficiency, and internal loss of 2.2 cm^{-1} [33.82].

Much work has been devoted to extending the emission wavelength towards 1.3 μm. An efficient means is overgrowth of the In(Ga)As QDs by an InGaAs QW with lower In content to form a dot-in-a-well (DWELL) structure [33.76, 100]. Since this QW reduces the strain on the buried QD layer with respect to a GaAs cap, it is also referred to as a strain-reducing layer (SRL). Note that the *overall* strain in the QD stack is increased by the QW. Furthermore, the *local* strain at the large, In-rich QDs required for long-wavelength emission is rather high. This easily leads to formation of large In-rich incoherent clusters, which degrade device characteristics. Using the DWELL approach and carefully adjusting the growth parameters for each individual QD layer in the stack to avoid defect for-

mation, lasing at 1250 nm with a very low threshold current density of 66 A/cm^2 and 94% internal quantum efficiency was recently obtained [33.101]. In a complementary approach a DWELL laser was subjected to a postgrowth annealing procedure at moderate 560 °C, yielding 1280 nm lasing at high threshold [33.102]. As an alternative, strain-compensating GaP layers were introduced into the stacking sequence. Lasing at 1249 nm with a reasonable threshold of 550 A/cm^2 was reported for partial strain compensation using 4 ML-thick GaP layers [33.103]. The presently (2006) most advanced MOVPE-grown lasers used 1.5 nm-thick lattice-matched InGaP buffers below the stacked InGaAs QD layers, yielding 1280 nm lasing at 200 A/cm^2 threshold with $T_0 = 210$ K [33.104]. The target of 1.3 μm lasing has been accomplished to date with DWELL devices grown using MBE, a technique much more widely employed in the field. By applying temperature ramping comparable to that described above for the MOVPE, operation at 1307 nm with an extraordinary low threshold of 33 A/cm^2 (17 A/cm^2 with facet coating) was achieved [33.105].

33.4 Concluding Remarks

Metalorganic vapor-phase epitaxy has become established as a versatile technique for advanced fabrication of heterostructures with thickness control down to the monolayer range. Besides applications in industrial mass production, fast progress is being made in the development of new fields. This chapter intended to deliver an insight into the technique and selected problems. Just two topics on current issues in semiconductor MOVPE were highlighted: epitaxy of GaAs-related dilute nitrides and quantum dots. Both aim at extending the well-developed GaAs-based technology to the long-wavelength spectral range and have recently achieved significant advances in the field of lasers. The performance of these demanding devices is comparable to the best fabricated devices using the complementary molecular-beam epitaxy described in Chap. 32. Both growth techniques presently provide the basis for novel low-dimensional structures in research and development. Note added in proof: Recently ground-state lasing at 1.35 μm and a threshold down to 328 A/cm^2 were accomplished by employing antimony-mediated MOVPE with tenfold stacked InAs/GaAs QDs [33.106].

References

33.1 W. Miederer, G. Ziegler, R. Dötzer: Verfahren zum tiegelfreien Herstellen von Galliumarsenidstäben aus Galliumalkylen und Arsenverbindungen bei niedrigen Temperaturen, German Patent 1176102, filed 25.9.1962; and: Method of crucible-free production of gallium arsenide rods from alkyl galliums and arsenic compounds at low temperatures, US Patent 3226270, filed 24.9.1963

33.2 H.M. Manasevit, W.I. Simpson: The use of metal-organics in the preparation of semiconductor materials on insulating substrates: I. Epitaxial III–V gallium compounds, J. Electrochem. Soc. **12**, 66C (1968)

33.3 R.W. Thomas: Growth of single crystal GaP from organometallic sources, J. Electrochem. Soc. **116**, 1449–1450 (1969)

33.4 H.M. Manasevit: The use of metalorganics in the preparation of semiconductor materials: Growth on insulating substrates, J. Cryst. Growth **13/14**, 306–314 (1972)

33.5 G.B. Stringfellow: Organometallic vapor phase epitaxy reaction kinetics. In: *Handbook of Crystal Growth*, ed. by D.R.T. Hurle (Elsevier, Amsterdam 1994) pp. 491–540

33.6 T.J. Mountziaris, K.F. Jensen: Gas-phase and surface reaction mechanisms in MOCVD of GaAs with trimethyl-gallium and arsine, J. Electrochem. Soc. **138**, 2426–2439 (1991)

33.7 G.B. Stringfellow: *Organometallic Vapor-Phase Epitaxy*, 2nd edn. (Academic, New York 1999)

33.8 R.T. Morrison, R.N. Boyd: *Organic Chemistry*, 5th edn. (Allyn & Bacon, New York 1987)

33.9 K.F. Jensen: Transport phenomena in vapor phase epitaxy reactors. In: *Handbook of Crystal Growth*, ed. by D.R.T. Hurle (Elsevier, Amsterdam 1994) pp. 541–599

33.10 D.H. Reep, S.K. Ghandhi: Deposition of GaAs epitaxial layers by organometallic CVD, J. Electrochem. Soc. **130**, 675–680 (1983)

33.11 R.B. Bird, W.E. Stewart, E.N. Lightfoot: *Transport Phenomena* (Wiley, New York 1962)

33.12 A.G. Bhuiyan, A. Hashimoto, A. Yamamoto: Indium nitride (InN): A review on growth, characterization, and properties, J. Appl. Phys. **94**, 2779–2808 (2003)

33.13 I. Vurgaftman, J.R. Meyer: Band parameters for nitrogen-containing semiconductors, J. Appl. Phys. **94**, 3675–3696 (2003)

33.14 M. Drago, T. Schmidtling, U.W. Pohl, S. Peters, W. Richter: InN metalorganic vapour phase epitaxy and ellipsometric characterization, Phys. Status Solidi (c) **0**(7), 2842–2845 (2003)

33.15 I. Vurgaftman, J.R. Meyer, L.R. Ram-Mohan: Band parameters for III–V compound semiconductors and their alloys, J. Appl. Phys. **89**, 5815–5875 (2001)

33.16 U.W. Pohl, C. Möller, K. Knorr, W. Richter, J. Gottfriedsen, H. Schumann, K. Rademann, A. Fielicke: Tertiarybutylhydrazine: A new precursor for the MOVPE of group III-nitrides, Mater. Sci. Eng. B **59**, 20–23 (1999)

33.17 I. Ho, G.B. Stringfellow: Solubility of nitrogen in binary III–V systems, J. Cryst. Growth **178**, 1–7 (1997)

33.18 J. Neugebauer, C.G. Van de Walle: Electronic structure and phase stability of $GaAs_{1-x}N_x$ alloys, Phys. Rev. B **51**, 10568–10571 (1995)

33.19 D.K. Gaskill, N. Bottka, M.C. Lin: Growth of GaN films using trimethylgallium and hydrazine, Appl. Phys. Lett. **48**, 1449–1451 (1986)

33.20 H. Tsuchiya, A. Takeuchi, M. Kurihara, F. Hasegawa: Metalorganic molecular beam epitaxy of cubic GaN on (100) GaAs substrates using triethylgallium and monomethylhydrazine, J. Cryst. Growth **152**, 21–27 (1995)

33.21 S. Miyoshi, K. Onabe, N. Ohkouchi, H. Yaguchi, R. Ito, S. Fukatsu, Y. Shiraki: MOVPE growth of cubic GaN on GaAs using dimethylhydrazine, J. Cryst. Growth **124**, 439–442 (1992)

33.22 H. Sato, H. Takanashi, A. Watanabe, H. Ota: Preparation of GaN films on sapphire by metalorganic chemical vapor deposition using dimethylhydrazine as nitrogen source, Appl. Phys. Lett. **68**, 3617–3619 (1996)

33.23 U.W. Pohl, K. Knorr, C. Möller, U. Gernert, W. Richter, J. Bläsing, J. Christen, J. Gottfriedsen, H. Schumann: Low-temperature metalorganic vapor phase epitaxy (MOVPE) of GaN using tertiarybutylhydrazine, Jpn. J. Appl. Phys. **38**, L105–L107 (1999)

33.24 U.W. Pohl, K. Knorr, J. Bläsing: Metalorganic vapor phase epitaxy of GaN on $LiGaO_2$ substrates using tertiarybutylhydrazine, Phys. Status Solidi (a) **184**, 117–120 (2001)

33.25 Z. Pan, T. Miyamoto, D. Schlenker, S. Sato, F. Koyama, K. Iga: Low temperature growth of GaInNAs/GaAs quantum wells by metalorganic chemical vapor deposition using tertiarybutylarsine, J. Appl. Phys. **84**, 6409–6411 (1998)

33.26 J. Derluyn, I. Moerman, M.R. Leys, G. Patriarche, G. Sęk, R. Kudrawiec, W. Rudno-Rudziński, K. Ryczko, J. Misiewicz: Control of nitrogen incorporation in Ga(In)NAs grown by metalorganic vapor phase epitaxy, J. Appl. Phys. **94**, 2752–2754 (2003)

33.27 S. Kurtz, J.F. Geisz, D.J. Friedman, W.K. Metzger, R.R. King, N.H. Karam: Annealing-induced-type conversion of GaInNAs, J. Appl. Phys. **95**, 2505–2508 (2004)

33.28 K. Volz, T. Torunski, B. Kunert, O. Rubel, S. Nau, S. Reinhard, W. Stolz: Specific structural and compositional properties of (GaIn)(NAs) and their influence on optoelectronic device performance, J. Cryst. Growth **272**, 739–747 (2004)

33.29 S.G. Spruytte, C.W. Coldren, J.S. Harris, W. Wampler, P. Krispin, K. Ploog, M.C. Larson: Incorporation of nitrogen in nitride-arsenides: Origin of improved luminescence efficiency after anneal, J. Appl. Phys. **89**, 4401–4406 (2001)

33.30 X. Yang, M.J. Jurkovic, J.B. Héroux, W.I. Wang: Molecular beam epitaxial growth of InGaAsN:Sb/GaAs quantum wells for long-wavelength semiconductor lasers, Appl. Phys. Lett. **75**, 178–180 (1999)

33.31 L.H. Li, V. Sallet, G. Patriarche, L. Largeau, S. Bouchoule, L. Travers, J.C. Harmand: Investigations on GaInNAsSb quinary alloy for 1.5 µm laser emission on GaAs, Appl. Phys. Lett. **83**, 1298–1300 (2003)

33.32 H.B. Yuen, S.R. Bank, H. Bae, M.A. Wistey, J.S. Harris: The role of antimony on properties of widely varying GaInNAsSb compositions, J. Appl. Phys. **99**, 093504 (2006)

33.33 H.A. McKay, R.M. Feenstra, T. Schmidtling, U.W. Pohl: Arrangement of nitrogen atoms in GaAsN alloys determined by scanning tunneling microscopy, Appl. Phys. Lett. **78**, 82–84 (2001)

33.34 H.A. McKay, R.M. Feenstra, T. Schmidtling, U.W. Pohl, J.F. Geisz: Distribution of nitrogen atoms in dilute GaAsN and InGaAsN alloys studied by scanning tunneling microscopy, J. Vac. Sci. Technol. B **19**, 1644–1649 (2001)

33.35 R. Duca, G. Ceballos, C. Nacci, D. Furlanetto, P. Finetti, S. Modesti, A. Cristofolio, G. Bais, M. Peccin, S. Rubini, F. Martelli, A. Franciosi: In-N and N-N correlation in $In_xGa_{1-x}As_{1-y}N_y$/GaAs quasi-lattice-matched quantum wells: A cross-sectional scanning tunneling microscopy study, Phys. Rev. B. **72**, 075311 (2005)

33.36 L. Bellaiche, A. Zunger: Effects of atomic short-range order on the electronic and optical properties of GaAsN, GaInN, and GaInAs, Phys. Rev. B **57**, 4425–4431 (1998)

33.37 O. Rubel, K. Volz, T. Torunski, S.D. Baranovskii, F. Grosse, W. Stolz: Columnar [001]-oriented nitrogen order in Ga(NAs) and (GaIn)(NAs) alloys, Appl. Phys. Lett. **85**, 5908–5910 (2004)

33.38 K. Volz, T. Torunski, W. Stolz: Detection of nanometer-sized strain fields in (GaIn)(NAs) alloys by specific dark field transmission electron microscopic imaging, J. Appl. Phys. **97**, 014306 (2005)

33.39 G. Ciatto, F. D'Acapito, L. Grenouillet, H. Mariette, D. De Salvador, G. Bisognin, R. Carboni,

L. Floreano, R. Gotter, S. Mobilio, F. Boscherini: Quantitative determination of short-range ordering in $In_xGa_{1-x}As_{1-y}N_y$, Phys. Rev. B **68**, 161201 (2003)
33.40 V. Lordi, H.B. Yuen, S.R. Bank, M.A. Wistey, J.S. Harris, S. Friedrich: Nearest-neighbor distributions in $Ga_{1-x}In_xN_yAs_{1-y}$ and $Ga_{1-x}In_xN_yAs_{1-y-z}Sb_z$ thin films upon annealing, Phys. Rev. B **71**, 125309 (2005)
33.41 K. Uno, M. Yamada, I. Tanaka, O. Ohtsuki, T. Takizawa: Thermal annealing effects and local atomic configurations in GaInNAs thin films, J. Cryst. Growth **278**, 214–218 (2005)
33.42 E.-M. Pavelescu, J. Wagner, H.-P. Komsa, T.T. Rantala, M. Dumitrescu, M. Pessa: Nitrogen incorporation into GaInNAs lattice-matched to GaAs: The effects of growth temperature and thermal annealing, J. Appl. Phys. **98**, 083524 (2005)
33.43 A.M. Teweldeberhan, S. Fahy: Effect of indium-nitrogen bonding on the localized vibrational mode in $In_yGa_{1-y}N_xAs_{1-x}$, Phys. Rev. B **73**, 245215 (2006)
33.44 H.B. Yuen, S.R. Bank, H. Bae, M.A. Wistey, J.S. Harris: Effects of strain on the optimal annealing temperature of GaInNAsSb quantum wells, Appl. Phys. Lett. **88**, 221913 (2006)
33.45 T. Kitatani, M. Kondow, M. Kudo: Transition of infrared absorption peaks in thermally annealed GaInNAs, Jpn. J. Appl. Phys. **40**, L750–L752 (2001)
33.46 P.R.C. Kent, A. Zunger: Evolution of III–V nitride alloy electronic structure: The localized to delocalized transition, Phys. Rev. Lett. **86**, 2613–2616 (2001)
33.47 J.D. Perkins, A. Mascarenhas, Y. Zhang, J.F. Geisz, D.J. Friedman, J.M. Olson, S.R. Kurtz: Nitrogen-activated transitions, level repulsion, and band gap reduction in $GaAs_{1-x}N_x$ with $x < 0.03$, Phys. Rev. Lett. **82**, 3312–3315 (1999)
33.48 W. Shan, W. Walukiewicz, J.W. Ager III, E.E. Haller, J.F. Geisz, D.J. Friedman, J.M. Olson, S.R. Kurtz: Effect of nitrogen on the band structure of GaInNAs alloys, J. Appl. Phys. **86**, 2349–2351 (1999)
33.49 J.B. Héroux, X. Yang, W.I. Wang: Photoreflectance spectroscopy of strained (In)GaAsN/GaAs multiple quantum wells, J. Appl. Phys. **92**, 4361–4366 (2002)
33.50 M. Galluppi, L. Geelhaar, H. Riechert, M. Hetterich, A. Grau, S. Birner, W. Stolz: Bound-to-bound and bound-to-free transitions in surface photovoltage spectra: determination of the band offsets for $In_xGa_{1-x}As$ and $In_xGa_{1-x}As_{1-y}N_y$ quantum wells, Phys. Rev. B **72**, 155324 (2005)
33.51 R. Kudrawiec, M. Motyka, M. Gladysiewicz, J. Misiewicz, H.B. Yuen, S.R. Bank, H. Bae, M.A. Wistey, J.S. Harris: Band gap discontinuity in $Ga_{0.9}In_{0.1}N_{0.027}As_{0.973-x}Sb_x$/GaAs single quantum wells with $0 \leq x < 0.06$ studied by contactless electroreflectance spectroscopy, Appl. Phys. Lett. **88**, 221113 (2006)
33.52 A.J. Ptak, D.J. Friedman, S. Kurtz, R.C. Reedy: Low-acceptor-concentration GaInNAs grown by molecular-beam epitaxy for high-current *p–i–n* solar cell applications, J. Appl. Phys. **98**, 094501 (2005)
33.53 M. Kondow, T. Kitatani, S. Nakatsuka, M.C. Larson, K. Nakahara, Y. Yazawa, M. Okai, K. Uomi: GaInNAs: A novel material for long-wavelength semiconductor lasers, IEEE J. Sel. Top. Quantum Electron. **3**, 719–730 (1997)
33.54 S. Sato, S. Sato: Metalorganic chemical vapor deposition of GaInNAs lattice matched to GaAs for long-wavelength laser diodes, J. Cryst. Growth **192**, 381–385 (1998)
33.55 F. Höhnsdorf, J. Koch, S. Leu, W. Stolz, B. Borchert, M. Druminski: Reduced threshold current densities of (GaIn)(NAs)/GaAs single quantum well lasers for emission wavelengths in the range 1.28–1.38 µm, Electron. Lett. **35**, 571–572 (1999)
33.56 G. Jaschke, R. Averbeck, L. Geelhaar, H. Riechert: Low threshold InGaAsN/GaAs lasers beyond 1500 nm, J. Cryst. Growth **278**, 224–228 (2005)
33.57 N. Tansu, N.J. Kirsch, L.J. Mawst: Low-threshold-current-density 1300-nm dilute-nitride quantum well lasers, Appl. Phys. Lett. **81**, 2523–2525 (2002)
33.58 W. Li, J. Turpeinen, P. Melanen, P. Savolainen, P. Uusimaa, M. Pessa: Growth of strain-compensated GaInNAs/GaAsP quantum wells for 1.3 µm lasers, J. Cryst. Growth **230**, 533–536 (2001)
33.59 N. Tansu, J.-Y. Yeh, L.J. Mawst: Low-threshold 1317-nm InGaAsN quantum-well lasers with GaAsN barriers, Appl. Phys. Lett. **83**, 2512–2514 (2003)
33.60 J.-Y. Yeh, N. Tansu, L.J. Mawst: Long wavelength MOCVD grown InGaAsN-GaAsN quantum well lasers emitting at 1.378–1.41 µm, Electron. Lett. **40**, 739–740 (2004)
33.61 Y.Q. Wei, M. Sadeghi, S.M. Wang, P. Modh, A. Larsson: High performance 1.28 µm GaInNAs double quantum well lasers, Electron. Lett. **41**, 1328–1329 (2005)
33.62 B. Damilano, J. Barjon, J.-Y. Duboz, J. Massies, A. Hierro, J.-M. Ulloa, E. Calleja: Growth and in situ annealing conditions for long-wavelength (GaIn)(NAs)/GaAs lasers, Appl. Phys. Lett. **86**, 071105 (2005)
33.63 M. Hopkinson, C.Y. Jin, H.Y. Liu, P. Navaretti, R. Airey: 1.34 µm GaInNAs quantum well lasers with low room-temperature threshold current density, Electron. Lett. **42**, 923–924 (2006)
33.64 S.R. Bank, H.P. Bae, H.B. Yuen, M.A. Wistey, L.L. Goddard, J.S. Harris: Room-temperature continuous-wave 1.55 µm GaInNAsSb laser on GaAs, Electron. Lett. **42**, 156–157 (2006)
33.65 A. Forchel, H. Leier, B.E. Maile, R. Germann: Fabrication and optical spectroscopy of ultra small III–V compound semiconductor structure. In: *Advances in Solid State Physics*, Vol. 28, ed. by U. Rössler (Vieweg, Braunschweig 1988) pp. 99–119
33.66 I. N. Stranski, L. Krastanow: Zur Theorie der orientierten Ausscheidung von Ionenkristallen

aufeinander. In: *Sitzungsberichte der Akademie der Wissenschaften in Wien*, Math.-naturwiss. Klasse, Abt. IIB, Vol. 146 (1937) pp. 797–810, in German
33.67 V.A. Shchukin, N.N. Ledentsov, P.S. Kop'ev, D. Bimberg: Spontaneous ordering of arrays of coherent strained islands, Phys. Rev. Lett. **75**, 2968–2971 (1995)
33.68 L.G. Wang, P. Kratzer, M. Scheffler, N. Moll: Formation and stability of self-assembled coherent islands in highly mismatched heteroepitaxy, Phys. Rev. Lett. **82**, 4042–4045 (1999)
33.69 D. Leonard, K. Pond, P. M. Petroff: Critical layer thickness for self-assembled InAs islands on GaAs, Phys. Rev. B **50**, 11687–11692 (1994)
33.70 L. Goldstein, F. Glas, J.Y. Marzin, M.N. Charasse, G. Le Roux: Growth by molecular beam epitaxy and characterization of InAs/GaAs strained-layer superlattices, Appl. Phys. Lett. **47**, 1099–1101 (1985)
33.71 D. Leonard, M. Krishnamurthy, C.M. Reaves, S.P. Denbaars, P.M. Petroff: Direct formation of quantum-sized dots from uniform coherent islands of InGaAs on GaAs surfaces, Appl. Phys. Lett. **63**, 3203–3205 (1993)
33.72 F. Heinrichsdorff, A. Krost, N. Kirstaedter, M.-H. Mao, M. Grundmann, D. Bimberg, A.O. Kosogov, P. Werner: InAs/GaAs quantum dots grown by metalorganic chemical vapor deposition, Jpn. J. Appl. Phys. **36**, 4129–4133 (1997)
33.73 J. Oshinowo, M. Nishioka, S. Ishida, Y. Arakawa: Area density control of quantum-size InGaAs/Ga(Al)As dots by metalorganic chemical vapor deposition, Jpn. J. Appl. Phys. **33**, L1634–L1637 (1994)
33.74 F. Heinrichsdorff, A. Krost, M. Grundmann, D. Bimberg, F. Bertram, J. Christen, A. Kosogov, P. Werner: Self-organization phenomena of InGaAs/GaAs quantum dots grown by metalorganic chemical vapour deposition, J. Cryst. Growth **170**, 568–573 (1997)
33.75 A.A. El-Emawy, S. Birudavolu, P.S. Wong, Y.-B. Jiang, H. Xu, S. Huang, D.L. Huffaker: Formation trends in quantum dot growth using metalorganic chemical vapor deposition, J. Appl. Phys. **93**, 3529–3534 (2003)
33.76 I.N. Kaiander, F. Hopfer, T. Kettler, U.W. Pohl, D. Bimberg: Alternative precursor growth of quantum dot-based VCSELs and edge emitters for near infrared wavelengths, J. Cryst. Growth **272**, 154–160 (2004)
33.77 N.N. Ledentsov, J. Böhrer, D. Bimberg, I.V. Kochnev, M.V. Maximov, P.S. Kop'ev, I.Z. Alferov, A.O. Kosogov, S.S. Ruvimov, P. Werner, U. Gösele: Formation of coherent superdots using metal-organic chemical vapor deposition, Appl. Phys. Lett. **69**, 1095–1097 (1996)
33.78 K. Mukai, N. Ohtsuka, M. Sugawara: Controlled quantum confinement potentials in self-formed InGaAs quantum dots grown by atomic layer epitaxy technique, Jpn. J. Appl. Phys. **35**, L262–L265 (1996)
33.79 K. Pötschke, L. Müller-Kirsch, R. Heitz, R.L. Sellin, U.W. Pohl, D. Bimberg, N. Zakharov, P. Werner: Ripening of self-organized InAs quantum dots, Physica E **21**, 606–610 (2004)
33.80 A. Krost, F. Heinrichsdorff, D. Bimberg, A. Darhuber, G. Bauer: High-resolution x-ray diffraction of self-organized InGaAs/GaAs quantum dot structures, Appl. Phys. Lett. **68**, 785–787 (1996)
33.81 A. Lenz, R. Timm, H. Eisele, C. Hennig, S.K. Becker, R.L. Sellin, U.W. Pohl, D. Bimberg, M. Dähne: Reversed truncated cone composition distribution of $In_{0.8}Ga_{0.2}As$ quantum dots overgrown by an $In_{0.1}Ga_{0.9}As$ layer in a GaAs matrix, Appl. Phys. Lett. **81**, 5150–5152 (2002)
33.82 R.L. Sellin, I. Kaiander, D. Ouyang, T. Kettler, U.W. Pohl, D. Bimberg, N.D. Zakharov, P. Werner: Alternative-precursor MOCVD of self-organized InGaAs/GaAs quantum dots and quantum dot lasers, Appl. Phys. Lett. **82**, 841–843 (2003)
33.83 L. Höglund, E. Petrini, C. Asplund, H. Malm, J.Y. Andersson, P.O. Holtz: Optimising uniformity of InAs/(InGaAs)/GaAs quantum dots by metal organic vapor phase epitaxy, Appl. Surf. Sci. **252**, 5525–5529 (2006)
33.84 F. Heinrichsdorff, A. Krost, D. Bimberg, A.O. Kosogov, P. Werner: Self-organized defect free InAs/GaAs and InAs/InGaAs/GaAs quantum dots with high lateral density grown by MOCVD, Appl. Surf. Sci. **123/124**, 725–728 (1998)
33.85 F. Heinrichsdorff, M. Grundmann, O. Stier, A. Krost, D. Bimberg: Influence of In/Ga intermixing on the optical properties of InGaAs/GaAs quantum dots, J. Cryst. Growth **195**, 540–545 (1998)
33.86 R. Sellin, F. Heinrichsdorff, C. Ribbat, M. Grundmann, U.W. Pohl, D. Bimberg: Surface flattening during MOCVD of thin GaAs layers covering InGaAs quantum dots, J. Cryst. Growth **221**, 581–585 (2000)
33.87 U.W. Pohl, K. Pötschke, A. Schliwa, F. Guffarth, D. Bimberg, N.D. Zakharov, P. Werner, M.B. Lifshits, V.A. Shchukin, D.E. Jesson: Evolution of a multimodal distribution of self-organized InAs/GaAs quantum dots, Phys. Rev. B **72**, 245332 (2005)
33.88 U.W. Pohl: InAs/GaAs quantum qots with multimodal size distribution. In: *Self-assembled Quantum Dots*, ed. by Z.M. Wang (Springer, New York 2008) pp. 43–66
33.89 U.W. Pohl, K. Pötschke, M.B. Lifshits, V.A. Shchukin, D.E. Jesson, D. Bimberg: Self-organized formation of shell-like InAs/GaAs quantum dot ensembles, Appl. Surf. Sci. **252**, 5555–5558 (2006)
33.90 U.W. Pohl, K. Pötschke, A. Schliwa, M.B. Lifshits, V.A. Shchukin, D.E. Jesson, D. Bimberg: Formation and evolution of multimodal size-distributions

of InAs/GaAs quantum dots, Physica E **32**, 9–13 (2006)
33.91 U.W. Pohl, A. Schliwa, R. Seguin, S. Rodt, K. Pötschke, D. Bimberg: Size-tunable exchange interaction in InAs/GaAs quantum dots. In: *Advances in Solid State Physics*, Vol. 46, ed. by R. Haug (Springer, Berlin 2008) pp. 41–54
33.92 R. Timm, H. Eisele, A. Lenz, T.-Y. Kim, F. Streicher, K. Pötschke, U.W. Pohl, D. Bimberg, M. Dähne: Structure of InAs/GaAs quantum dots grown with Sb surfactant, Physica E **32**, 25–28 (2006)
33.93 Y. Arakawa, H. Sakaki: Multidimensional quantum well laser and temperature dependence of its threshold current, Appl. Phys. Lett. **40**, 939–941 (1982)
33.94 M. Asada, Y. Miyamoto, Y. Suematsu: Gain and the threshold of three-dimensional quantum-box lasers, IEEE J. Quantum Electron. **22**, 1915–1921 (1986)
33.95 N. Kirstaedter, N.N. Ledentsov, M. Grundmann, D. Bimberg, V.M. Ustinov, S.S. Ruvimov, M.V. Maximov, P.S. Kop'ev, I.Z. Alferov, U. Richter, P. Werner, U. Gösele, J. Heydenreich: Low threshold, large T_0 injection laser emission from (InGa)As quantum dots, Electron. Lett. **30**, 1416–1417 (1994)
33.96 N. Kirstaedter, O.G. Schmidt, N.N. Ledentsov, D. Bimberg, V.M. Ustinov, A.Y. Egorov, A.E. Zhukov, M.V. Maximov, P.S. Kop'ev, I.Z. Alferov: Gain and differential gain of single layer InAs/GaAs quantum dot injection lasers, Appl. Phys. Lett. **69**, 1226–1228 (1996)
33.97 F. Heinrichsdorff, M.-H. Mao, N. Kirstaedter, A. Krost, D. Bimberg, A.O. Kosogov, P. Werner: Room-temperature continuous-wave lasing from stacked InAs/GaAs quantum dots grown by metal-organic chemical vapor deposition, Appl. Phys. Lett. **71**, 22–24 (1997)
33.98 M.V. Maximov, I.V. Kochnev, Y.M. Shernyakov, S.V. Zaitsev, N.Y. Gordeev, A.F. Tsatsul'nikov, A.V. Sakharov, I.L. Krestnikov, P.S. Kop'ev, I.Z. Alferov, N.N. Ledentsov, D. Bimberg, A.O. Kosogov, P. Werner, U. Gösele: InGaAs/GaAs quantum dot lasers with ultrahigh characteristic temperature ($T_0 = 385$ K) grown by metal organic chemical vapour deposition, Jpn. J. Appl. Phys. **36**, 4221–4223 (1997)
33.99 R.L. Sellin, C. Ribbat, M. Grundmann, N.N. Ledentsov, D. Bimberg: Close-to-ideal device characteristics of high-power InGaAs/GaAs quantum dot lasers, Appl. Phys. Lett. **78**, 1207–1209 (2001)
33.100 L.F. Lester, A. Stintz, H. Li, T.C. Newell, E.A. Pease, B.A. Fuchs, K.J. Malloy: Optical characteristics of 1.24 μm InAs quantum-dot laser diodes, IEEE Photon. Technol. Lett. **11**, 931–933 (1999)
33.101 A. Strittmatter, T.D. Germann, T. Kettler, K. Posilovic, U.W. Pohl, D. Bimberg: Alternative precursor metal-organic chemical vapor deposition of InGaAs/GaAs quantum dot laser diodes with ultra-low threshold at 1.25 μm, Appl. Phys. Lett. **88**, 262104 (2006)
33.102 J. Tatebayashi, N. Hatori, M. Ishida, H. Ebe, M. Sugawara, Y. Arakawa, H. Sudo, A. Kuramata: 1.28 μm lasing from stacked InAs/GaAs quantum dots with low-temperature-grown AlGaAs cladding layer by metalorganic chemical vapor deposition, Appl. Phys. Lett. **86**, 053107 (2005)
33.103 N. Nuntawong, Y.C. Xin, S. Birudavolu, P.S. Wong, S. Huang, C.P. Hains, D.L. Huffaker: Quantum dot lasers based on a stacked and strain-compensated active region grown by metal-organic chemical vapor deposition, Appl. Phys. Lett. **86**, 193115 (2005)
33.104 S.M. Kim, Y. Wang, M. Keever, J.S. Harris: High-frequency modulation characteristics of 1.3 μm InGaAs quantum dot lasers, IEEE Photon. Technol. Lett. **16**, 377–379 (2004)
33.105 I.R. Sellers, H.Y. Liu, K.M. Groom, D.T. Childs, D. Robbins, T.J. Badcock, M. Hopkinson, D.J. Mowbray, M.S. Skolnick: 1.3 μm InAs/GaAs multilayer quantum-dot laser with extremely low room-temperature threshold current density, Electron. Lett. **40**, 1412–1413 (2004)
33.106 D. Giumard, M. Ishida, N. Hatori, Y. Nakata, H. Sudo, T. Yamamoto, M. Sugawara, Y. Arakawa: CW lasing at 1.35 μm from ten InAs-Sb:GaAs quantum-dot layers grown by metal-organic chemical vapor deposition, IEEE Photon. Technol. Lett. **20**, 827–829 (2008)

34. Formation of SiGe Heterostructures and Their Properties

Yasuhiro Shiraki, Akira Sakai

The Si/Ge system provides a lot of varieties of materials growth due to the lattice mismatch between Si and Ge. From the point of view of device applications, both pseudomorphic growth and strain-relaxed growth are important. Not only the layer growth but also dot formation is now attracting much attention from both the scientific community and for device applications. Comprehensive studies on the growth mechanisms have resulted in the development of novel formation techniques of SiGe heterostructures and enable us to implement strain effects in Si devices. It is obvious that the device applications largely depend on the material growth, particularly control of surface reaction and formation of dislocations and surface roughness that strongly affect device performances. Here we review the fabrication technology of SiGe heterostructures aiming at growth of high-quality materials. The relaxation of strain of SiGe buffer layers grown on Si substrates is discussed in detail, since many devices are formed on the strain-relaxed buffer layers that are sometimes called *virtual substrates*. Carbon incorporation and dot formation that are now studied to extend the possibilities of SiGe are discussed in this chapter too.

34.1 Background 1153
34.2 Band Structures of Si/Ge Heterostructures 1154
34.3 Growth Technologies 1156
34.3.1 Molecular-Beam Epitaxy 1156
34.3.2 Chemical Vapor Deposition 1157
34.4 Surface Segregation 1157
34.5 Critical Thickness 1161
34.6 Mechanism of Strain Relaxation 1163
34.7 Formation of Relaxed SiGe Layers 1165
34.7.1 Graded Buffer 1165
34.7.2 Low-Temperature Method 1166
34.7.3 Chemical-Mechanical Polishing Method 1167
34.7.4 Ion Implantation Method 1168
34.7.5 Ge Condensation Method 1169
34.7.6 Dislocation Engineering for Buffer Layers 1170
34.7.7 Formation of SiGeC Alloys 1172
34.8 Formation of Quantum Wells, Superlattices, and Quantum Wires 1173
34.9 Dot Formation 1177
34.10 Concluding Remarks and Future Prospects 1184
References 1184

34.1 Background

SiGe heterostructures have great potential to improve state-of-the-art Si devices, particularly very large-scale integrated circuits (VLSIs), and add such new functions as optics. Moreover, they provide a new scientific field of materials growth and characterization related to the lattice mismatch between Si and Ge. Band modification due to the strain coming from the lattice mismatch brings about the increase of the mobility of both electrons and holes. Since heterostructure bipolar transistors (HBTs) became commercially available, many people have become engaged in the research and development of field-effective transistor (FET)-type devices which have a much wider range of applications. SiGe-based optical devices including optical interconnection and optoelectronic integrated circuits (OEICs) will provide new possibilities to enhance the performance and functions of Si VLSIs. Heterostructures such as quantum wells and dots make it possible to real-

ize light-emitting devices even with indirect-bandgap materials. These fascinating applications obviously depend on the material growth, particularly the control of surface reaction and formation of dislocations and surface roughness, which strongly affect device performance.

In this chapter, we review the fabrication technology of SiGe heterostructures developed so far and discuss the physics and chemistry behind the process. The relaxation of strain of SiGe buffer layers grown on Si substrates is very important in this field, since many devices are formed on the strain-relaxed buffer layers which are sometimes called *virtual substrates*. The formation and properties of these layers are discussed here in detail. To extend the possibilities of SiGe, carbon incorporation is applied, which is also very interesting from the point of view of material growth. The large lattice mismatch in this material system causes the formation of misfit dislocations as well as dot formation. The latter is now one of the hottest topics in semiconductor physics and technology and will be discussed in this chapter too.

34.2 Band Structures of Si/Ge Heterostructures

Since there exists about 4.2% lattice mismatch between Si and Ge, the strain induced by the lattice mismatch modifies the band structures of SiGe layers [34.1–3]. The bandgap energy of SiGe layers grown on Si substrates is changed by the strain, as shown in Fig. 34.1 as a function of Ge composition. The bandgap is decreased from the bulk value under lateral compressive strain, and the lowest point of the conduction band of strained SiGe is the delta valley, as in Si crystals, over the whole composition range.

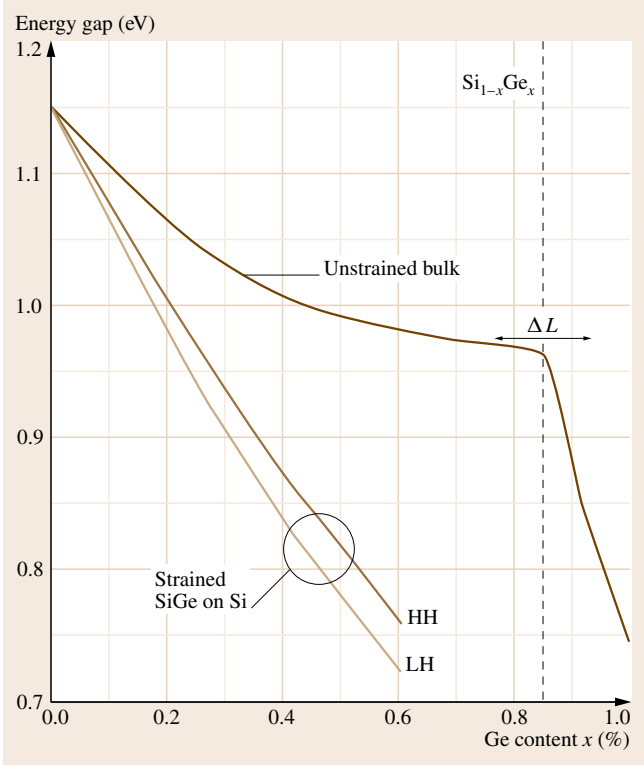


Fig. 34.1 Ge content dependence of the energy bandgap of strained SiGe grown on Si substrates. HH and LH represent bandgaps for heavy and light hole bands. That of unstrained SiGe is also shown as a reference

Figure 34.2 shows the change in the conduction and valence bands due to strain. The degenerate sixfold conduction bands are separated into two groups, that is, twofold degenerated bands and fourfold degenerated bands. The degenerated heavy-hole (HH) and light-hole (LH) bands are also separated, and the energy difference of the spin–orbit splitting band is changed. This change effectively modifies the transport properties of electrons and holes, providing opportunities to improve transistor performance. In particular, the mobility enhancement of both electrons and holes is highly attractive from the point of view of device applications such as complementary metal–oxide–semiconductor (MOS) FETs, i.e., CMOS circuits.

The band alignment at Si/Ge heterointerfaces is also significantly modified and type I and type II alignments are realized by changing the strain distribution, as illustrated schematically in Fig. 34.3, i.e., when $Si_{1-x}Ge_x$ ($x < 0.4$) layers are grown on Si substrates and are laterally compressed, the band alignment becomes type I and electrons and holes are confined to the same SiGe region. On the other hand, when Si layers are grown on unstrained SiGe layers and are under lateral tensile strain, type II alignment is realized and electrons and holes are confined sepa-

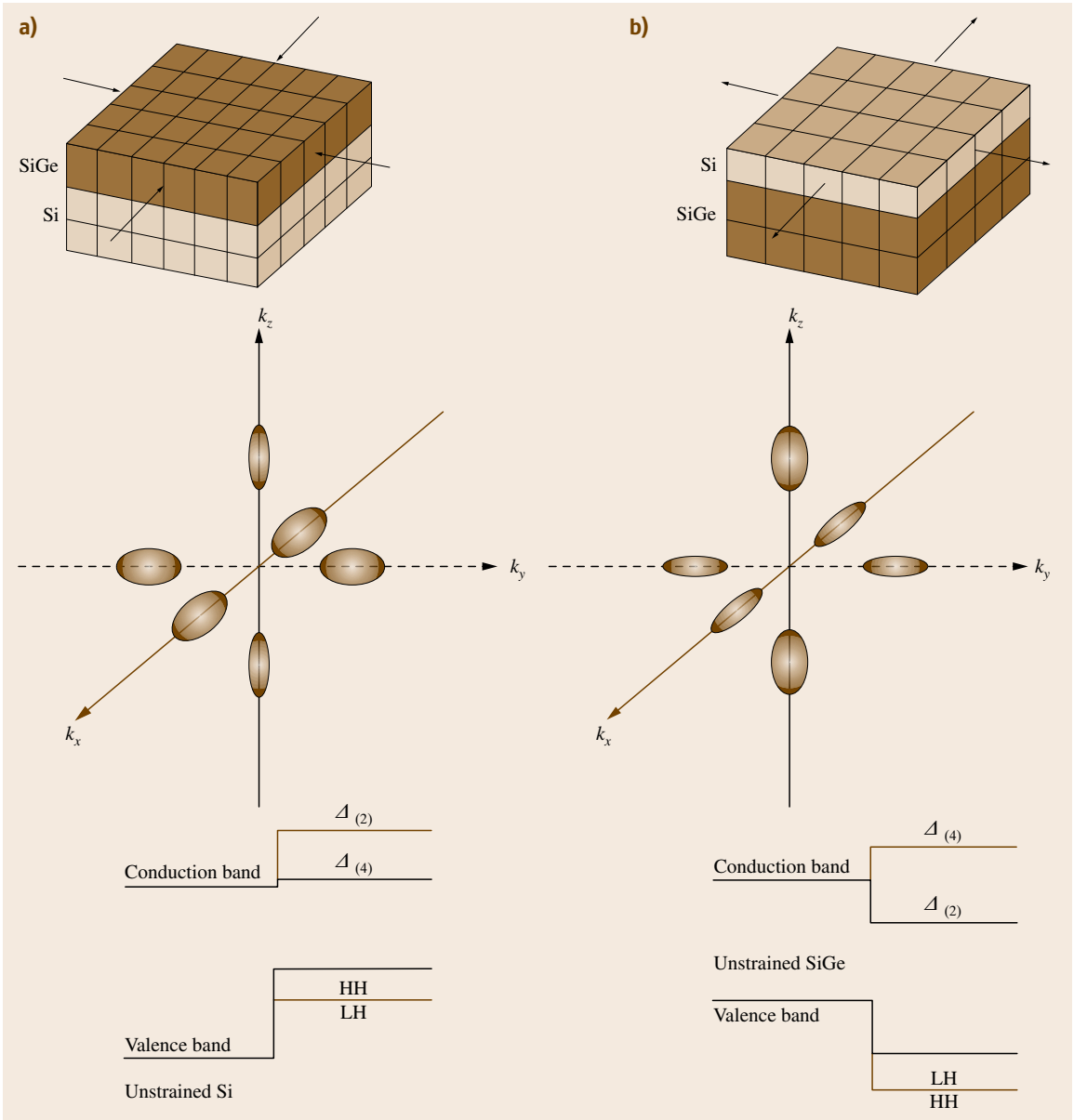


Fig. 34.2a,b Band modification of tensilely strained Si **(a)** and compressively strained SiGe **(b)**

rately. When Ge is laterally compressed by growth on SiGe layers, type II alignment also appears, as shown in the figure. Both band alignments are very useful from the point of view of device applications. Especially, since the band discontinuity at the conduction band of the former is very small, type II alignment is important when band engineering for electrons is intended.

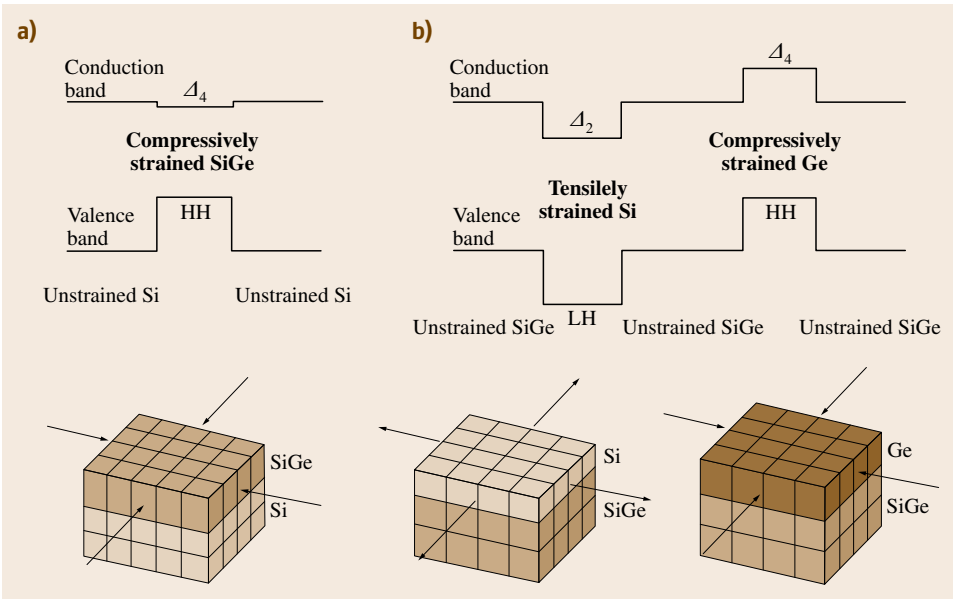


Fig. 34.3a,b Band alignment of Si/Ge heterostructures under various strains: **(a)** compressively strained SiGe on Si substrate (type I), **(b)** tensilely strained Si and compressively strained Ge on unstrained SiGe (type II)

34.3 Growth Technologies

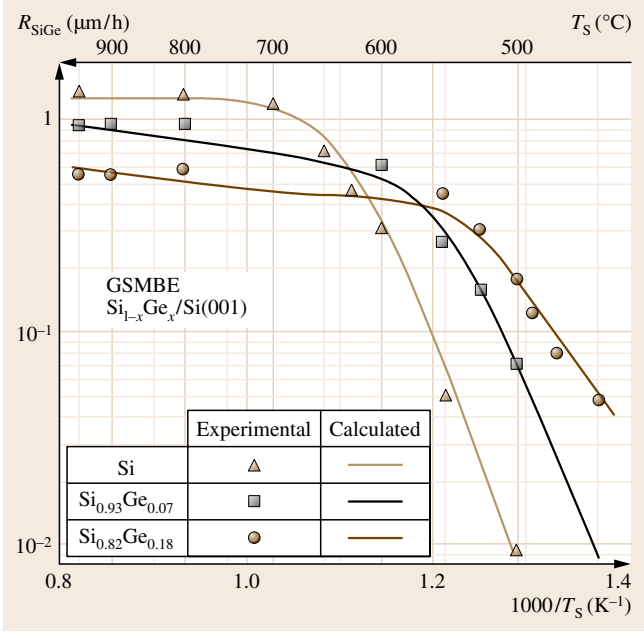


Fig. 34.4 Growth rate R_{SiGe} of GSMBE $Si_{1-x}Ge_x(001)$ layers with $x = 0$, 0.07, and 0.18 as a function of temperature T_s. The *solid lines* are calculated. For details, see [34.4]

34.3.1 Molecular-Beam Epitaxy

Molecular-beam epitaxy (MBE) is broadly used for depositing semiconductor and metallic materials to form thin films and multilayers. For the growth of Si films by MBE, a Si molecular beam is irradiated onto a clean surface of a Si substrate in a stainless-steel chamber in which the base pressure is reduced to the ultrahigh-vacuum (UHV) range, typically of the order 10^{-10} Torr.

An electron gun evaporator is usually used as a solid source of Si in order to achieve the vapor pressure required for practical growth. The evaporation of Ge is performed using a conventional Knudsen cell (K-cell), which is surrounded by liquid-nitrogen-cooled shrouds to condense unwanted evaporants and improve the vacuum in the sample region. The K-cell is also used for evaporating dopant materials: Ga and B for p-type doping, and Sb for n-type. Solid-source MBE is so simple and safe that it is also widely employed for fundamental research into thin-film growth mechanisms. Due to the reduced-pressure environment during growth, in situ and real-time diagnostic tools, such as reflection high-energy electron diffraction, scanning electron microscopy, and scanning tunneling microscopy, can be readily incorporated into the vacuum system to moni-

tor the dynamics occurring on the surface of a growing film.

On the other hand, growth is often achieved using gas sources of Si and Ge, such as SiH_4, Si_2H_6, and GeH_4, in the same type of UHV chambers as are used for solid-source MBE. An advantage of gas-source MBE is its ability for selective epitaxial growth on SiO_2-mask-patterned Si substrates [34.5,6]. The growth kinetics of epitaxy on clean Si surfaces has been extensively studied for SiH_4 [34.7,8], Si_2H_6 [34.4,9–13], GeH_4 [34.11,13–15], Ge_2H_6 [34.4,14], and organosilanes [34.16–18]. Typical trends of SiGe growth on Si in gas-source MBE can be seen in Fig. 34.4 [34.4], which shows the growth rate of SiGe layers on Si(001) substrates as a function of substrate temperature when using Si_2H_6 and GeH_4. In the low-temperature regime, surface-reaction-limited growth takes place and the growth rate increases with increasing Ge concentration. On the contrary, the growth rate saturates at high temperature, where impingement-flux-limited growth occurs, and decreases with increasing Ge concentration. These phenomena are also observed when using monohydride sources of SiH_4 and GeH_4, and can be well modeled in terms of several kinetic parameters such as the sticking probability of gas species at adsorption sites, hydrogen desorption, and Ge segregation [34.4,19,20].

34.3.2 Chemical Vapor Deposition

Chemical vapor deposition (CVD) is most frequently used for the growth of semiconductor, metal, and insulator films for device production. In general, either a hot- or cold-wall reactor is used, with pumping and sophisticated gas control systems which enable exact supply of reactive gases onto heated substrates to produce thin, epitaxial SiGe and SiGeC alloy films. CVD can ordinarily be categorized according to operating pressure into: UHV, low pressure (LP), reduced pressure (RP), and atmospheric pressure (AP) CVD.

UHV-CVD was firstly achieved by *Meyerson* et al. for growth of SiGe films on Si substrates using SiH_4 and GeH_4 [34.21]. They developed a hot-wall-type system consisting of a quartz reactor tube surrounded by an electrical heater with pumping systems to realize the UHV condition. Most commercially applicable processing is LP-CVD and RP-CVD, typically operating at pressures ranging from 1 Torr to a few tens of Torr. Gaseous sources of SiH_4, SiH_2Cl_2, SiH_3CH_3, GeH_4, CH_4, and $SiCH_6$ are usually used for growth of Si, SiGe, and SiGeC films, and PH_3 and B_2H_6 for doping [34.22–25].

As mentioned earlier for gas-source MBE, during CVD growth of SiGe layers on Si, dramatic acceleration of the growth rate was also observed with the introduction of GeH_4 into the Si source gas, as compared with that of Si layer growth alone [34.21–23,25–27]. This is mainly due to the lowering of the hydrogen desorption energy caused by the presence of Ge on the Si surface. Selective epitaxial growth of Si and SiGe on Si surfaces has also been widely studied in CVD. For growth at pressure of more than 10 Torr, besides SiH_2Cl_2 and GeH_4 as gaseous sources, HCl was effectively used to control deposition selectivity with respect to SiO_2-masked regions [34.28,29], whereas it was achieved without HCl in the case of very low pressure, less than 1 Torr [34.30,31]. Practical applications of selective epitaxial growth technology have mainly been the formation of SiGe channels for high-performance MOSFETs, elevated source/drain regions in shallow-junction electrode MOSFETs, and base regions of high-speed SiGe heterojunction bipolar transistors.

34.4 Surface Segregation

It is well known that the real heterointerface is not atomically flat and abrupt. There are several effects that deteriorate interface abruptness, of which surface segregation is considered to be the main one. Surface segregation is a reaction between impinging atoms and surface atoms of substrates, which then exchange positions to reduce the total energy of the system. This phenomenon was first recognized to be important when MBE layers were doped with some kinds of impurities [34.32,33]. It should be pointed out, however, that surface segregation takes place not only in cases of doping, but also during heterointerface formation. Ge and In are well known to segregate when forming Si-on-(Si)Ge and GaAs-on-In(Ga)As heterostructures, respectively.

Figure 34.5 shows a typical example of the surface segregation phenomenon [34.34]. This figure shows secondary-ion mass spectroscopy (SIMS) profiles for the case when Si atoms are deposited on Si substrate covered with submonolayer Sb atoms. It is seen that Sb atoms do not remain at the original position and that

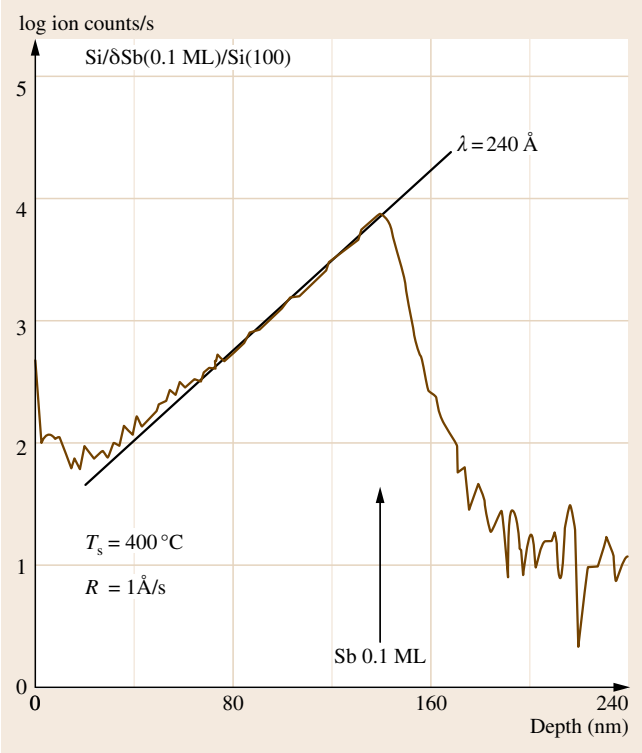


Fig. 34.5 SIMS profile of Sb atoms with 0.1 ML deposited on Si substrates measured after overgrowth of Si epitaxial layer. Due to surface segregation, the profile shows a tail with segregation length of $\lambda = 240$ Å towards the surface

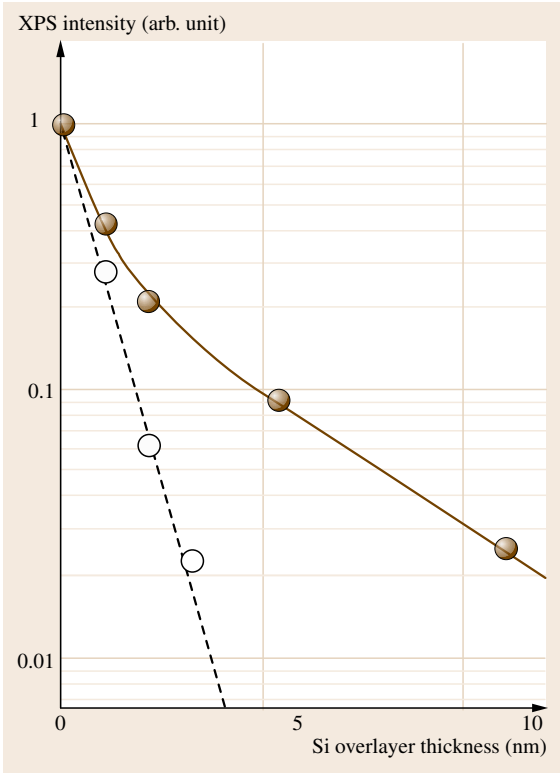


Fig. 34.6 XPS intensity of Ge atoms as a function of the thickness of overgrown Si layer. *Filled circles* are experimental results and the *dotted line* represents the calculated dependence without surface segregation. *Open circles* are for the case of surfactant-mediated growth

they show an exponential distribution towards the surface with a characteristic parameter of the segregation length.

Another example is seen in Fig. 34.6, which shows how surface segregation occurs when Si overlayer is grown on the Ge layers [34.35]. If Ge atoms sat on their original sites and site exchange between Ge and impinging Si did not occur, the x-ray photoemission (XP) intensity of Ge atoms should decrease exponentially with increasing Si overlayer thickness, as shown by the dashed line. It is, however, seen that the x-ray photoemission spectroscopy (XPS) intensity does not follow the exponential decay but is much stronger even after growth of Si layers with thickness of 10 nm, which is much larger than the escape depth of photoexcited electrons.

Surface segregation of Ge is also clearly seen in photoluminescence (PL) of SiGe/Si quantum wells, the details of which will be described later. Due to the quantum confinement effect, the PL peak positions shift to higher energies with decreasing well width. The solid symbols in Fig. 34.7 show no-phonon (NP) and transverse optic (TO) phonon replica peak positions as a function of well width [34.34]. The dashed lines in the figure show the well width dependence of the peaks calculated based on the square well, that is, when surface segregation does not take place. It is seen that the peaks shift to higher energies from the expected positions. This is because the quantum levels are lifted by the deformation of the well shape due to Ge surface segregation.

It is confirmed [34.36] that only the topmost Ge atoms are mainly incorporated in the segregation and that we may neglect underlying atoms to a first approximation. The surface segregation phenomenon is therefore described in terms of a two-state exchange model where only exchange of atoms in surface and subsurface states is taken into account, as shown in

Fig. 34.8 [34.32, 37]. The exchange is described by the following rate equation

$$\frac{dn_1}{dt} = -pn_1 + qn_2 \,, \tag{34.1}$$

$$\frac{dn_2}{dt} = -qn_2 + pn_1 \,, \tag{34.2}$$

$$n_1 + n_2 = n_0 \,,$$

where $p = p_0 \exp(-E_a/k_B T)$, $q = q_0 \exp[-(E_a + E_b)/k_B T]$, E_a is a potential for atoms to jump into the surface state from the underlying site and is less than the bulk thermal diffusion potential, E_b is the energy gain for the surface segregation and a measure of the segregation strength, and p_0 and q_0 are pre-exponential factors corresponding to the attempt frequency of atomic jumps and that may be considered to be on the order of the lattice vibration of 10^{12}–10^{13} s^{-1}. Under the equilibrium condition, the equation can be numerically solved as

$$n_2(t) = n_0 \left[1 - \frac{1}{p+q}\left(q + e^{(-p+q)t}\right)\right] . \tag{34.3}$$

This solution shows the exponential distribution of atoms segregating towards the surface, well representing the results of SIMS experiments. Figure 34.9 shows a logarithmic plot of segregated atoms as a function of inverse temperature, in which the line labeled *equilibrium segregation* corresponds to the above solution. It is noted that segregation becomes smaller as the temperature increases. At first sight, this contradicts the observation that segregation is enhanced with increasing temperature. However, it should be pointed out that the experiments are not conducted under conditions of thermal equilibrium but under kinetically limited con-

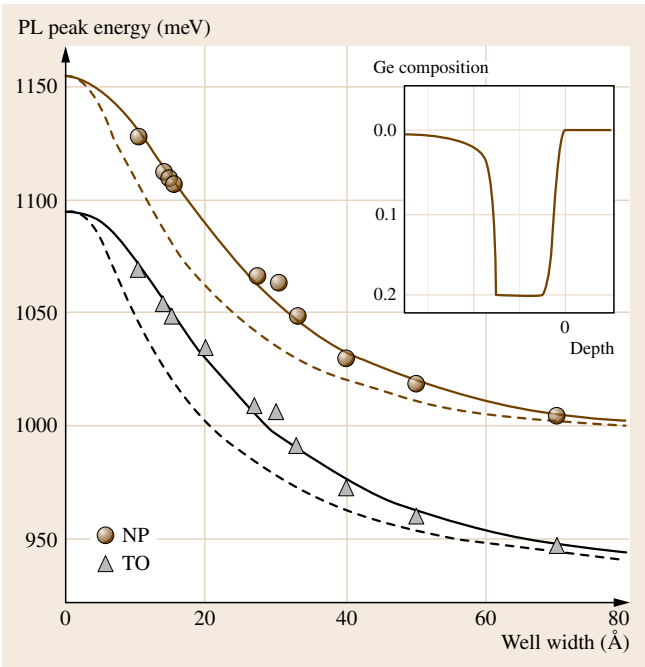


Fig. 34.7 Well width dependence of NP and TO peaks of PL from SiGe/Si quantum wells. *Solid symbols* are experimental results and the *dotted lines* represent the calculated well width dependence without surface segregation. The *solid lines* represent the dependence obtained by taking into account surface segregation in the two-state exchange model. The *inset* shows a schematic of the well shape distorted by surface segregation

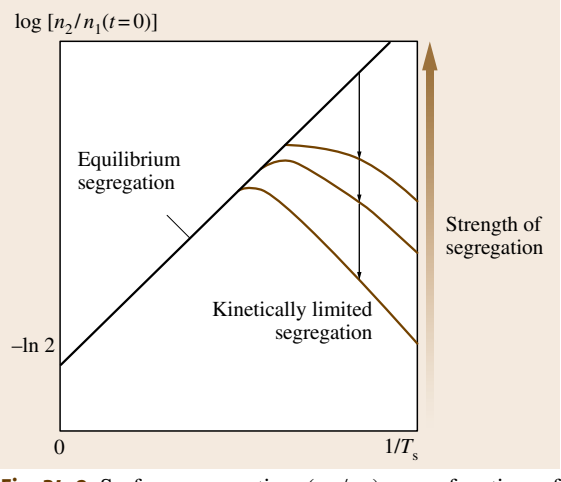


Fig. 34.9 Surface segregation (n_2/n_1) as a function of inverse temperature. Equilibrium segregation occurs under thermal equilibrium conditions and kinetically limited segregation occurs in the case of real crystal growth, particularly at low temperatures

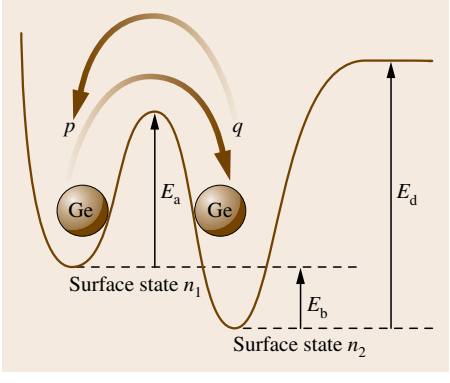


Fig. 34.8 Energy diagram of a two-state exchange model. E_a is the potential for atoms to jump from the subsurface position, and E_b is the energy gain for surface segregation. E_d is the desorption energy

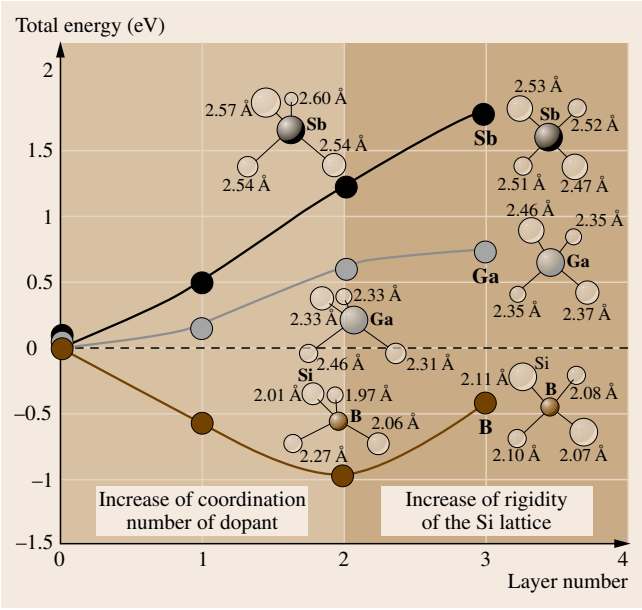


Fig. 34.10 Total energy of the system of Si with impurity atoms as a function of impurity position

ditions, that is, surface atoms are embedded before they reach the equilibrium as the crystal growth proceeds continuously. This situation is called kinetically limited segregation, as shown in Fig. 34.9, and segregation is largely suppressed at lower temperatures.

The solid line in Fig. 34.7 shows the well width dependence of quantum well (QW) edge emissions calculated by taking into account the distortion of the well shape based on the two-state exchange model. It is seen that the energy position is sufficiently represented by the calculation, indicating that the model describes the surface segregation well, even though it is phenomenological.

More microscopically, the local strain energy calculation is very useful to understand the segregation [34.38]. Figure 34.10 shows the total energy of the atomic arrangement in Si as a function of foreign atom site. In the case of Sb in Si, the energy is seen to increase as the site becomes deeper in the Si substrate. This indicates that Sb tends to move to the surface region to reduce the total energy, resulting in surface segregation. Interestingly, the energy minimum of B atoms is the third layer, and therefore B always tends to sit in this layer as crystal growth proceeds.

The trend for surface segregation can be predicted by considering the bond energy with Si atoms. The bond energy of Sb–Si, Ga–Si, and Ge–Si are smaller than that of Si itself. These atoms are, therefore, rejected from the Si matrix to form strong chemical bonds and show surface segregation.

To avoid segregation and obtain abrupt heterointerfaces, low-temperature growth has been thought to be essential. However, low-temperature growth deteriorates crystal quality. Surfactant-mediated growth (SMG) is now well known to avoid interface smearing and obtain abrupt interfaces. In this method, a small amount of foreign atoms that show strong segregation, sometimes called segregants, are deposited before formation of heterointerfaces. For this purpose, As, Sb, Ga, Bi, and H are well known to suppress Ge segregation for the formation of Si/SiGe/Si structures. The principle of this method is as follows: before the Si overlayer of the Si/SiGe/Si heterostructure is grown, the SiGe layer is covered with strong segregants such as Sb so that Ge is no longer the topmost surface atom. When Si atoms are deposited on the Sb-covered Ge surface, the position of Sb becomes the subsurface site and Sb exchanges position with impinging Si atoms sitting at surface sites. However, site exchange between Si and Ge atoms does not follow. This is because Ge atoms do not occupy surface sites or subsurface site, that is, they are in bulk sites once Si is deposited. Between atoms in bulk sites, there is a high potential barrier that they cannot climb at the growth temperature, which is much lower than thermal diffusion temperatures, and therefore they do not exchange their positions.

In Fig. 34.6, the open circles show the XPS intensity of Ge atoms when 0.75 ML of Sb atoms are introduced as suppressors of segregation at the heterointerface before Si layers are overgrown [34.35]. It is seen that the XPS intensity follows the exponential decay and that surface segregation is effectively suppressed. Suppression of Ge segregation is also seen in SIMS profiles of Si/Ge superlattices and, once Sb atoms are deposited on Ge layers, the long tail towards the surface is seen to disappear [34.35]. The amount of Sb atoms required to suppress surface segregation is then an important question, and it was found that the effect becomes pronounced at around 0.5 ML and is optimized at 0.75 ML [34.34].

Gas-source MBE (GSMBE) is another important example of SMG, or pseudo-SMG where atomic hydrogen decomposed from hydrogenated gases such as Si_2H_6 and GeH_4 acts as a suppressor of Ge segregation. The well width dependence of the PL peak positions of QWs grown by GSMBE is found to coincide well with the square-well potential calculation, indicating that GSMBE provides heterostructures with-

out significant interfacial smearing [34.39]. It is also reported that Ge segregation is almost absent in UHV-CVD growth, in which hydrogen and/or hydrogenated compounds are speculated to act as suppressors [34.40]. It is, however, noted that hydrogen desorbs at high temperatures such as those used in GSMBE. Therefore, hydrogen is thought to have a relatively long residence time on Si surfaces before desorption and can stay to act as a surfactant during a period of monolayer growth.

34.5 Critical Thickness

In epitaxial growth of overlayers that have a lattice parameter slightly different from that of the substrate, pseudomorphic (or coherent or commensurate) overlayers are formed when the thickness is not large. Since, in this case, the in-plane lattice constant of the pseudomorphic film is equivalent to that of the substrate, the film is elastically deformed according to Poisson's ratio to accommodate the lattice mismatch between the film and the substrate. The elastic energy due to the strain in the film increases linearly with increasing film thickness. However, there is a film thickness beyond which the introduction of misfit dislocations into the film becomes energetically favorable even though the energy increase due to the self-energy of the dislocations is taken into account. This minimum value of the film thickness is called the *critical thickness*.

Early theoretical works elucidating the critical thickness in lattice-mismatched heteroepitaxial growth were done by *Frank* [34.41–43] and *van der Merwe* [34.44, 45]. They gave a fundamental approach for predicting the critical thickness on the basis of the elastic energy and dislocation energy in strained epitaxial systems. The model proposed by *van der Merwe* was modified to be applied more practically to the diamond lattice in low-lattice-mismatch systems such as SiGe/Si [34.46–48]. Minimization of total energy given as a sum of the elastic energy stored in a homogeneously strained film and the energy associated with the misfit dislocations leads to the following equation for the critical thickness h_c

$$h_c = \frac{b}{8\pi(1+\nu)f}\left[1+\ln\left(\frac{2h_c}{r}\right)\right] . \qquad (34.4)$$

In (34.4), b is the magnitude of the active component of the misfit dislocation Burgers vector, ν is Poisson's ratio, and r is the inner cutoff radius of the misfit dislocation. Using the spacing between dislocations d and the elastic strain ε, the misfit of the overlayer with the substrate f is defined to be

$$f = |\varepsilon| + \frac{b}{d} . \qquad (34.5)$$

In (34.5), when $d \to \infty$, $\varepsilon = f$. A more detailed expression including numerical values associated with a SiGe/Si(001) system is also given in [34.48]. On the other hand, a simple and helpful model was developed by *Matthews* and *Blakeslee* and has often been used for interpreting various types of lattice-mismatched epitaxial systems [34.49–51]. Contrary to the aforementioned energy balance model, they considered the balance of forces exerted on a propagated dislocation with misfit and threading segments in the strained film, as shown in Fig. 34.11. The physical concept of the derivation of critical thickness in the force balance model is essentially the same as that in the energy balance model. The critical thickness in this case is given as

$$h_c = \frac{b(1-\nu\cos^2\alpha)}{8\pi(1+\nu)f\cos\theta}\left[1+\ln\left(\frac{h_c}{b}\right)\right] , \qquad (34.6)$$

where α is the angle between the misfit dislocation line and its Burgers vector, and θ is the angle between the misfit dislocation Burgers vector and a line in the interface drawn perpendicular to the dislocation line direction.

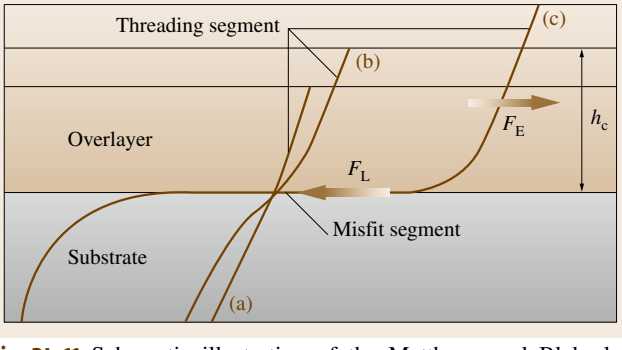


Fig. 34.11 Schematic illustration of the Matthews and Blakeslee model of critical thickness. A pre-existing threading dislocation in a coherent interface (a), critical interface (b), and incoherent interface (c). Critical thickness h_c is determined by the equality of the force exerted in the dislocation line by misfit stress F_E and the line tension in the dislocation line F_L

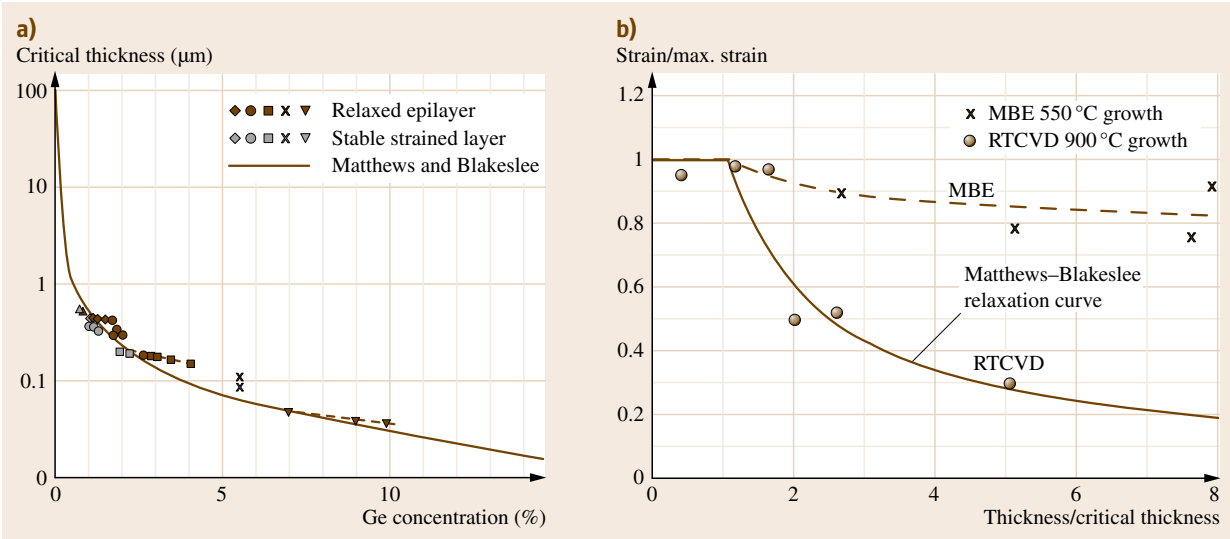


Fig. 34.12 **(a)** Experimentally determined critical values for thickness and composition obtained from annealed SiGe samples where the thickness and composition were laterally graded. An abrupt transition from a stable strained layer to a relaxed epilayer was detected to give a value of critical thickness at a certain composition. The *solid curve* is based on the Matthews and Blakeslee model, showing critical thickness as strain converted to Ge concentration in SiGe layers [34.56]. **(b)** Strain in the SiGe film divided by the theoretical maximum strain for that film versus film thickness divided by the critical thickness for that film grown by RTCVD. For comparison, the data of MBE-grown SiGe films [34.52] are also shown [34.57]

The discrepancy between theoretically predicted values and experimentally observed ones for the critical thickness has been reported in early works of molecular-beam epitaxy (MBE) of SiGe films on Si(001) substrates [34.52]. To explain the considerably larger critical thicknesses observed experimentally for SiGe films grown by MBE at 550 °C, *People* and *Bean* [34.53] firstly attempted to derive an expression of the critical thickness by taking into account the extra energy required for the generation of misfit dislocations. They obtained the critical thickness by setting the area density of strain energy associated with a film equal to the energy density of dislocations. Although this model showed quite good agreement with the experimental values of the critical thickness depending on the Ge content in the film, their interpretation is rather ad hoc, and the estimation of dislocation core energy is opposed to the elasticity treatment which is ordinarily employed for conventional semiconductor materials. On the other hand, it was pointed out that finite experimental resolution for detecting dislocation introduction in x-ray diffraction measurements for lattice parameters led to spurious results, and the critical thickness obtained by high-resolution measurements, such as stimulated-emission characteristics and photoluminescence, approaches the equilibrium theory proposed by *Matthews* and *Blakeslee* [34.54]. High-sensitivity detection for misfit dislocations was also performed using electron-beam-induced current imaging in scanning electron microscopy (SEM) and transmission electron microscopy (TEM) [34.55]. The obtained critical thicknesses were revealed to be less than those determined in [34.52], although the SiGe films were grown under similar conditions.

For another dominant reason causing the discrepancy between the theory and experiment, nonequilibrium growth conditions for strain-relaxed SiGe films seemed to be responsible. High-temperature (900 °C) postgrowth annealing for SiGe films grown at 500–600 °C [34.56, 58] and rapid-thermal chemical vapor deposition (RTCVD) at 900 °C [34.57] led to excellent agreement of the experimental critical thickness with the predicted values of the Matthews and Blakeslee model, as shown in Fig. 34.12a,b, respectively. In general, the kinetic barrier for the formation or motion of misfit dislocations gives rise to nonequilibrium behavior of the critical thickness. This kinetic effect on strain-relaxation behavior was the focus of the early stage of development of the Matthews and Blakeslee theory, in which *Matthews* et al. proposed that Peierls stress

results in a friction force on the dislocation [34.59]. Contrary to this model, assuming only a velocity-dependent friction force, a model taking the static Peierls barrier into account was developed [34.60]. Other models including precise expression of arrays of misfit dislocations [34.61] and surface-relaxation effects [34.62] have been proposed so far. As mentioned above, prediction of the critical thickness requires the consideration of kinetic effects relevant to practical growth conditions, such as growth temperature, initial dislocation density, and growth rate. This simultaneously means that strain relaxation of SiGe/Si(001) systems is strongly influenced by the kinetics, which will be explained in more detail in the next section.

34.6 Mechanism of Strain Relaxation

SiGe alloy films epitaxially grown on Si substrates have a maximum lattice mismatch of approximately 4.2% (when the film is pure Ge). This lattice mismatch induces strain mainly into the film, and elastic strain energy is accumulated with increasing film thickness. The relief of the elastic energy in the strained film occurs mainly through two mechanisms: elastic deformation accompanying surface evolution of the film, and plastic deformation with the introduction of misfit dislocations. In this section, to explain strain-relaxation mechanisms in SiGe/Si systems, we first mention the former process in which surface perturbation mainly contributes to the reduction of elastic energy and then the latter process including nucleation, propagation, and reaction/multiplication of dislocations.

It is well known that Ge grows on Si in Stranski–Krastanov mode in which layer-by-layer growth is followed by islanding of Ge. In practice, at the initial growth stage of Ge on Si(001) surfaces, dislocation introduction is preceded by the formation of dislocation-free (coherently strained) Ge islands [34.63] and/or *hut clusters* [34.64], i. e., {501}-faceted Ge islands, each of which was revealed by TEM and scanning tunneling microscopy (STM), respectively, and has a slightly different shape with respect to each other. Further TEM analysis revealed the hut cluster to be also a coherently strained Ge island and clarified the interplay between surface morphological variation and defect introduction concomitant with Ge islanding, which is critically dependent on the growth temperature [34.65, 66].

On the other hand, islanding phenomena occurring in SiGe/Si systems can be interpreted as strain relaxation based on Asaro–Tiller–Grinfeld instability in elastically stressed solids [34.67, 68]. The formation of perturbation on the stressed solid surface significantly reduces the stored elastic energy at the peak of perturbation but, at the same time, costs additional surface energy. Thus, the critical wavelength of the surface undulation can be drawn, which determines the stability of perturbation. The critical wavelength λ_c for a semi-infinite uniaxially stressed solid with a sinusoidal surface profile was derived as follows [34.69]

$$\lambda_c = \frac{\pi M \gamma}{\sigma^2} , \tag{34.7}$$

where M is the elastic modulus appropriate to the local surface orientation, γ is the surface tension, and σ is the mean stress in the solid. A more realistic model for the morphological instability of both growing and static epitaxially strained films was further developed, including differences in lattice parameters and elastic constants between the films and the substrate [34.70]. For experimental studies, wafer curvature measurements [34.71], direct observation using scanning transmission electron microscopy (STEM) [34.62], and STM [34.72] have been effectively applied to analyze the strain-related morphological evolution in SiGe/Si(001).

Surface evolution for elastic strain relaxation is generally followed by the introduction of misfit dislocations. A model describing total processes of strain relaxation in SiGe/Si systems was proposed by *Dodson* and *Tsao* [34.73]. On the basis of a phenomenological model of the dislocation dynamics and the plastic flow in bulk diamond-lattice semiconductors [34.74], they established a kinetic model of strain relaxation in which propagation and multiplication of pre-existing dislocations were treated systematically. In their model, a differential equation for the time-dependent degree of strain relaxation $\chi(t)$ is expressed as

$$\frac{d\chi(t)}{dt} = C\mu^2 \left[f_0 - \chi(t) - s(h) \right]^2 \left[\chi(t) + \chi_0 \right] . \tag{34.8}$$

In this equation, f_0 is the misfit, μ is the shear modulus, $s(h)$ is the thickness-dependent homogeneous strain retained by the overlayer, and χ_0 is the dislocation source density. The choice of appropriate C and χ_0 values

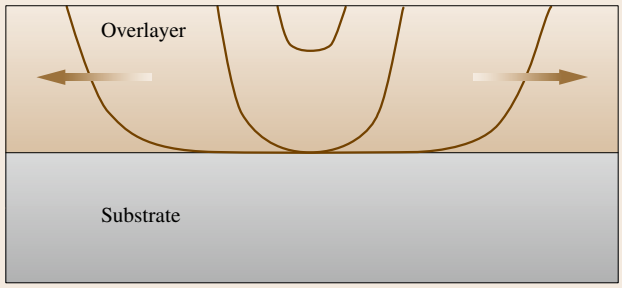
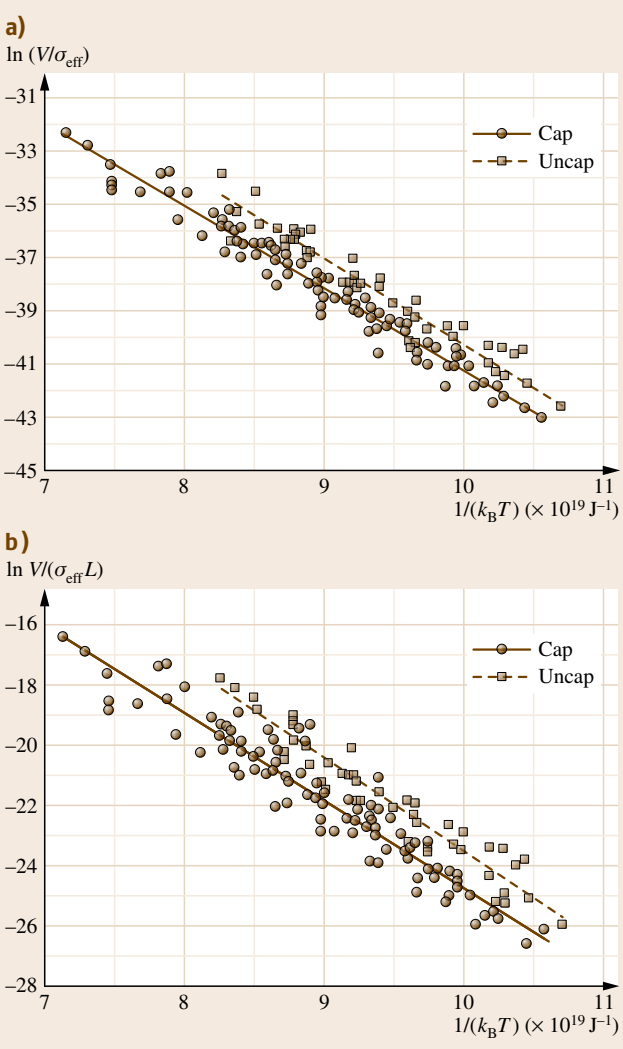


Fig. 34.13 Schematic illustration of nucleation and growth of a dislocation half-loop. Each threading part moves toward the edges of the sample, leaving a misfit segment at the interface, when the film thickness exceeds the critical thickness ◀

as fitting parameters resulted in remarkable agreement with the previous experimental results [34.52, 75]. According to the Dodson and Tsao model, even though the film thickness exceeds the critical thickness, both the effective stress exerted on the overlayer and the density of dislocations available for multiplication are small and this leads to initial slow relaxation of strain. Later on, the strain relaxation proceeds with the increase in the dislocation density exponentially with time, but then becomes sluggish due to the reduction of effective stress accompanied by the strain relaxation. They also explained when predicting the critical thickness that the initial slow relaxation combined with the later sluggish relaxation allows the growth of overlayers with a thickness considerably larger than the equilibrium critical thickness. Quantitative data obtained from in situ TEM analysis were also systematically applied to develop a model of strain relaxation including the effects of dislocation nucleation, propagation, and interaction [34.76].



Although detailed mechanisms for misfit dislocation nucleation in SiGe/Si systems are still controversial, the phenomena generally obey either heterogeneous or homogeneous nucleation mechanism. Earlier works looked for the source of heterogeneous nucleation, such as SiC precipitates formed by incomplete substrate cleaning [34.77], $a/4\langle 114\rangle$ stacking fault regions [34.78, 79], surface half-loops, and faulted dislocation loops associated with metallic contamination [34.80, 81]. However, all of these heterogeneous sites appear dependently of the growth conditions and cannot account for the strain-relaxation behavior occurring in the initial stage because of their relatively low density. An activation barrier for homogeneous nucleation of surface dislocation half-loops was calculated for SiGe/Si(001) systems and was shown to be sensitively influenced by Ge contents, i. e., strain, and the dislocation core energy parameter [34.82]. A con-

Fig. 34.14a,b Dislocation velocities in capped and uncapped (Si)/$Si_{1-x}Ge_x$/Si(001) heterostructures measured by TEM. The measured velocities V are normalized to equivalent velocities in pure Si by assuming a linear interpolation of the activation energy of Si and Ge, i. e., $E_{\mathrm{a}} = (2.2–0.6x)$ eV. The velocities are also normalized to the effective stress σ_{eff} **(a)** and the product $\sigma_{\mathrm{eff}}L$, where L is the length of the propagated dislocation line **(b)** ◀

sensus of qualitative arguments was reached in which homogeneous nucleation of half-loops is physically plausible at moderate and high mismatches, while heterogeneous nucleation relevant to a neighboring defect operates at low mismatch [34.83]. Furthermore, the aforementioned perturbation formed on the film surface plays a crucial role in dislocation nucleation. On the basis of STEM images showing SiGe cusp profiles, the condition that the local stress concentration at the surface cusp overcomes the nucleation barrier of half-loop dislocations was quantitatively discussed and the critical geometry for barrierless nucleation of 60° half-loops was derived [34.82, 84]. Similar analysis has been carried out for interfacial prismatic dislocation loops resulting from subnanometer-sized Ge-rich clusters as an efficient nucleation source [34.85].

As shown in Fig. 34.13, once half-loops nucleate on the film surface, they grow until they reach the substrate–overlayer interface. Then, each threading part of the dislocation moves toward the edges of the sample, leaving a misfit dislocation in the plane of the interface. The kinetics of such misfit dislocation propagation has been extensively investigated by two categorized methods: either direct observation of dislocation velocities by TEM or measurements of maximum length of dislocations by Nomarski interference microscopy of defect-etched surfaces [34.86–90]. These studies, irrespective of the difference in the measurements, show reasonable quantitative agreement with each other. In principle, the velocity of dislocation propagation V is of the form of

$$V = V_0 \sigma_{\mathrm{eff}} \exp\left(-\frac{E_\mathrm{a}}{k_\mathrm{B}T}\right) , \tag{34.9}$$

where V_0 is a prefactor containing an attempt frequency, σ_{eff} is the effective stress acting on the threading dislocation arm, and E_a is an activation energy of dislocation glide [34.73]. Systematic measurements of dislocation velocities were achieved using in situ TEM [34.91]. Assuming a linear interpolation of the activation energies of Si and Ge, i. e., $(2.2–0.6x)$ eV, where x denotes the Ge content in $Si_{1-x}Ge_x$, the measured velocity was normalized to an equivalent velocity under a stress of 1 Pa in pure Si. Consequently, fairly good correlation is seen between samples across all ranges of epilayer composition and thickness (Fig. 34.14).

In the final stage of strain relaxation, dislocation multiplication events are dominant. The Hagen–Strunk source [34.92], which was first observed in orthogonal dislocation configurations in Ge/GaAs systems, is commonly cited for dislocation multiplication. However, later on, theoretical and experimental analyses showed that it seems an unlikely source [34.93, 94]. At present, several multiplication sources are proposed, such as so-called modified Frank–Read sources [34.95], cross-slip and pinning [34.96], and cross-slip for branch formation [34.97]; the latter two events are also followed by Frank–Read-type dislocation multiplication mechanisms.

34.7 Formation of Relaxed SiGe Layers

34.7.1 Graded Buffer

There are various types of strain-relaxed SiGe buffer layers developed so far for strained-layer channels in metal–oxide–semiconductor field-effect transistors (MOSFETs) and modulation-doped field-effect transistors (MODFETs). A compositionally graded SiGe buffer layer is most successfully applicable to such devices [34.98–102]. In the formation procedure, the Ge concentration x in the $Si_{1-x}Ge_x$ alloy gradually increases with increasing film thickness. The in-depth composition profile of Ge is generally set to be either linear or step-like. Since this structure is considered to be the sum of low-mismatch interfaces, misfit dislocations are successively introduced during growth, resulting in total strain relaxation of the film. Contrary to SiGe films without a gradient, since each atomic plane tends to have its own equilibrium lattice parameter, the dislocations accommodating the lattice parameter difference between the substrate and the top layer are distributed over the thickness of the graded region. This specific configuration of dislocations results in a much lower density of threading dislocations than that of the constant-composition film. Typically, a threading dislocation density on the order of $10^5–10^7\,\mathrm{cm}^{-2}$ is obtained, depending on the grading rate and final Ge concentration in the top layer. One of the reasons for the absence of threading dislocations in the top layer is high dislocation velocities at which the dislocation can bypass pinning points such as a site of dislocation intersection [34.95]. Since all dislocations are not attracted to only a single interface in the compositionally graded film, there is great freedom for a dislocation to move

onto another (001) plane reaching the edges of the wafers.

However, severe surface roughness, referred to as a cross-hatch pattern, has been pointed out from the beginning of the research on SiGe graded buffers [34.103–107]. *Fitzgerald* et al. have shown that such cross-hatch patterns are related to the strain field in the epilayer caused by an inhomogeneous distribution of misfit dislocations [34.106]. According to their calculation, the critical thickness h_{gc} for the introduction of dislocations into a graded layer is

$$h_{gc}^2 = \frac{3\mu b}{2\pi Y C_f (1-\nu)} \left(1-\frac{\nu}{4}\right) \left[1+\ln\left(\frac{h_{gc}}{b}\right)\right] , \quad (34.10)$$

where C_f is the grading rate and Y is the Young's modulus. This formula clearly explains that larger grading rate leads to smaller h_{gc}, meaning that dislocations are introduced much closer to the surface and more strongly affect the evolution of the surface. On the other hand, a different mechanism was proposed, by which surface steps arising from the single and multiple 60° dislocations at the film–substrate interface directly influence the surface morphology of films [34.108]. Note that these two cases employ different growth conditions: a steeper (slower) grading rate of 50% Ge μm^{-1} (10% Ge μm^{-1}) and a lower (higher) growth temperature of 560 °C (900 °C) for the latter (former). Recently, a simulation study was performed to model the development of cross-hatch patterning on the basis of a combination of dislocation-assisted strain relaxation and surface steps during growth [34.109].

In graded buffers, there is a strong correlation between surface roughening and residual threading dislocation density. In general, the higher the final Ge concentration, the higher the threading dislocation density, despite the same grading rate. This results from a mechanism by which the combination of strain fields of underlying multiple misfit dislocations and the resultant surface roughness blocks the motion of a propagating threading dislocation, leading to dislocation pileups [34.110]. More recently, dislocation glide and blocking kinetics in compositionally graded SiGe buffers grown by UHV-CVD were reported and the dislocation density on the order of 10^4 cm^{-2} for 30% Ge was achieved at relatively high growth temperature of 900 °C [34.111].

In principle, graded buffers with low threading dislocation density require slow grading rate, which results in large thickness of the films and requires from time and material. Furthermore, poor thermal conductivity relevant to included Ge severely affects the performance of devices fabricated on graded buffers. Thus several techniques other than growth of graded buffers have also been attempted for strain-relaxed SiGe buffers on Si(001) substrates.

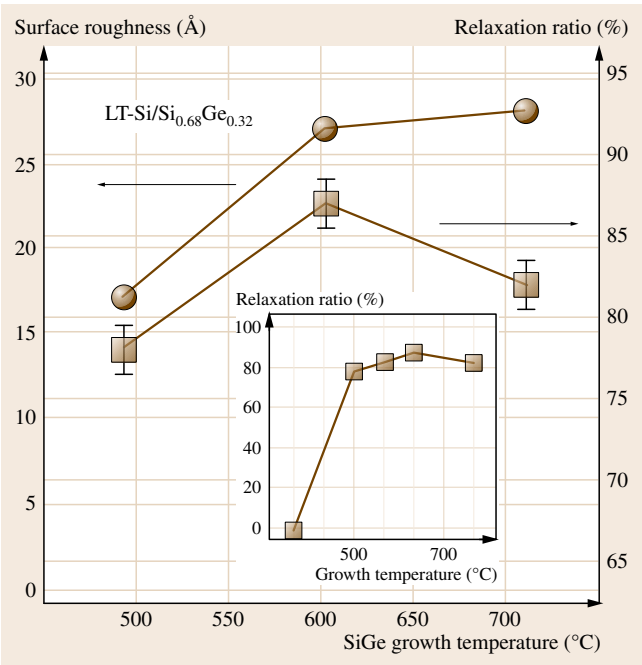


Fig. 34.15 Growth temperature dependence of surface roughness and relaxation ratio of $Si_{0.68}Ge_{0.32}$ buffer layers grown on low-temperature-grown Si (LT-Si). The inset shows relaxation ratio in the wider temperature range

34.7.2 Low-Temperature Method

To reduce the layer thickness and improve the crystal quality of buffer layers, the low-temperature (LT) method was proposed [34.112, 113]. In this method, before growing relatively thin SiGe buffer layers (about 500 nm), a thin Si layer, e.g., 50 nm thick, is grown at a low temperature such as 400 °C. Although the thickness of the SiGe buffer is much smaller than that required for the graded buffer method, it was found that the surface roughness is much better and the dislocation density is much smaller than those of the graded buffer method, while the relaxation ratio is almost equal, more than 80% [34.114]. That is, when the Ge composition is 30%, the surface roughness is about 1.5 nm and the dislocation density is around 10^5 cm^{-2}, while they are more than 10 nm and 10^7 cm^{-2} for the graded buffer method. It is known tthe hat surface roughness and re-

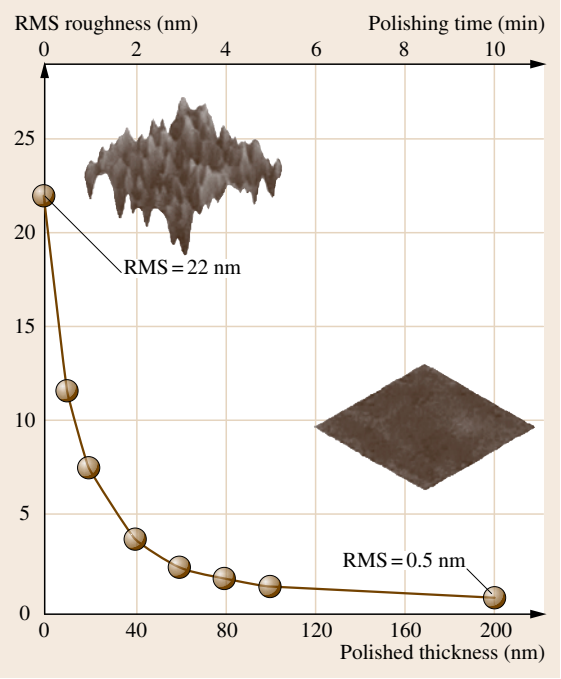


Fig. 34.16 CMP-polished thickness dependence of root-mean-square (RMS) surface roughness of relaxed SiGe buffer layers

laxation ratio do not change much when the growth temperature of LT Si layers is varied between 300 and 400 °C. However, the quality of buffer layers grown on LT Si largely depends on the buffer layer growth temperature. The relaxation ratio increases from 0 to 90% as the temperature increases from 400 to 700 °C and the roughness also increases, from 1 to 3 nm, with increasing temperature as shown in Fig. 34.15.

The quality of buffer layers largely depends on the Ge composition as well. Surface roughness as well as threading dislocation density increase with increasing Ge content up to 40% and then decrease.

From positron annihilation experiments, it is revealed that the Si layer grown at 400 °C contains a lot of defects, particularly vacancy clusters [34.115]. So, it is reasonably considered that these defects act as dislocation sources and confine dislocations in the vicinity of LT Si layers.

When high-Ge-content buffer layers are grown, the two-step growth of LT layers is found to provide strained overgrown layers with higher quality than the one-step growth method. Although the surface roughness of the buffer is hardly distinct between the two- and one-step methods, short-period roughness is found to be superimposed in the case of the one-step method. This roughness may degrade the transport properties of the structures grown on it, and higher mobility was obtained in the case of the two-step method.

34.7.3 Chemical–Mechanical Polishing Method

Since obtaining high-quality relaxed SiGe buffer layers is a critical issue and the quality of the layers obtained by the methods described above is still not satisfactory for production, there have been a lot of attempts to develop better techniques. Chemical–mechanical polishing (CMP) of SiGe buffer layers with large roughness is one of the promising techniques [34.116–120]. CMP consists of mechanical polishing by small particles and chemical etching and is well established for preparation of Si wafers. The surface flatness of SiGe buffers after proper CMP polishing is almost equal to that of Si wafers, as shown in Fig. 34.16 where surface roughness is plotted as a function of polishing time [34.116]. It is demonstrated that the mo-

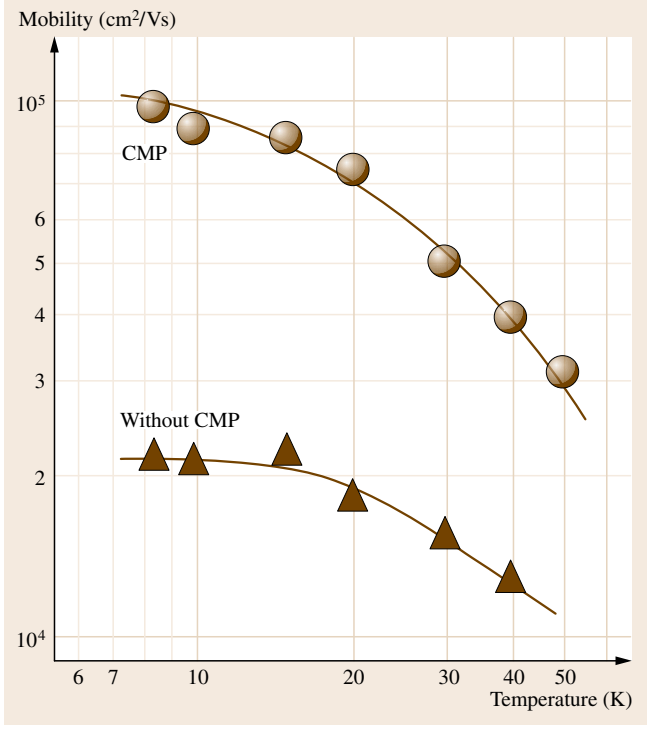


Fig. 34.17 Temperature dependence of electron mobility of SiGe/Si modulation-doped structures with and without CMP polishing

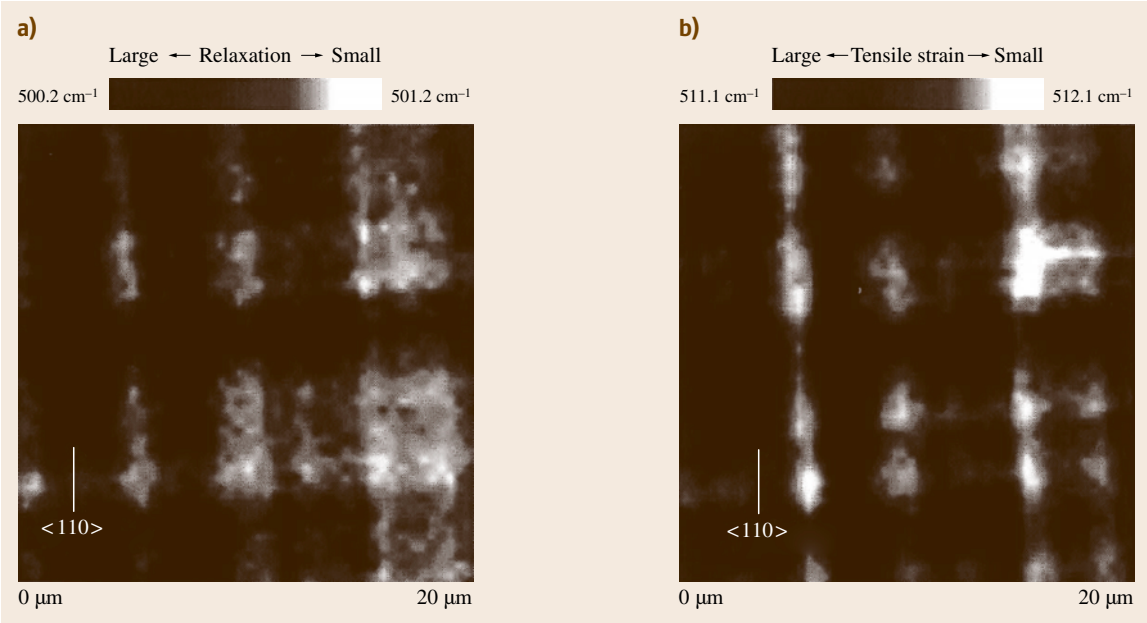


Fig. 34.18a,b Raman mapping of (a) relaxed SiGe buffer layer and (b) strained Si layer grown on SiGe buffers

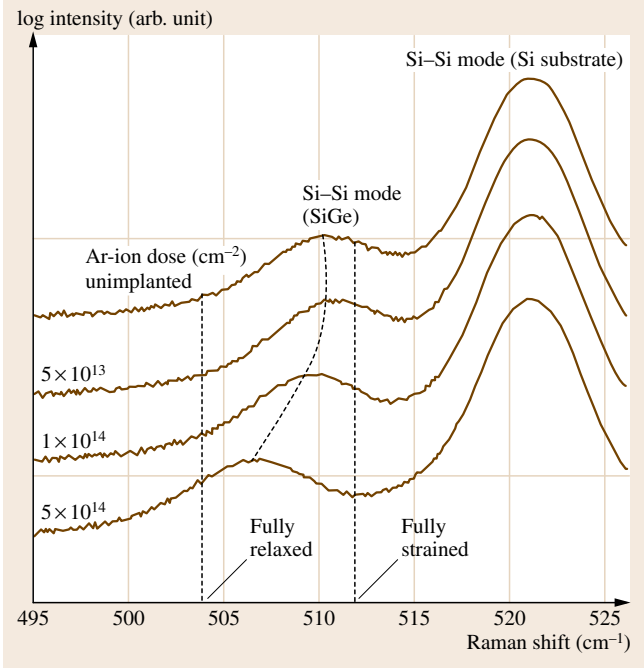


Fig. 34.19 Raman spectra of SiGe buffer layers grown on unimplanted and implanted Si substrates. Si–Si mode of SiGe layer is seen to shift to smaller wavenumber with increasing Ar-ion dose

bility of MOD structures formed on the flat surface is enhanced due to the reduction of interface scattering [34.117]. Figure 34.17 shows the temperature dependence of the mobility of the MOD structures formed on CMP polished and unpolished SiGe buffer layers. It is seen that the CMP sample provides an electron mobility four times higher than that of the unpolished sample. The mobility is shown to increase to about 600 000 $cm^2/(V\,s)$ at very low temperatures. As for hole mobility, it has also been demonstrated recently that p-type Ge channel grown on CMP-polished SiGe buffers provides mobility eight times higher than that on unpolished ones [34.120].

However, it is noted that, even though CMP provides very flat surfaces, there still exists strain fluctuation coming from the underlying misfit dislocations in the SiGe buffer [34.121]. Figure 34.18 shows Raman mapping of SiGe buffer as well as strained Si grown on the buffer, from which it is clear that strain fluctuations similar to the cross-hatch pattern exist. It is also known that this strain fluctuation affects the growth rate of the overlayer and causes surface roughening [34.121].

34.7.4 Ion Implantation Method

Another interesting approach is ion bombardment of Si substrates. It was demonstrated [34.122–126] that

proton or helium ion implantation into Si substrates through epitaxially grown SiGe layers and annealing provided good strain relaxation of SiGe layers, and misfit dislocations were generated at the interface thanks to the defects introduced by the ion bombardment. To control defect formation well in the surface region of Si substrates, ion implantation with heavy ions such as Ar and Ge before epitaxial growth of SiGe layers has been shown to be very effective to relax the SiGe buffer layers [34.127–130]. Figure 34.19 shows Raman spectra of implanted and unimplanted samples, showing clearly that the implanted sample shows good strain relaxation, even though the SiGe buffer is much thinner than that produced by the conventional graded method. Many defects, mainly consisting of vacancy clusters, act as nucleation centers as well as dislocation absorbers similar to the role of the LT buffer layers. A thickness of only 100 nm is good enough to obtain fully relaxed SiGe buffer layers, which is attractive from the point of view of production. However, the surface roughness is not very low, being almost the same as that of the LT method.

Surface flatness comparable to that of Si wafers was found to be obtained when thin SiGe layers were pseudomorphically deposited at low temperatures on ion-implanted Si substrates and postannealing was carried out at relatively high temperatures [34.131]. As seen in Fig. 34.20, the surface of the SiGe buffer formed by this method is very smooth, almost the same as for Si wafers, and the relaxation ratio is more than 80%. It is also noted that the strain distribution is much more uniform and no cross-hatch-like pattern is observed. Since the cross-hatch pattern is not seen in this sample, uniform distribution of misfit dislocation without bunching or strain-relief mechanisms due to point defects, different from the conventional misfit dislocation formation, may occur in this sample.

This the ion implantation method is very useful for growth methods such as CVD and GSMBE where low-temperature growth cannot be performed to decompose source gases.

34.7.5 Ge Condensation Method

The ultimate application of strained Si may be as silicon-on-insulator (SOI). SOI has a lot of advantages over the bulk devices, such as suppression of short-channel effects and so on. To realize high-quality strained Si on relaxed SiGe-on-insulator (SGOI), a very unique and attractive method called the Ge condensation method was proposed [34.132–135]. The principle

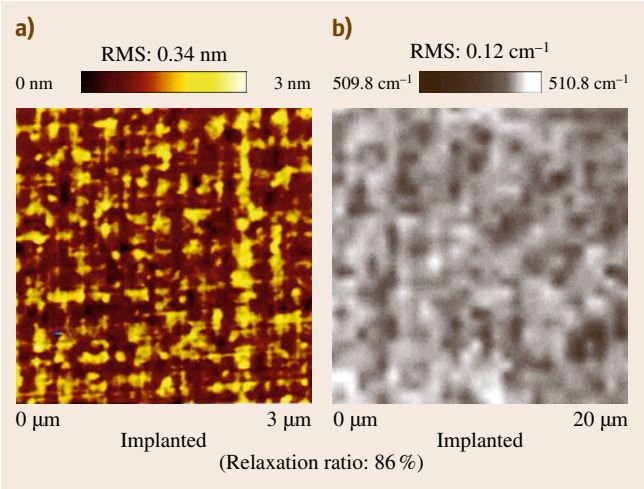


Fig. 34.20 (a) AFM image and (b) Raman mapping of SiGe buffers grown on ion-implanted Si substrate

of this method is shown in Fig. 34.21. When SiGe layers are oxidized, only Si atoms are consumed to form SiO_2 while Ge atoms are rejected from the oxide film. Therefore, when one deposits SiGe layers on SOI substrates and oxidizes the SiGe layers, the Ge content of the unoxidized SiGe layers is increased as the oxidation proceeds. It is also found that strain relaxation of SiGe grown on SOI simultaneously takes place during oxidation. As a result, relaxed SiGe layers with higher Ge content are formed on the buried oxide, that

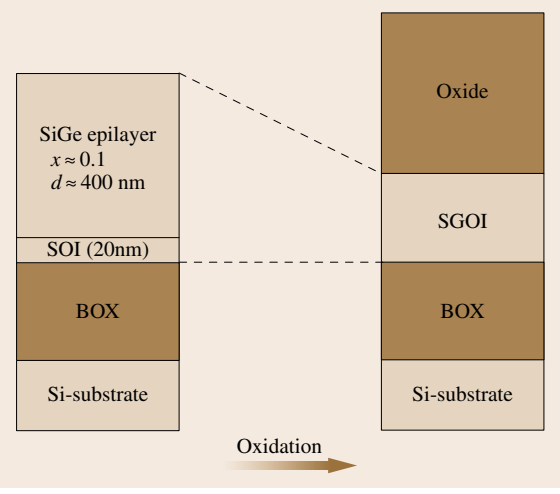


Fig. 34.21 Schematic illustration of Ge condensation method applied on SiGe epitaxial layer grown on SOI substrates. BOX is the buried oxide

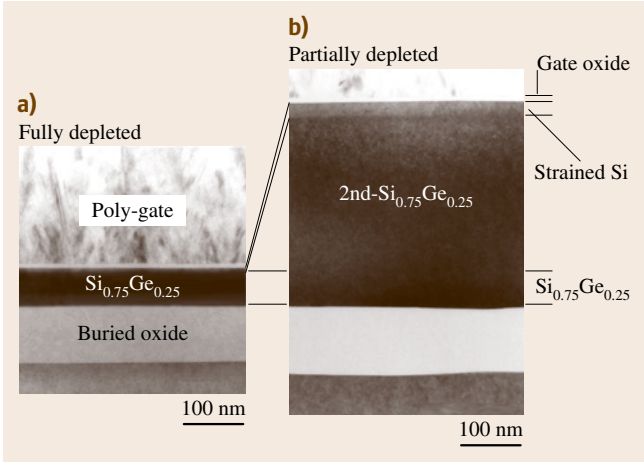


Fig. 34.22a,b TEM images of strained Si grown on (a) thin (fully depleted) and (b) thick (partially depleted) relaxed SiGe buffer layers formed by Ge condensation method

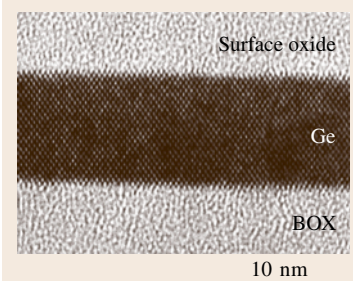


Fig. 34.23 TEM picture of Ge-on-insulator (GOI) structure fabricated by Ge condensation method

is, SGOI structures are formed. Strained Si layers can, therefore, be grown on this SGOI. Figure 34.22 shows TEM images of SGOI with strained Si layers; both fully (thin) and partially depleted SGOI (thick) structures are seen to be formed [34.136]. In the extreme case, pure Ge layers can be formed by this Ge condensation method and Ge-on-insulator (GOI) is realized, as shown in Fig. 34.23 where the Ge content of the layer on the insulator becomes almost 100% after complete oxidation of SiGe layers [34.137].

34.7.6 Dislocation Engineering for Buffer Layers

In general, glide dislocations, i. e., 60° dislocations, are predominantly introduced into the SiGe/Si(001) interface as a result of the operation of the ⟨110⟩/{111} slip system. Although misfit strain is mainly relaxed by an edge component of the Burgers vector of the 60° dislocation parallel to the interface, a screw component and components nonparallel to the interface simultaneously induce nonisotropic strain relaxation [34.138], as well as crystallographic tilting and rotation of SiGe. Such crystallographic inhomogeneity severely influences the lattice deformation of the channel Si and eventually leads to nonuniformity of the energy band structure. In this subsection, recent experimental results on epitaxial growth in which dislocation generation and propagation are precisely controlled during strain relaxation of SiGe are presented.

The two-step strain-relaxation procedure [34.139] was demonstrated to realize fully strain-relaxed thin buffer layers with low threading dislocation densities. This procedure consists of, first, annealing of a metastable pseudomorphic SiGe layer with a Si cap layer and, second, subsequent growth of SiGe on that layer. The thin cap layer effectively suppresses surface roughening during the annealing due to reduction of the surface stress of the film [34.140, 141]. More than 90% relaxation was achieved after growth of only a 100 nm thick second $Si_{0.7}Ge_{0.3}$ layer. Figure 34.24 shows a representative cross-sectional TEM image of a sample having a $Si_{0.7}Ge_{0.3}$(200 nm)/Si-cap(5 nm)/$Si_{0.7}Ge_{0.3}$(50 nm)/Si(001) structure. Note that threading dislocations are almost absent from the observed area and almost all misfit dislocations are confined at the first SiGe/Si substrate interface. These dislocations tend to be dispersed at the interface and pileup of the dislocations, which is often observed in compositionally graded layers [34.108], is hardly observed. The observed periodic surface undulation comes



Fig. 34.24 Cross-sectional TEM image of a sample grown by the two-step strain-relaxation procedure. The sample has a $Si_{0.7}Ge_{0.3}$(200 nm)/Si-cap(5 nm)/$Si_{0.7}Ge_{0.3}$(50 nm)/Si(001) structure. Threading dislocations are almost absent from the observed area and almost all misfit dislocations are dispersively confined at the first SiGe–Si substrate interface. The observed periodic surface undulation comes from aligned SiGe islands formed at the earlier stage on the misfit dislocation network buried at the first SiGe–substrate interface as a template

from aligned SiGe islands formed at the earlier stage on the misfit dislocation network buried at the first SiGe–substrate interface as a template. Similar preferential nucleation over dislocation was reported by *Xie* et al. [34.142]. As mentioned in Sect. 34.6, since a cusp in the surface undulation acts as a preferential nucleation site for misfit dislocations [34.143], dislocation half-loops are likely introduced at every cusp on the surface to relax the strain during the growth of the second layer. Therefore, strain relaxation is dominated by the introduction of new dislocations from the surface, and a regular strain field created by the periodic undulation greatly enhances the propagation of the introduced dislocations so that the threading segments have the opportunity to travel long distances.

Due to the intrinsic structure of a 60° dislocation, a SiGe film strain-relaxed by 60° dislocations often exhibits mosaicity [34.144] and cross-hatch patterns [34.106, 108]. In order to prevent such degradation, the introduction of pure-edge dislocations is crucial. A novel approach based on strain relaxation predominantly by a pure-edge dislocation network buried at the SiGe/Si(001) interface was recently demonstrated [34.145]. Employing pure-Ge thin-film growth prior to SiGe formation effectively restrains the introduction of 60° dislocations; instead, a high density of pure-edge dislocations can be generated [34.146]. Figure 34.25a shows a plan-view TEM image of the Ge layer, which was grown at 200 °C and then subjected to annealing at 700 °C. A dislocation network, consisting of pure-edge dislocations aligned along two orthogonal ⟨110⟩ directions, can be clearly observed at the Ge–Si(001) interface. For forming SiGe layers, solid-phase

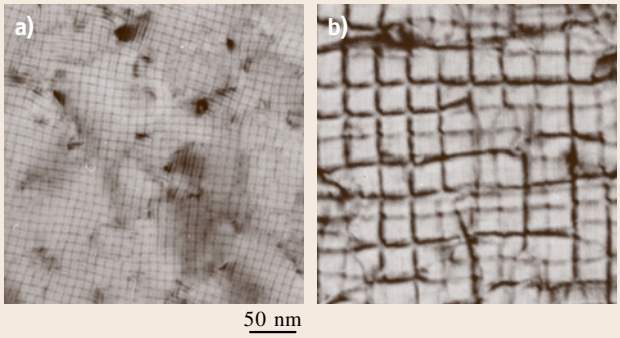


Fig. 34.25 **(a)** Plan-view TEM image of a Ge layer on Si(001), which was grown at 200 °C and then subjected to annealing at 700 °C, showing a dislocation network consisting of pure-edge dislocations aligned along two orthogonal ⟨110⟩ directions. **(b)** Plan-view TEM image of an a-Si (17 nm)/Ge(30 nm)/Si(001) sample annealed at 1100 °C (the scale is the same as that in **(a)**). The morphology of the pure-edge dislocation network is retained even after the high-temperature annealing, but the dislocation spacing increases

intermixing of amorphous Si (a-Si) deposited on the Ge layer was performed. Figure 34.25b shows a plan-view TEM image of an a-Si (17 nm)/Ge(30 nm)/Si(001) sample annealed at 1100 °C. Note that the morphology of the pure-edge dislocation network is explicitly retained even after the high-temperature annealing but the dislocation spacing is found to increase. Figure 34.26a shows an x-ray diffraction two-dimensional reciprocal-space map around a SiGe(115) diffraction peak of the sample annealed at 1100 °C. From the peak position,

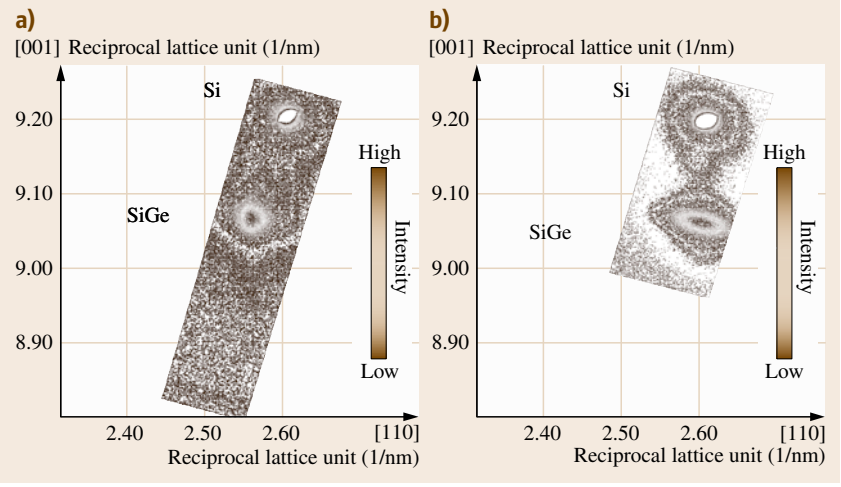


Fig. 34.26a,b XRD two-dimensional reciprocal maps around Si(115) and SiGe(115) diffraction peaks of **(a)** the sample annealed at 1100 °C, which has a pure-edge dislocation network at the SiGe–Si(001) interface, and **(b)** the sample with 60° dislocations grown by the two-step strain-relaxation procedure

it is found that an in-plane lattice constant corresponding to that of a SiGe crystal with a Ge fraction of 0.48 is obtained, enabling practical use as a SiGe buffer. A remarkable feature is found in the peak shape being sharp and highly symmetric, in marked contrast to that of the reference sample shown in Fig. 34.26b, which is prepared by the two-step strain relaxation procedure and predominantly has 60° dislocations at the interface. This result clearly demonstrates that the mosaicity, such as lateral finite sizes and microscopic in-plane tilts of the SiGe crystal [34.147], is greatly reduced when introducing pure-edge dislocations.

34.7.7 Formation of SiGeC Alloys

Since the first epitaxial growth of high-quality $Si_{1-y}C_y$ and $Si_{1-x-y}Ge_xC_y$ films on Si(001) was reported in 1992 [34.148–150], the introduction of C into Si and SiGe films has attracted increasing interest for application of these films to electric and optoelectronic devices composed of group IV semiconductor materials. Due to the different covalent radius of Si, Ge, and C of 0.117, 0.122, and 0.077 nm, respectively, substitutional introduction of C atoms into SiGe lattice sites gives rise to a substantial decrease of the lattice parameter and a reduction of strain in SiGe layers on Si substrates [34.151–154]. The addition of C into $Si_{1-x}Ge_x$ films is, therefore, expected to yield several advantageous effects in terms of device performance. In Si/SiGe/Si HBTs, when C is incorporated into a SiGe base layer, the strain caused by the SiGe/Si lattice mismatch can be alleviated and thus the generation of misfit dislocations can be prevented, realizing improved crystalline quality of the base layer and resultant reliability of the devices. Furthermore, C incorporation can significantly suppress boron outdiffusion from a p-type base layer due to the undersaturation of interstitial Si atoms [34.155–158]. On the other hand, it has been reported that the bandgap and band offsets with respect to the conduction band and valence band sensitively vary with C contents and strain in $Si_{1-x-y}Ge_xC_y$ films [34.159–169]. On the basis of the control over the band alignment in $Si_{1-x-y}Ge_xC_y$ heterostructures, $Si_{1-x-y}Ge_xC_y$ epitaxial films were applied to channel layers in MOSFET [34.170, 171], high-electron-mobility transistor (HEMT), and optical devices [34.172–174]. However, substitutional incorporation of C exceeding a few percent is very difficult to achieve. One dominant reason seems to be the thermal equilibrium solubilities of C into Si and Ge, of the order of 10^{17} and 10^8 atoms/cm^3, respectively [34.175, 176]. The crystalline quality of the films is degraded with increasing C fraction and exhibits nonplanar morphology, SiC polytype precipitates, and extended defects such as stacking faults and dislocations. Therefore, film growth techniques in which the growth mode is governed not by thermodynamics but kinetics are now widely employed, such as MBE [34.150, 151, 177, 178], UHV-CVD [34.179, 180], and RT-CVD [34.181, 182]. However, this still remains an essential issue in the growth of high-quality epitaxial $Si_{1-x-y}Ge_xC_y$ films with a substitutional C content higher than 3% and this limits the potential of these films to be applied widely to various kinds of devices [34.178, 183]. Enhanced solubility of C was theoretically predicted [34.184], and it was experimentally demonstrated that $Si_{1-y}C_y$ layers with $y \approx 0.2$ can be grown pseudomorphically on Si(001) due to the formation of low-energy ordered structures [34.185]. In the case of $Si_{1-x-y}Ge_xC_y$ films, high repulsive interaction between Ge and C in the Si lattice plays a dominant role in determining the composition profiles of the film; both theoretical and experimental works on this matter have been performed [34.152, 186, 187]. Furthermore, attractive interaction between Si and C affects the final film morphology. Figure 34.27a shows an STM image of a surface of a $Si_{0.478}Ge_{0.478}C_{0.044}$ film with a thickness of four monolayers (ML) [34.188]. The surface exhibits the onset of three-dimensional (3-D) islanding, suggesting the local increase of Ge fraction around the island where the film locally exceeds the critical thickness for islanding. It should be noted that these islands are different from those

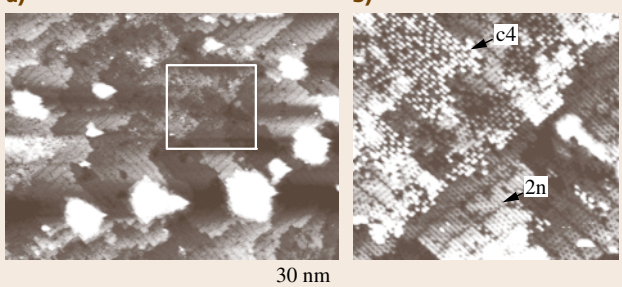


Fig. 34.27a,b STM images of a $Si_{0.478}Ge_{0.478}C_{0.044}$ film of 4 ML thickness grown on Si(001). **(a)** Islands with a height of about 2 nm on average are frequently observed on top of protruding regions of the terraces. **(b)** Magnified 30×30 nm^2 image of the *boxed area* in **(a)**. Examples of the $c(4\times4)$ and $(2\times n)$ reconstructions are indicated by the labels "c4" and "2n," respectively

due to C-induced Ge islanding [34.189, 190], since they were not formed at the beginning of the growth. As shown in Fig. 34.27b, a close-up of the fairly flat terraces reveals the $c(4\times4)$ reconstructed structure caused by significant C condensation on the growing surface [34.191–193] and the $(2\times n)$ reconstruction consisting of buckled dimers and missing dimer rows. These results clearly show the formation of C-rich and Ge-rich regions in the film, which is presumably driven by the phase separation between Si-C and Si-Ge during growth when the film contains high concentration of C.

Control over the initial growth stage was performed to improve the film morphology [34.188, 193]. The addition of a thin (1–2 ML) SiGe interlayer between the $Si_{1-x-y}Ge_xC_y$ film and the Si substrate drastically improves the film structure, leading to a planar morphology even with large C fractions present in the film. Figure 34.28 shows an STM image of a sample which has a structure of $Si_{0.473}Ge_{0.473}C_{0.054}$(6 ML)/$Si_{0.5}Ge_{0.5}$(1 ML)/Si(001). Note that a planar surface morphology consisting of a step and terrace structure is formed even though the film contains a 4.6% average C fraction. Three-dimensional islanding partially appears at the step edges with increasing $Si_{1-x-y}Ge_xC_y$ film thickness [34.194, 195] but no $c(4\times4)$ reconstructions were observable in any growth stage. This clearly demonstrates that the $Si_{1-x}Ge_x$ interlayer explicitly plays a role in suppressing C condensation and Si-C/Si-Ge phase separation during the film evolution.

More recently, sequential alternate deposition of 1 ML thick $Si_{0.793}Ge_{0.207}$ and 0.048 ML thick C layer on Si(001) was attempted in order to suppress local phase separation, 3-D island growth, and defect formation [34.196]. A comparison of the surface atomic morphologies between the 5 ML thick $Si_{0.769}Ge_{0.183}C_{0.048}$ layer formed by the codeposition of Si, Ge, and C and that by the alternate deposition is shown in Fig. 34.29. In Fig. 34.29a, the 3-D islands, with a height of approximately 2 nm, and a rough surface consisting of small terraces are observed. Many defects, observed as dark spots on the surface, which are sink sites of C atoms, seem to prevent conformal step-flow growth due to the positive (normal) Ehrlich–Schwoebel barrier [34.197–199]. On the other hand, in Fig. 34.29b, no 3-D islands are observed and aligned steps are still formed. The density of defects is relatively low compared with the codeposition case. In the alternate deposition case, the migration of C on the growing surface is effectively restrained because the Ge atoms, which give rise to the repulsive interaction force to the deposited C atoms, are uniformly distributed on the $Si_{1-x}Ge_x$ surface. This effect leads to suppression of defect formation at the initial stage of $Si_{1-x-y}Ge_xC_y$ growth and consequently increases the critical thickness of layer-by-layer growth of $Si_{1-x-y}Ge_xC_y$.

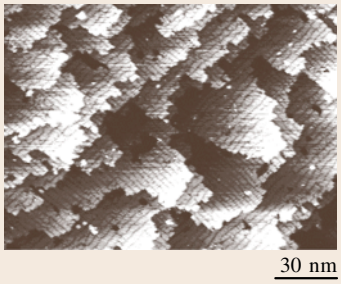


Fig. 34.28 STM image of a sample with a structure of $Si_{0.473}Ge_{0.473}C_{0.054}$ (6 ML)/$Si_{0.5}Ge_{0.5}$ (1 ML)/Si(001). A planar surface morphology is realized by the formation of the SiGe interlayer, in spite of the large C fraction

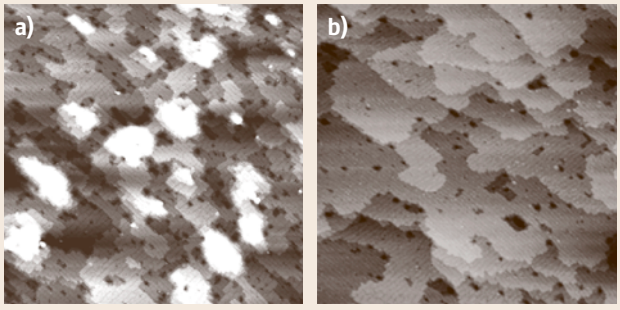


Fig. 34.29a,b STM images of the 5 ML thick $Si_{0.769}Ge_{0.183}C_{0.048}$ layer formed by **(a)** the codeposition of Si, Ge, and C; **(b)** the alternating deposition of 1 ML thick $Si_{0.793}Ge_{0.207}$ and 0.048 ML thick C layers. In the codeposition, 3-D islands with a height of approximately 2 nm, a rough surface consisting of small terraces, and defects as dark spots are observed on the surface, while no 3-D islands with reduced defects are seen in the case of the alternating deposition

34.8 Formation of Quantum Wells, Superlattices, and Quantum Wires

The formation of double heterostructures results in quantum well (QW) formation. Type I QWs are easily realized by growing strained SiGe layers on Si substrates and sandwiching it with unstrained Si lay-

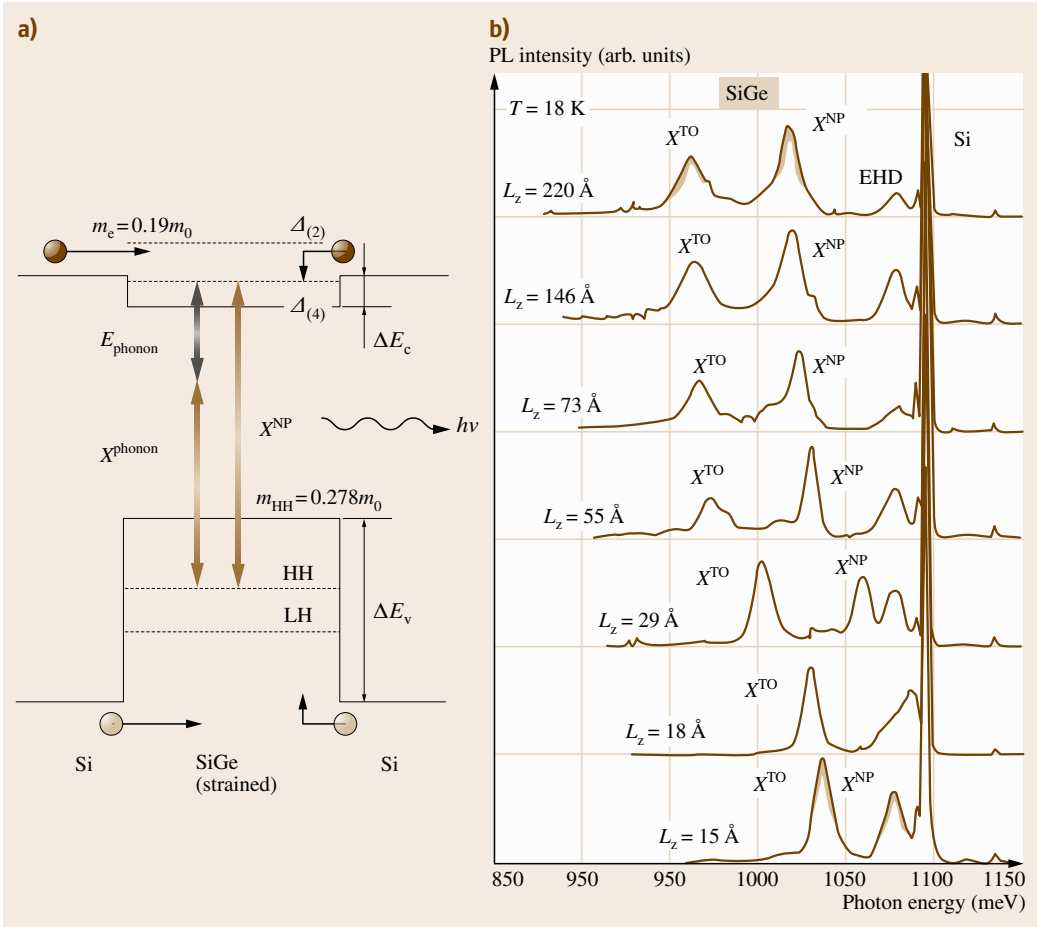


Fig. 34.30a,b Schematics of PL from strained SiGe/Si quantum well **(a)** and well width dependence of PL spectra of the quantum well **(b)**. NP represents luminescence peaks without phonon assistance, and phonon and TO represent phonon replica peaks

ers. Peak shift due to the quantum confinement effect characteristic of QWs is clearly seen in the PL spectra of QWs grown by gas-source MBE, as shown in Fig. 34.30 [34.200]. There are two prominent PL peaks, no-phonon (NP) and the TO-phonon replica, which are characteristic of indirect-bandgap transition in this system, and both peaks shift to higher energies with decreasing well width. The theoretical calculation based on square QWs well represents the well width dependence of the quantum confinement energy.

Although the luminescence peaks become broader with increasing temperature, their integrated intensity is stable up to 100 K. Above 100 K, however, the intensity decreases rapidly with an activation energy of around 100 meV [34.34]. This activation energy corresponds to the energy difference between the valence band of Si barriers and the ground quantum level in the SiGe QW. That is, carrier confinement of QWs is quite efficient at low temperatures but holes begin to escape from the well above the temperature corresponding to the confinement energy.

Since the band alignment changes depending on the strain condition, type II QWs are also formed. When tensile-strained Si layer is sandwiched by unstrained SiGe barrier layers, holes are trapped in the Si well layers while electrons are located in the SiGe barrier layers and therefore the transition is indirect in real space as well as in momentum space. However, luminescence, due to both NP and TO peaks, is clearly observed and the energy shift coming from the quantum confinement effect is confirmed [34.201].

Although the band alignment is type I when strained SiGe layers with Ge composition less than 30% is set between unstrained Si layers, the band discontinuity at the conduction band is too small to confine electrons well in the well region. To overcome this problem, there

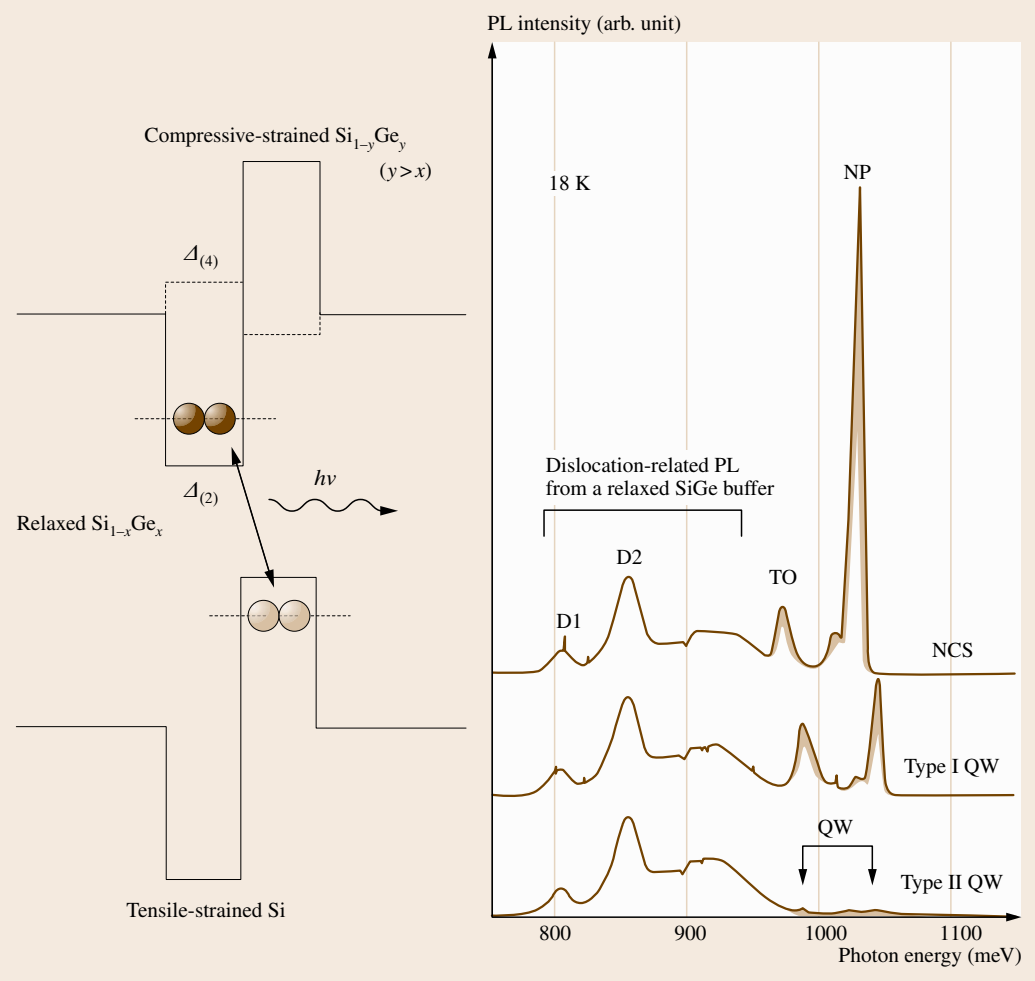


Fig. 34.31 Band structure of Si/SiGe neighboring confinement structure (NCS) and PL spectra of NCS, type I and type II quantum wells

are several ideas to improve the carrier confinement, one of which is the neighboring confinement structure (NCS), in which a pair of compressively strained and tensile-strained layers is sandwiched by strain-relaxed SiGe barrier layers, as shown in Fig. 34.31 [34.202]. Although electrons and holes are separately confined in the well region, the PL intensity, particularly the NP peak, is significantly enhanced in this structure and is stronger than that in conventional type I QWs. This comes from the significant wavefunction overlap between electrons in Si and holes in Ge layers, but it is noted here that the thickness of the pair layers should be thin to keep sufficient wavefunction overlap for the transition.

The coupling of QWs results in the formation of superlattices, which is seen in the energy shift of PL peaks, with the peaks shifting to lower energies with decreasing distance between the QWs [34.203]. The peak energies of coupled QWs agree well with the calculated results based on ideal square-shaped wells, in accordance with the theoretical prediction [34.1] that the band alignment is type I when Ge content is lower than 0.3. This also indicates that wells without deformation due to surface segregation are formed by GSMBE.

The evolution of superlattices is seen when the number of coupled wells is increased [34.204]. Since the increase in the number of coupled wells lowers the ground-state energy and finally forms a miniband, the PL peak energy, corresponding to the miniband edge, decreases with increasing well number. The peak energy is found to follow a simple Kronig–Penny-type calculation for the superlattices [34.204].

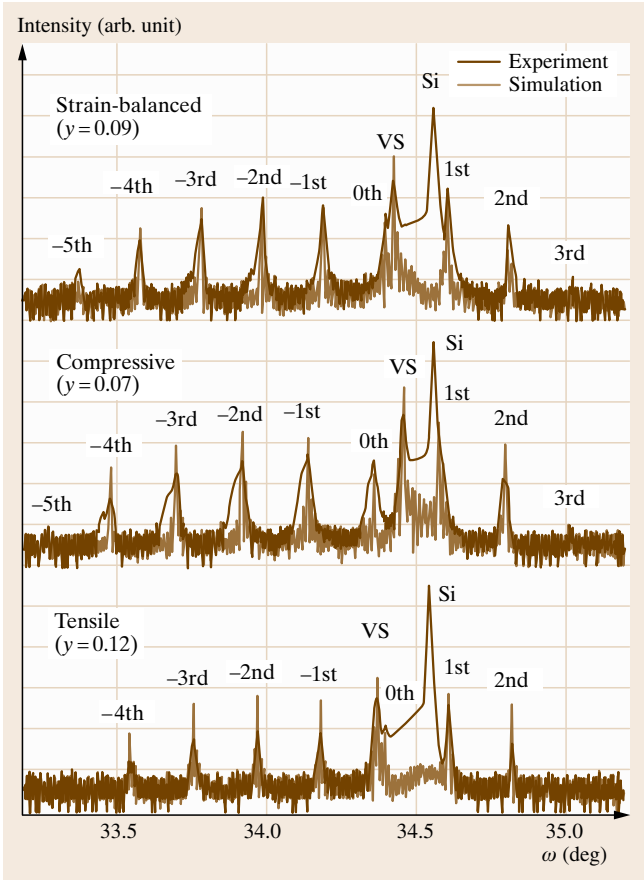


Fig. 34.32 XRD intensity profiles of SiGe/Si multilayers (DBR structures) grown under strain-balanced, compressive, and tensile conditions

Multilayer structures, in which each layer is much thicker than in superlattices, are also important from the point of view of optical device applications. The typical example of such structures for optical applications is the distributed Bragg reflector (DBR) mirror, which consists of multilayers of materials with different refractive index, such as GaAs/AlAs and Si/Ge. Since the period of DBR structures should be comparable to the wavelength of light in semiconductors, the thickness of each layer is as large as ten times that in a superlattice. Therefore, the growth of such structures with a strained system is very difficult due to the limitation of the critical thickness. To overcome this problem, strain-balanced structures are very useful [34.205], and DBR mirrors consisting of Si and SiGe layers have been successfully fabricated [34.206–209]. In this method, a pair of layers with lattice constants a_1 and a_2 are grown on a substrate with lattice constant a_0, where $a_1 < a_0 < a_2$. This condition is satisfied when a pair of Si layers with tensile strain and SiGe layers with compressive strain are grown on relaxed SiGe with Ge content less than that of the pair. The thickness and Ge composition are selected so that the strain energy of the structure is minimized.

Figure 34.32 shows x-ray diffraction measurements where the Ge content of the virtual substrate is changed. It is seen that the peak of the sample with $y = 0.09$ coincides well with the position of the strain-balanced condition, showing that strain-balanced structures are formed by carefully choosing the layer thickness and Ge contents. Figure 34.33 shows an SEM image of a DBR structure grown under the strain-balanced condition, and it is seen that high-quality structures without detectable defects are obtained.

Figure 34.34 shows the reflection spectrum of SiGe DBR mirrors fabricated by this method. In this structure, Si and SiGe layers with thickness of 94 and 90 nm, respectively, were grown on fully relaxed SiGe buffer layers to compensate the strain. Reflectivity of about 90%, which is quite large for the strained system, is be obtained.

Selective epitaxial growth, which takes place in growth techniques such as gas-source MBE, leads to the formation of wire and dot structures. This occurs, for instance, between Si and SiO_2 substrates, and epitaxy occurs only on Si. When SiGe/Si QWs are formed on V-groove-patterned Si(100) substrates with (111) facets [34.210, 211], crescent-shaped SiGe features are grown in the bottom of the V-groove. This feature gives rise to PL with large blue-shift in the spectrum compared with the reference QW sample. It is also seen that the cross-sectional emission in the case of the wire

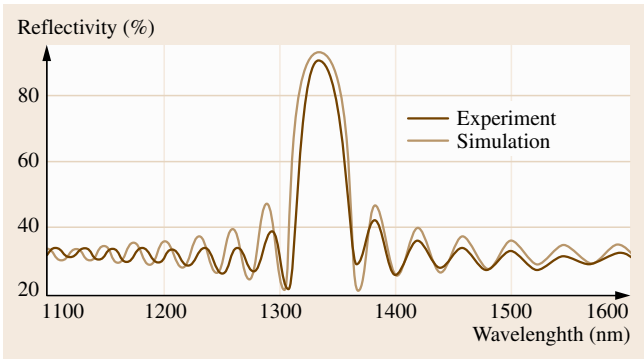


Fig. 34.34 Reflectivity spectrum of SiGe/Si DBR mirror formed by the strain-balanced method ◀

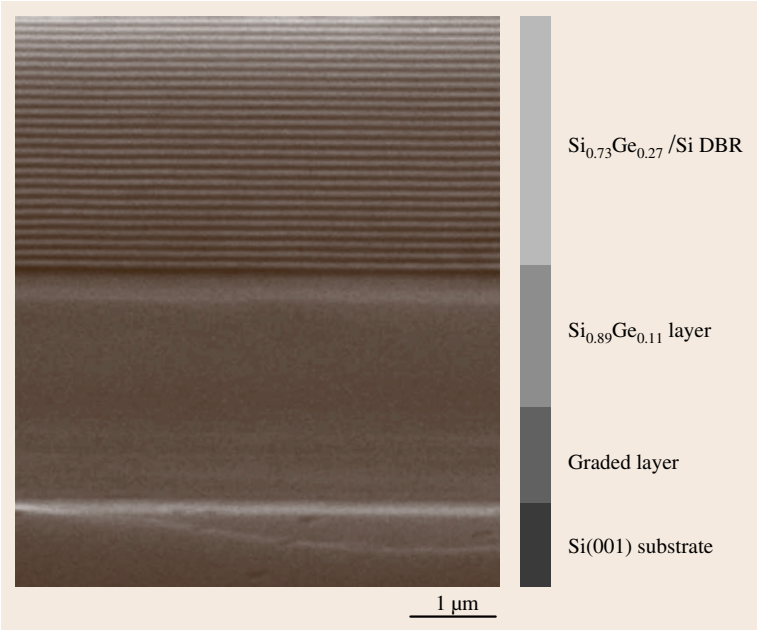


Fig. 34.33 Cross-sectional SEM image of $Si_{0.73}Ge_{0.27}$/Si DBR mirror grown on $Si_{0.89}Ge_{0.11}$ buffer layers formed by the graded-buffer technique

structure shows polarization characteristics while no polarization is seen in the emission from the QWs. This result suggests that the wire structure in the V-grooved sample is likely to be a quantum wire. However, the energy shift in quantum wires is too large and does not agree with a simple estimation based on their size. The main cause of the energy shift may be the change in the strain distribution. The wire surrounded by Si crystal is considered to be under hydrostatic-like pressure that causes bandgap broadening comparable to the observed energy shift. Since spatial variation of the Ge composition in the wire is also likely to occur, more detailed study is required to clarify the nature of quantum wires in the SiGe/Si system.

34.9 Dot Formation

There are a large number of studies on Ge and SiGe dot formation, which strongly depends on the growth methods and conditions. Since there are several good reviews [34.212–215] concerning Ge dot formation, some general features of the formation, which are observed in the case of gas-source MBE (GSMBE), are described here [34.216].

Figure 34.35 shows the thickness dependence of the PL spectrum from Si/pure-Ge/Si quantum-well structures grown by GSMBE [34.217]. Up to 3.7 ML, the PL spectrum shows a conventional quantum confinement effect of quantum wells and the peaks shift to lower energies with increasing Ge layer thickness. Above 3.7 ML, however, the peaks originating from QWs stop the energy shift and a new broad peak is seen to appear. The appearance of this broad peak corresponds well to the formation of Ge islands observed by TEM measurements. That is, above the critical thickness, Ge begins to form islands on Ge wetting layers to release the strain energy as described in the previous section. It is remarkable that these Ge dots give rise to significant luminescence, and therefore their application as quantum dots is eagerly awaited.

The critical thickness in the case above is 3.7 ML and depends on the growth temperature, as shown in Fig. 34.36. It is seen from this figure that the critical thickness increases with decreasing growth temperature.

It is well known that Ge dot formation takes place in a bimodal fashion, that is, small pyramidal-shape and large dome-shape dots are formed simultaneously, as shown in Fig. 34.37, and that their relative numbers

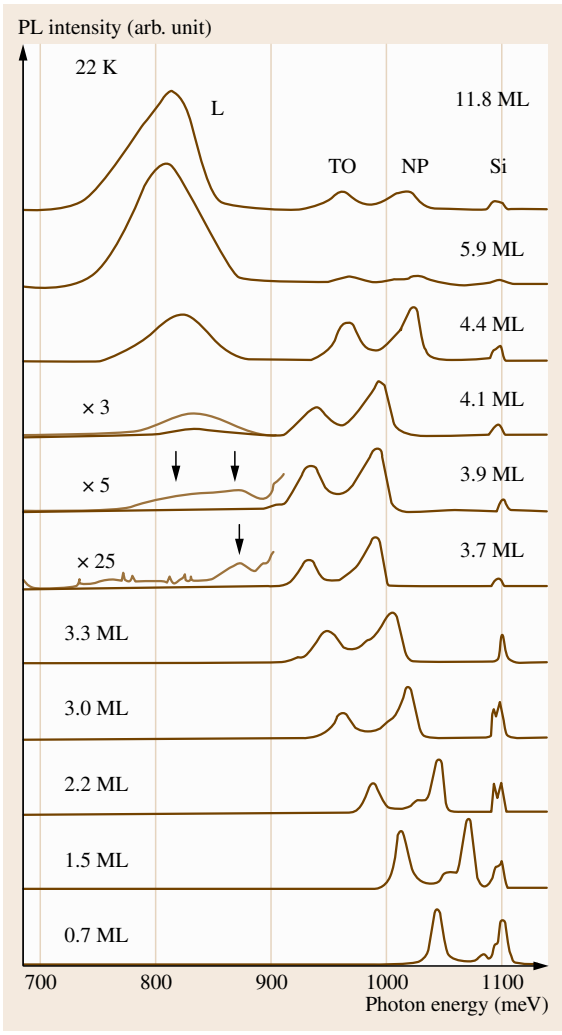


Fig. 34.35 Ge layer thickness dependence of PL spectrum of Si/pure-Ge/Si structures. Above 3.7 ML, a new broad PL peak (L) corresponding to Ge dot formation appears and grows with increasing Ge amount, while the peaks corresponding to quantum wells are seen to stop the energy shift once the dot is formed ◀

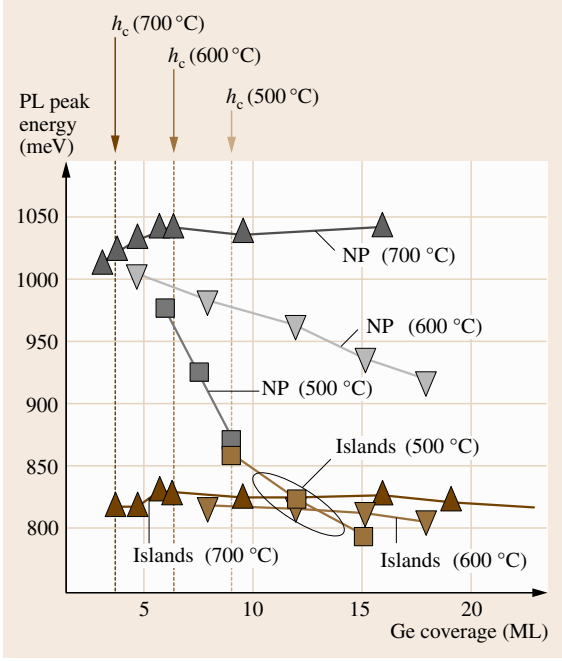


Fig. 34.36 Ge coverage dependences of NP and island PL peaks as a function of growth temperature. The critical thickness h_c of Ge dot formation is seen to increase with decreasing growth temperature

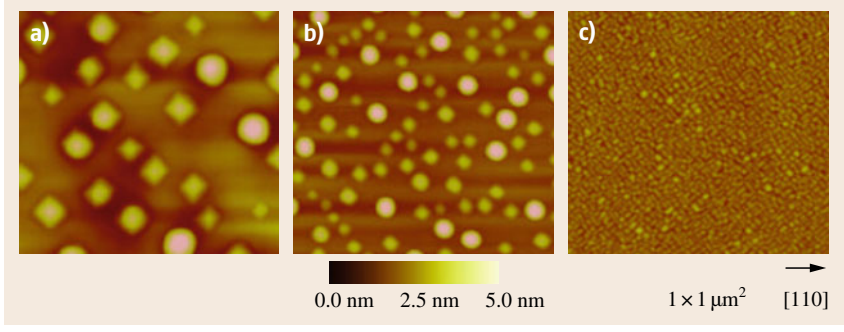


Fig. 34.37a–c AFM images of Ge dots grown at **(a)** 700 °C, **(b)** 600 °C, and **(c)** 500 °C. Scale is $1 \times 1\ \mu m^2$

change depending on the growth conditions. When the growth temperature is decreased, the number of dots becomes larger and their size becomes smaller; in particular, the density of pyramids becomes larger rather

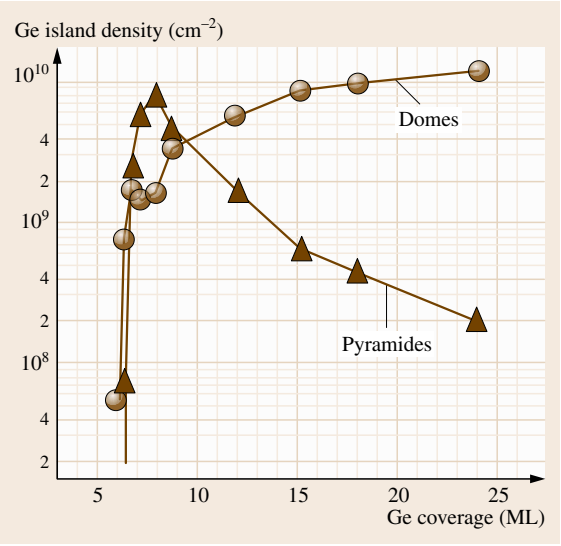


Fig. 34.38 Ge coverage dependence of island density. Morphological change is seen from pyramid to dome shapes when coverage exceeds about 8 ML

than that of domes. At lower temperatures, that is, 500 °C, domes and pyramids disappear and quite a large number of small elongated dots known as *hut clusters* appear, as seen in Fig. 34.37c. Energetically, domes are stable, whereas pyramids are metastable and therefore appear at rather higher temperatures. When the growth temperature is low, atom migration is suppressed and unstable hut clusters are formed.

Figure 34.38 shows the Ge coverage dependence of dot formation. It is interesting that a morphological change from pyramidal to dome shape is observed with increasing coverage. This is probably due to accumulating strain. Since the domes have a much larger degree of strain relaxation than the pyramids, this shape change may occur to reduce the total energy, and it dominates at higher coverage.

The relative number of pyramids to domes is also dependent on the growth rate. The density of pyramids is drastically decreased by decreasing the growth rate, and it is also found that inserting growth interruption or annealing decreases the number of pyramids. This reflects the fact that the dome is energetically more stable than the pyramid, and that the shape change takes place more easily under near-equilibrium conditions.

The formation of SiGe alloy dots is quite different from the case of pure Ge dots at first sight. As seen in Fig. 34.39, a large number of pyramids are formed compared with pure-Ge islands, and domes are hardly observed. The number of pyramids increases with increasing SiGe coverage, in contrast to the case of pure Ge, and the shape change does not take place. This feature is very similar to that of low-temperature growth of pure-Ge dots, in which hut clusters are formed and the shape change does not occur. This difference results from the growth mechanism. It is noted that, when the dots are formed at 600 °C by gas-source MBE with Si_2H_6 and GeH_4, the growth mode is in the reaction-limited regime, where growth is strictly limited by the chemical reaction of impinging gases. In the reaction-limited regime, the surface migration of atoms is suppressed and the shape change from pyramid to dome hardly occurs since the energy barrier is too high compared to the surface migration energy. On the other hand, the growth of pure Ge proceeds in the supply-limited regime at this temperature, and the suppression of surface migration is quite small, which allows the formation of stable domes. If the growth of SiGe dots is performed in the supply-limited regime, that is, at higher growth temperatures, the situation changes greatly and very similar behavior to that of pure Ge dots is seen, as shown in Fig. 34.40. In this figure,

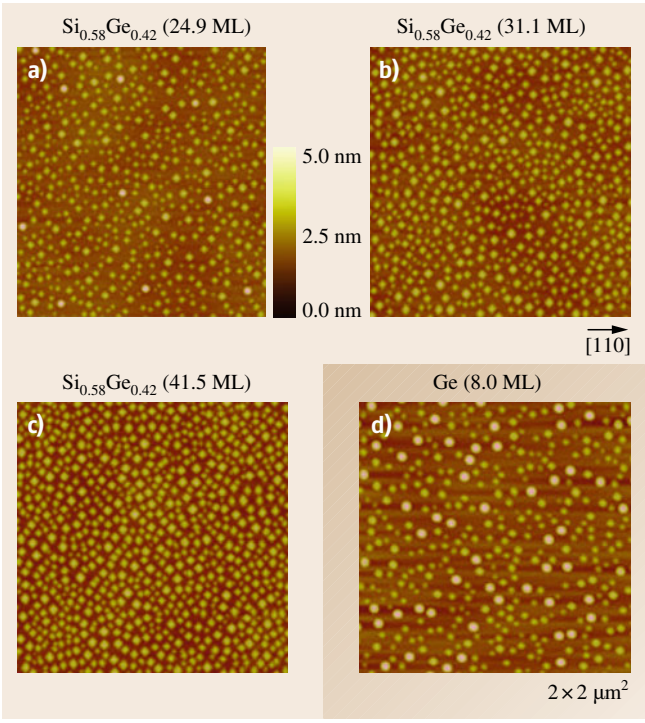


Fig. 34.39a–d Coverage dependence **(a–c)** of AFM images of $Si_{0.58}Ge_{0.42}$ dots and **(d)** Ge dots with 8 ML. Scale is 2×2 μm^2

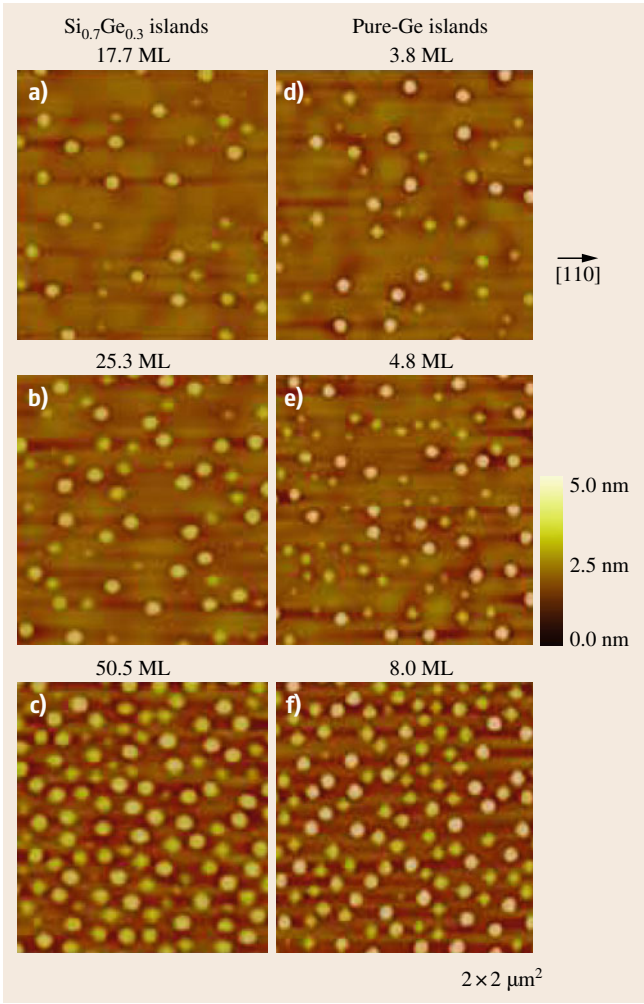


Fig. 34.40a–f Coverage dependence of AFM images of $Si_{0.7}Ge_{0.3}$ dots grown at high temperatures **(a–c)** and pure Ge dots grown at low temperatures **(d–f)**. Scale is $2\times2\,\mu m^2$

bimodal growth of islands and the shape change are clearly seen for SiGe islands. This similarity suggests intermixing of Si and Ge even for pure-Ge dot formation. It is now known that the intermixing effect of Si and Ge, particularly at high temperatures, is important for dot formation and reduces the formation barrier for dots, resulting in a smaller critical thickness than that for low-temperature growth.

The formation of Ge dots on a strained SiGe layer is also important from the device application point of view. Although the critical thickness for dot formation is different from that on Si substrates, the bimodal growth and the shape change are very similar to those on Si. The Ge coverage dependence of the dot density is just shifted to lower coverage, which can be understood if we consider the strain energy of the two-dimensional (2-D) underlying strained layers. That is, the strain energy of the Ge thickness difference between SiGe and Si substrates for dot formation is almost equal to that of the underlying strained SiGe layer. This implies that the same dot formation mechanism is in operation and that the strain of the underlying layers contributes to dot formation.

Stacking of dot layers is attracting great attention from the point of view of crystal growth as well as device applications. As mentioned above, the strain originating from the underlying layer affects dot formation greatly. If the thickness of the separation layers is properly selected, it is well known that dots are aligned vertically, as shown in Fig. 34.41. This is because the strain coming from the underlying dots provides energetically favorable sites for dot formation. However, it should be noted that the kinetics is very complicated and that dot alignment does not always occurs, that is, the dot formation strongly depends on the growth conditions, the dot distribution, and the distance to the underlying dot layer. As a typical example of dot alignment, the distance dependence is shown for high-temperature growth in Fig. 34.42 [34.218]. As seen in this figure, the distribution of dot size and position depends on the Si interlayer thickness. It is natural that the distribution is almost similar to that of single-layer formation when the Si spacer is thick, since the strain effect of the underlying layer does not reach the upper surface. In the sample with a thinner Si spacer layer, a very large size distribution and a drastic increase of dome size appear. However, in the case of intermediate thickness such as 39 nm, it is quite interesting that very uniform distribution and ordering of dots are realized. This tendency can also be understood in terms of the change of the strain distribution due to the underlying dots with spacer thickness.

To enhance the quantum effect of Ge dots, small island formation is favorable. For this purpose, the low-temperature growth where small hut-cluster dots are formed is suitable. However, crystal quality is sacrificed at lower growth temperatures. To overcome this problem, two-step growth or stacking of dots is proposed and highly luminescent dots with relatively small size are successfully formed, as seen in Fig. 34.43 [34.219]. Here, the first layer of Ge dots is formed at a low temperature of 500 °C to obtain small dots and the second layer is grown at a higher temperature of 600 °C

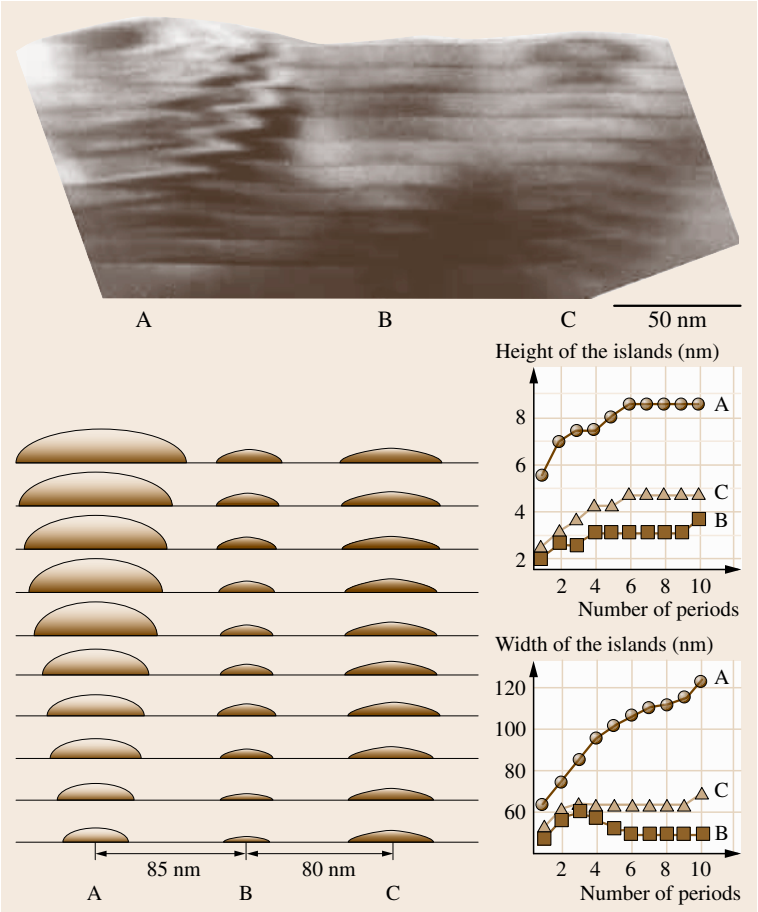


Fig. 34.41 TEM image of stacked Ge dots with 10 nm Si spacer and their schematic illustration. The *right-hand side inset* shows the height and width evolution of the dots as a function of the period number of the Ge layer

to obtain high-quality dots. Thanks to the strain field from the underlying small dots, the size and shape are almost repeated in the second layer, even though the growth is performed at higher temperatures where dome formation is favorable. It is confirmed that these dots have high luminescence efficiency, comparable to that of domes grown at higher temperatures. By using this method, relatively small dots with small size distribution and high luminescence efficiency can be obtained.

However, the size distribution is still large for device applications and the position is not perfectly controlled. Combination of selective epitaxial growth (SEG) and electron-beam (EB) lithography is a promising approach for sufficient control of the dots [34.220, 221]. As mentioned above, GSMBE has the advantage of providing selectivity between Si and SiO_2 surfaces and Ge dots can be grown only on Si surfaces. Therefore, when windows are opened in SiO_2 films on Si substrates, Ge dots are formed only in the windows. If the window size is smaller than the Ge migration length, only one Ge dot grows in a window and the dot size decreases with decreasing window size, as shown in Fig. 34.44 [34.220]. It is noted that the Ge dot is formed on the Ge wetting layer, that is, by the SK growth mode, even in SEG. These controlled Ge dots give rise to luminescence, and two well-resolved peaks are observed in contrast to the disordered Ge dots.

In order to enhance the quantum effects, several attempts to reduce Ge dot size and increase the dot density are now under investigation. Predeposition of elements such as C [34.189] and B [34.222] has been shown to be very effective to reduce the size. Incorporation of C is also found to be effective when dots are formed by using GSMBE [34.223]. Figure 34.45 shows dot formation when $(CH_3)_3SiH$ (TMS) is incorporated into the GeH_4 gas. There are some interesting features, different

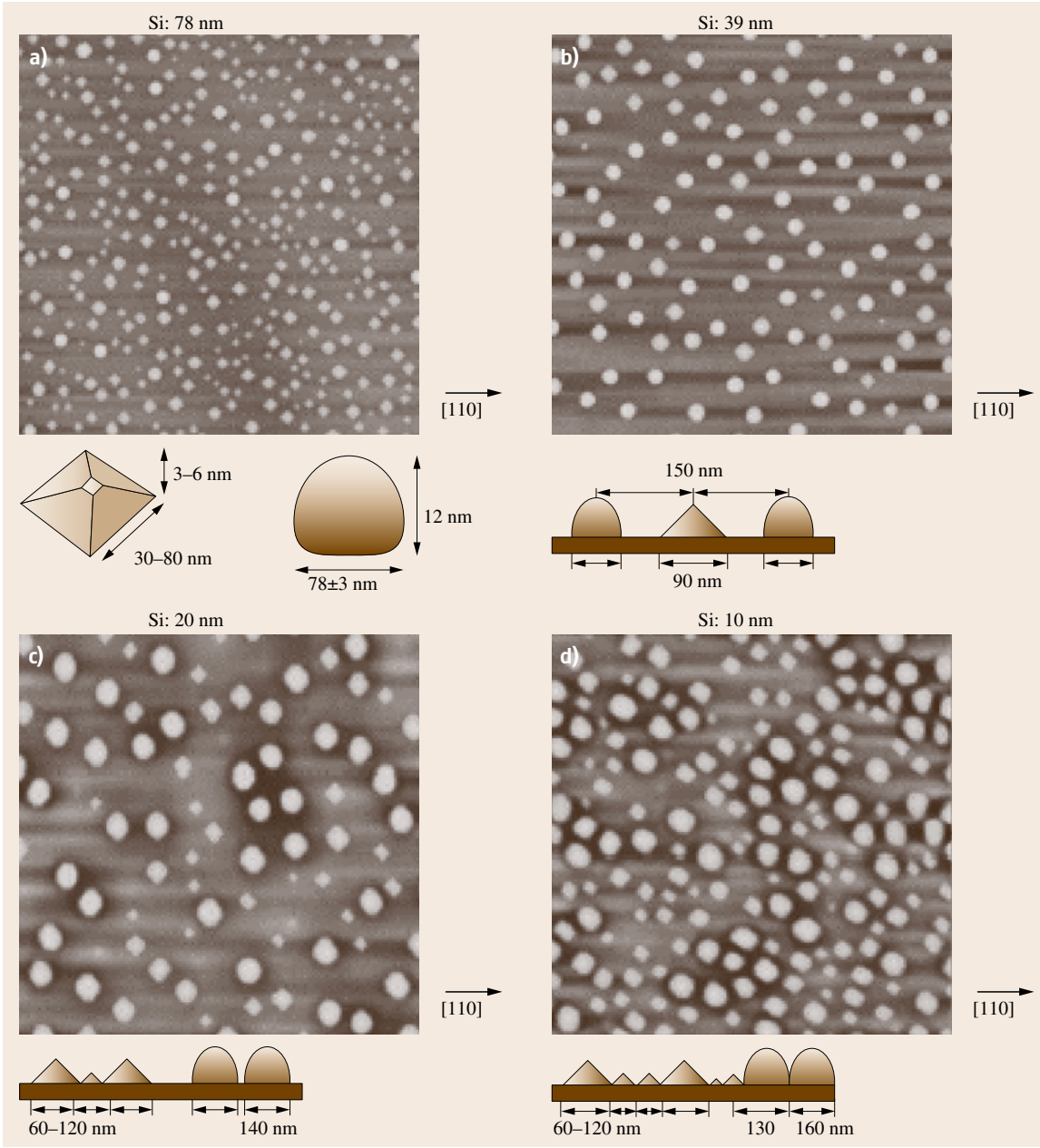


Fig. 34.42a–d AFM images of stacked Ge dots with Si spacer thickness as a parameter, from 78 nm **(a)** to 10 nm **(d)**. Scale is $2 \times 2\,\mu m^2$. Subfigures show schematics of formed Ge dots

from pure-Ge dot formation with GeH_4: (1) the critical thickness for dot formation increases to 7.5 ML from about 4 ML, (2) the dots strongly reduce in size and increase in number by as much as three times that of pure-Ge dots and (3) monomodal formation of dome-like dots occurs instead of the bimodal formation of pure-Ge dots. Dots with C provide luminescence as well, but the peak shift as a function of the deposition is different from that of Ge dots grown at the same growth temperature and very similar to the behavior of Ge dots

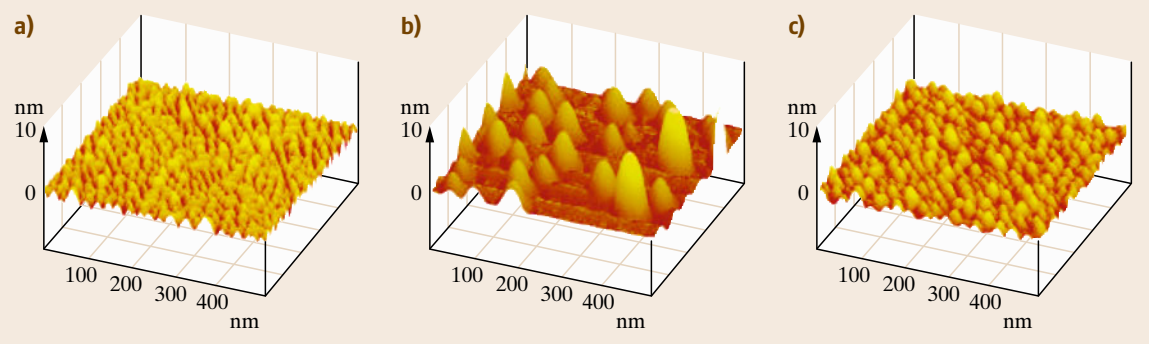


Fig. 34.43a–c AFM results of single Ge layers grown at **(a)** 500 °C and **(b)** 600 °C, and **(c)** stacked layer where the first and second layers are grown at 500 and 600 °C, respectively. The stacked one shows formation of small dots similar to the low-temperature growth one, even though the growth temperature is high ▲

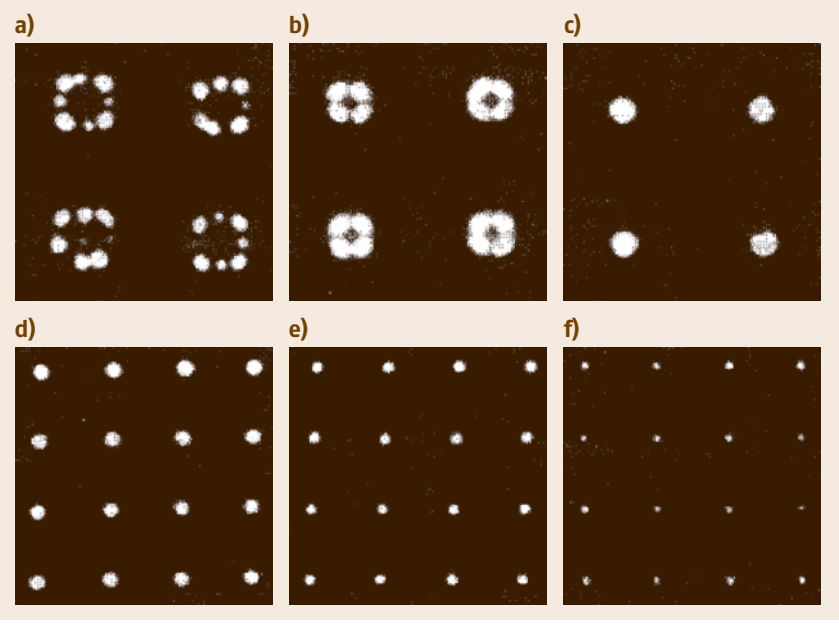


Fig. 34.44a–f SEM images of Ge dots formed in SiO_2 windows on Si substrates. The diameters of the windows are **(a)** 580 nm, **(b)** 440 nm, **(c)** 300 nm, **(d)** 180 nm, **(e)** 130 nm, and **(f)** 90 nm

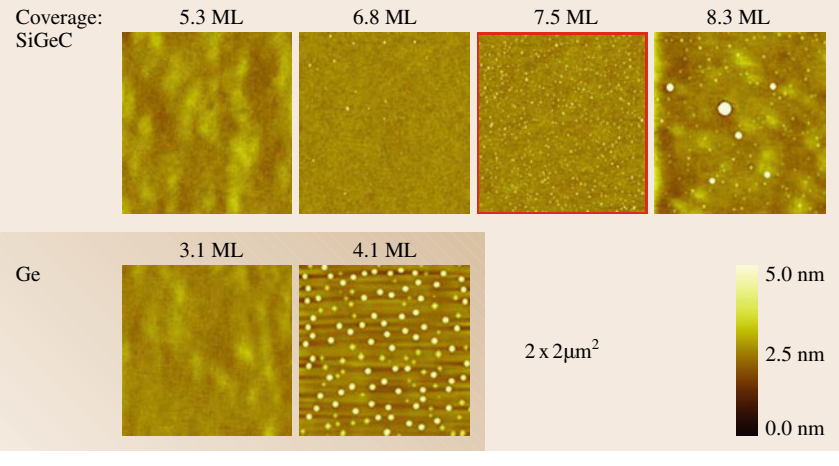


Fig. 34.45 AFM images showing the coverage dependence of dots grown with and without mixing of $(CH_3)_3SiH$ in the GeH_4 gas. Scale is $2 \times 2\,\mu m^2$

grown at lower temperature. This indicates that migration of atoms is impeded by the presence of C atoms, and therefore the low-temperature growth mode occurs even at higher temperatures. This corresponds well to the small dot formation resulting from the reduction of the migration length of atoms on the surface.

34.10 Concluding Remarks and Future Prospects

The Si/Ge system provides a lot of varieties of materials growth, from pseudomorphic growth and strain-relaxed growth to dot formation, and the strain coming from the lattice mismatch between Si and Ge plays an important role in the growth. Comprehensive studies on the growth mechanisms have resulted in the development of novel formation techniques for SiGe heterostructures and enable us to implement strain effects into Si devices.

It is now widely accepted that the performance of Si VLSIs is greatly improved with the aid of SiGe heterostructures and a lot of work on material growth and device design is being conducted. It is therefore forecast that SiGe heterostructures will be implemented in some important VLSI devices soon.

Recently, much attention has also been paid to the photonic properties of Si-related materials, particularly SiGe heterostructures. This is because the performance of VLSIs is improved by adding optical functions such as optical interconnection and parallel signal processing. In order to realize optical functions in Si VLSIs, various kinds of optical elements based on Si, including light sources, should be developed. Among these, photonic crystals are now attracting much attention, since not only can light emission be well controlled but also light waveguides with high flexibility can be realized. If one can apply the photonic crystal to Si devices, waveguides for optical interconnection may be realized on Si VLSIs. Moreover, by combining VLSIs and sophisticated Si, new devices with optical functions and multiprocessing of signals, which is highly desired for such devices as image processors, can be realized based on Si. In this case, SiGe heterostructures will play the main role, and therefore, more intensive studies on new formation techniques for SiGe heterostructures are required.

References

34.1 C.G. Van de Walle, R.M. Martin: Theoretical calculations of heterojunction discontinuities in the Si/Ge system, Phys. Rev. B **34**, 5621–5634 (1986)

34.2 C.G. Van de Walle, R.M. Martin: Theoretical study of band offsets at semiconductor interfaces, Phys. Rev. B **35**, 8154–8165 (1987)

34.3 C.G. Van de Walle: Band lineups and deformation potentials in the model-solid theory, Phys. Rev. B **39**, 1871–1883 (1989)

34.4 H. Kim, N. Taylor, T.R. Bramblett, J.E. Greene: Kinetics of $Si_{1-x}Ge_x$(001) growth on Si(001) 2×1 by gas-source molecular-beam epitaxy from Si_2H_6 and GeH_4, J. Appl. Phys. **84**, 6372–6381 (1998)

34.5 H. Hirayama, T. Tatsumi, A. Ogura, N. Aizaki: Gas source silicon molecular beam epitaxy using silane, Appl. Phys. Lett. **51**, 2213–2215 (1987)

34.6 H. Hirayama, T. Tatsumi, N. Aizaki: Selective growth condition in disilane gas source silicon molecular beam epitaxy, Appl. Phys. Lett. **52**, 2242–2243 (1988)

34.7 M. Suemitsu, F. Hirose, Y. Takakuwa, N. Miyamoto: Growth kinetics in silane gas-source molecular beam epitaxy, J. Cryst. Growth **105**, 203–208 (1990)

34.8 Y. Tsukidate, M. Suemitsu: Infrared study of SiH_4-adsorbed Si(100) surface: Observation and mode assignment of new peaks, Jpn. J. Appl. Phys. **40**, 5206–5210 (2001)

34.9 H. Hirayama, T. Tatsumi, N. Aizaki: Gas source silicon molecular beam epitaxy using disilane, Appl. Phys. Lett. **52**, 1484–1486 (1988)

34.10 D. Lubben, R. Tsu, T.R. Brambelett, J.E. Greene: Mechanisms and kinetics of Si atomic-layer epitaxy on Si(001) 2×1 from Si_2H_6, J. Vac. Sci. Technol. A **9**, 3003–3011 (1991)

34.11 S.H. Li, S.W. Chung, J.K. Rhee, P.K. Bhattacharya: Gas-source molecular-beam epitaxy using Si_2H_6 and GeH_4 and x-ray characterization of $Si_{1-x}Ge_x$ ($0 < x < 0.33$) alloys, J. Appl. Phys. **71**, 4916–4919 (1992)

34.12 T.R. Bramblett, Q. Lu, T. Karasawa, M.-A. Hasan, S.K. Jo, J.E. Greene: Si(001) 2×1 gas-source molecular-beam epitaxy from Si_2H_6: Growth kinetics and boron doping, J. Appl. Phys. **76**, 1884–1888 (1994)

34.13 R. Chelly, T. Angot, P. Louis, D. Bolmont, J.J. Koulmann: In situ monitoring of growth rate param-

eters in hot-wire assisted gas source-molecular beam epitaxy using a quartz microbalance, Appl. Surf. Sci. **115**, 299–306 (1997)
34.14 A.M. Lam, Y.-J. Zheng, J.R. Engstrom: Gas-source reactivity in mixed-crystal systems: The reaction of GeH_4 and Ge_2H_6 on Si surfaces, Surf. Sci. **393**, 205–221 (1997)
34.15 T. Murata, M. Suemitsu: GeH_4 adsorption on Si(001) at RT: Transfer of H atoms to Si sites and atomic exchange between Si and Ge, Appl. Surf. Sci. **224**, 179–182 (2004)
34.16 M. Foster, B. Darlington, J. Scharff, A. Campion: Surface chemistry of alkylsilanes on Si(100) 2×1, Surf. Sci. **375**, 35–44 (1997)
34.17 J. Xu, W.J. Choyke, J.T. Yates Jr.: Role of the $-SiH_3$ functional group in silane adsorption and dissociation on Si(100), J. Phys. Chem. B **101**, 6879–6882 (1997)
34.18 K. Senthil, H. Nakazawa, M. Suemitsu: Adsorption and desorption kinetics of organosilanes at Si(001) surfaces, Jpn. J. Appl. Phys. **42**, 6804–6808 (2003)
34.19 S. Gu, R. Wang, R. Zhang, Y. Zheng: Simulation model to very low pressure chemical vapor deposition of SiGe alloy, J. Vac. Sci. Technol. A **14**, 3256–3260 (1996)
34.20 D.J. Robbins, J.L. Glasper, A.G. Cullis, W.L. Leong: A model for heterogeneous growth of $Si_{1-x}Ge_x$ films from hydrides, J. Appl. Phys. **69**, 3729–3732 (1991)
34.21 B.S. Meyerson, K.J. Uram, F.K. LeGoues: Cooperative growth phenomena in silicon/germanium low-temperature epitaxy, Appl. Phys. Lett. **53**, 2555–2557 (1988)
34.22 J. Murota, S. Ono: Low-temperature epitaxial growth of $Si/Si_{1-x}Ge_x/Si$ heterostructure by chemical vapor deposition, Jpn. J. Appl. Phys. **33**, 2290–2299 (1994)
34.23 J. Murota, T. Matsuura, M. Sakuraba: Atomically controlled processing for group IV semiconductors, Surf. Interface Anal. **34**, 423–431 (2002)
34.24 V. Loup, J.M. Hartmann, G. Rolland, P. Holliger, F. Laugier, C. Vannuffel, M.N. Séméria: Reduced pressure chemical vapor deposition of $Si_{1-x-y}Ge_xC_y$ and $Si_{1-y}C_y/Si$ heterostructures, J. Vac. Sci. Technol. B **20**, 1048–1054 (2002)
34.25 J.M. Hartmann, V. Loup, G. Rolland, M.N. Séméria: Effects of temperature and HCl flow on the SiGe growth kinetics in reduced pressure-chemical vapor deposition, J. Vac. Sci. Technol. B **21**, 2524–2529 (2003)
34.26 P.M. Garone, J.C. Sturm, P.V. Schwartz, S.A. Schwartz, B.J. Wilkens: Silicon vapor phase epitaxial growth catalysis by the presence of germane, Appl. Phys. Lett. **56**, 1275–1277 (1990)
34.27 J.L. Hoyt, C.A. King, D.B. Noble, C.M. Gronet, J.F. Gibbons, M.P. Scott, S.S. Laderman, S.J. Roser, K. Kauka, J. Turner, T.I. Kamins: Limited reaction processing: Growth of $Si_{1-x}Ge_x/Si$ for heterojunction bipolar transistor applications, Thin Solid Films **184**, 93–106 (1990)
34.28 T.O. Sedgwick, D.A. Grutzmacher, A. Zaslavsky, V.P. Kesan: Selective SiGe and heavily As doped Si deposited at low temperature by atmospheric pressure chemical vapor deposition, J. Vac. Sci. Technol. B **11**, 1124–1128 (1993)
34.29 S. Bodnar, E. de Berranger, P. Bouillon, M. Mouis, T. Skotnocki, J.L. Regolini: Selective Si and SiGe epitaxial heterostructures grown using an industrial low-pressure chemical vapor deposition module, J. Vac. Sci. Technol. B **15**, 712–718 (1997)
34.30 W.-C. Wang, J.P. Denton, G.W. Neudeck, I.-M. Lee, C.G. Takoudis, M.T.K. Koh, E.P. Kvam: Selective epitaxial growth of $Si_{1-x}Ge_x/Si$ strained-layers in a tubular hot-wall low pressure chemical vapor deposition system, J. Vac. Sci. Technol. B **15**, 138–141 (1997)
34.31 L. Vescan, K. Grimm, C. Dieker: Facet investigation in selective epitaxial growth of Si and SiGe on (001) Si for optoelectronic devices, J. Vac. Sci. Technol. B **16**, 1549–1554 (1998)
34.32 J.J. Harris, D.E. Ashenford, C.T. Foxon, P.J. Dobson, B.A. Joyce: Kinetic limitations to surface segregation during growth of III–V compounds: Sn in GaAs, Appl. Phys. A **33**, 87–92 (1984)
34.33 R.A. Metzger, F.G. Allen: Evaporative antimony doping of silicon during molecular beam epitaxial growth, J. Appl. Phys. **55**, 931–940 (1984)
34.34 S. Fukatsu: Growth of group-IV semiconductor heterostructures with controlled interfaces and observation of band-edge luminescence from strained SiGe/Si quantum wells. Ph.D. Thesis (University of Tokyo, Tokyo 1992)
34.35 K. Fujita, S. Fukatsu, H. Yaguchi, T. Igarashi, Y. Shiraki, R. Ito: Realization of abrupt interfaces in Si/Ge superlattices by suppressing Ge surface segregation with submonolayer of Sb, Jpn. J. Appl. Phys. **29**, L1981–L1983 (1990)
34.36 K. Fujita, S. Fukatsu, H. Yaguchi, Y. Shiraki, R. Ito: Involvement of the topmost Ge layer in the Ge surface segregation during Si/Ge heterostructure formation, Appl. Phys. Lett. **59**, 2240–2241 (1991)
34.37 S. Fukatsu, K. Fujita, H. Yaguchi, Y. Shiraki, R. Ito: Self-limitation in the surface segregation of Ge atoms during Si molecular beam epitaxial growth, Appl. Phys. Lett. **59**, 2103–2105 (1991)
34.38 J. Ushio, K. Nakagawa, M. Miyao, T. Maruizumi: Surface segregation behavior of Ge in comparison with B, Ga, and Sb: Calculations using a first-principles method, J. Cryst. Growth **201/202**, 81–84 (1999)
34.39 S. Fukatsu, N. Usami, Y. Kato, H. Sunamura, Y. Shiraki, H. Oku, T. Ohnishi, Y. Ohmori, K. Okumura: Gas-source molecular beam epitaxy and luminescence characterization of strained $Si_{1-x}Ge_x/Si$ quantum wells, J. Cryst. Growth **136**, 315–321 (1994)

34.40 M. Copel, C. Reuter, E. Kaxiras, R.M. Tromp: Surfactants in epitaxial growth, Phys. Rev. Lett. **63**, 632–635 (1989)

34.41 F.C. Frank, J.H. van der Merwe: One-dimensional dislocations. I. Static theory, Proc. R. Soc. Lond. Ser. A **198**, 205–216 (1949)

34.42 F.C. Frank, J.H. van der Merwe: One-dimensional dislocation. II. Misfitting monolayers and oriented overgrowth, Proc. R. Soc. Lond. Ser. A **198**, 216–225 (1949)

34.43 F.C. Frank, J.H. van der Merwe: One-dimensional dislocation. III. Influence of the second harmonic term in the potential representation on the properties of the model, Proc. R. Soc. Lond. Ser. A **200**, 125–134 (1949)

34.44 J.H. van der Merwe: Crystal interface. Part I. Semi-infinite crystals, J. Appl. Phys. **34**, 117–122 (1963)

34.45 J.H. van der Merwe: Crystal interface. Part II. Finite overgrowths, J. Appl. Phys. **34**, 123–127 (1963)

34.46 E. Kasper, H.-J. Herzog: Elastic and misfit dislocation density in $Si_{0.92}Ge_{0.08}$ films on silicon substrates, Thin Solid Films **44**, 357–370 (1977)

34.47 E. Kasper: Growth and properties of Si/SiGe superlattices, Surf. Sci. **174**, 630–639 (1986)

34.48 E. Kasper, H.-J. Herzog, H. Daembkes, G. Abstreiter: Equally strained Si/SiGe superlattices on Si substrates, Mater. Res. Soc. Proc. **56**, 347–357 (1986)

34.49 J.W. Matthews, A.E. Blakeslee: Defects in epitaxial multilayers. I. Misfit dislocations, J. Cryst. Growth **27**, 118–125 (1974)

34.50 J.W. Matthews, A.E. Blakeslee: Defects in epitaxial multilayers. II. Dislocation pile-ups, threading dislocations, slip lines and cracks, J. Cryst. Growth **29**, 273–280 (1975)

34.51 J.W. Matthews, A.E. Blakeslee: Defects in epitaxial multilayers. III. Preparation of almost perfect multilayers, J. Cryst. Growth **32**, 265–273 (1976)

34.52 J.C. Bean, L.C. Feldman, A.T. Fiory, S. Nakahara, I.K. Robinson: Ge_xSi_{1-x}/Si strained-layer superlattice grown by molecular beam epitaxy, J. Vac. Sci. Technol. A **2**, 436–440 (1984)

34.53 R. People, J.C. Bean: Calculation of critical layer thickness versus lattice mismatch for Ge_xSi_{1-x}/Si strained-layer heterostructures, Appl. Phys. Lett. **47**, 322–324 (1985)

34.54 I.J. Fritz: Role of experimental resolution in measurements of critical layer thickness for strained-layer epitaxy, Appl. Phys. Lett. **51**, 1080–1082 (1987)

34.55 Y. Kohama, Y. Fukuda, M. Seki: Determination of the critical layer thickness of $Si_{1-x}Ge_x$/Si heterostructures by direct observation of misfit dislocations, Appl. Phys. Lett. **52**, 380–382 (1988)

34.56 D.C. Houghton, C.J. Gibbings, C.G. Tuppen, M.H. Lyons, M.A.G. Halliwell: Equilibrium critical thickness for $Si_{1-x}Ge_x$ strained layers on (100) Si, Appl. Phys. Lett. **56**, 460–462 (1990)

34.57 M.L. Green, B.E. Weir, D. Brasen, Y.F. Hsieh, G. Higashi, A. Feygenson, L.C. Feldman, R.L. Headrick: Mechanically and thermally stable Si-Ge films and heterojunction bipolar transistors grown by rapid thermal chemical vapor deposition at 900 °C, J. Appl. Phys. **69**, 745–751 (1991)

34.58 D.C. Houghton, D.D. Perovic, J.-M. Baribeay, G.C. Weatherly: Misfit strain relaxation in Ge_xSi_{1-x}/Si heterostructures: The structural stability of buried strained layers and strained-layer superlattices, J. Appl. Phys. **67**, 1850–1862 (1990)

34.59 J.W. Matthews, S. Mader, T.B. Light: Accommodation of misfit across the interface between crystals of semiconducting elements or compounds, J. Appl. Phys. **41**, 3800–3804 (1970)

34.60 B.A. Fox, W.A. Jesser: The effect of frictional stress on the calculation of critical thickness in epitaxy, J. Appl. Phys. **68**, 2801–2808 (1990)

34.61 T.J. Gosling, S.C. Jain, J.R. Willis, A. Atkinson, R. Bullough: Stable configurations in strained epitaxial layers, Philos. Mag. A **66**, 119–132 (1992)

34.62 A. Fischer, H. Kuhne, M. Eichler, F. Hollander, H. Richter: Strain and surface phenomena in SiGe structures, Phys. Rev. B **54**, 8761–8768 (1996)

34.63 D.J. Eaglesham, M. Cerullo: Dislocation-free Stranski–Krastanov growth of Ge on Si(100), Phys. Rev. Lett. **16**, 1943–1946 (1990)

34.64 Y.-W. Mo, D.E. Savage, B.S. Swartzentruber, M.G. Lagally: Kinetic pathway in Stranski–Krastanov growth of Ge on Si(001), Phys. Rev. Lett. **65**, 1020–1023 (1990)

34.65 A. Sakai, T. Tatsumi: Defect-mediated island formation in Stranski–Krastanov growth of Ge on Si(001), Phys. Rev. Lett. **71**, 4007–4010 (1993)

34.66 A. Sakai, T. Tatsumi: Defect and island formation in Stranski–Krastanov growth of Ge on Si(001), Mater. Res. Soc. Symp. Proc. **317**, 343–348 (1994)

34.67 R.J. Asaro, W.A. Tiller: Interface morphology development during stress-corrosion cracking. Part I. Via surface diffusion, Metall. Mater. Trans. **3**, 1789–1796 (1972)

34.68 M.A. Grinfeld: Instability of the separation boundary between a non-hydrostatically stressed elastic body and a melt, Sov. Phys. Dokl. **31**, 831–834 (1986)

34.69 D.J. Srolovitz: On the stability of surfaces of stressed solids, Acta Metall. **37**, 621–625 (1989)

34.70 B.J. Spencer, P.W. Voorhees, S.H. Davis: Morphological instability in epitaxially strained dislocation-free solid films, Phys. Rev. Lett. **67**, 3696–3699 (1991)

34.71 J.A. Floro, G.A. Lucadamo, E. Chason, L.B. Freund, M. Sinclair, R.D. Twesten, R.Q. Hwang: SiGe island shape transition induced by elastic repulsion, Phys. Rev. Lett. **80**, 4717–4720 (1998)

34.72 G. Medeiros-Ribeiro, A.M. Bratkovski, T.I. Kamins, D.A.A. Ohlberg, R.S. Williams: Shape transition of germanium nanocrystals on a silicon (001) surface from pyramids to domes, Science **279**, 353–355 (1998)

34.73 B.W. Dodson, J.Y. Tsao: Relaxation of strained-layer semiconductor structures via plastic flow, Appl. Phys. Lett. **51**, 1325–1327 (1987)

34.74 H. Alexander, P. Haasen: Dislocations and plastic flow in the diamond structure, Solid State Phys. **22**, 22–158 (1968)

34.75 E. Kasper, H.J. Herzog, H. Kibbel: One-dimensional SiGe superlattice grown by UHV epitaxy, Appl. Phys. **8**, 199–205 (1975)

34.76 R. Hull, J.C. Bean, C. Büscher: A phenomenological description of strain relaxation in Ge_xSi_{1-x}/Si(100) heterostructures, J. Appl. Phys. **66**, 5837–5843 (1989)

34.77 D.D. Perovic, G.C. Whetherly, J.M. Baribeau, D.C. Houghton: Heterogeneous nucleation sources in molecular beam epitaxy-grown $Ge_{1-x}Si_x$/Si strained layer superlattices, Thin Solid Films **183**, 141–156 (1989)

34.78 D.J. Eaglesham, E.P. Kvam, D.M. Maher, C.J. Humphrey, J.C. Bean: Dislocation nucleation near the critical thickness in GeSi/Si strained layers, Philos. Mag. A **59**, 1059–1073 (1989)

34.79 C.J. Humphreys, D.M. Maher, D.J. Eagleshum, E.P. Kvam, I.G. Salisbury: The origin of dislocations in multilayers, J. Phys. III **1**, 1119–1130 (1991)

34.80 V. Higgs, P. Kightley, P. Goodhew, P. Augustus: Metal-induced dislocation nucleation for metastable SiGe/Si, Appl. Phys. Lett. **59**, 829–831 (1991)

34.81 M.D. de Coteau, P.R. Wilshaw, R. Falster: Gettering of copper in silicon – precipitation at extended surface-defects, Inst. Phys. Conf. Ser. **117**, 231–234 (1991)

34.82 R. Hull, J.C. Bean: Nucleation of misfit dislocations in strained-layer epitaxy in the Ge_xSi_{1-x}/Si system, J. Vac. Sci. Technol. A **7**, 2580–2585 (1989)

34.83 U. Jain, S.C. Jain, A.H. Harker, R. Bullough: Nucleation of dislocation loops in strained epitaxial layers, J. Appl. Phys. **77**, 103–109 (1995)

34.84 D.E. Jesson, S.J. Pennycook, J.-M. Bribeau, D.C. Houghton: Surface stress, morphological development, and dislocation nucleation during $Si_{1-x}Ge_x$ epitaxy, Scanning Microsc. **8**, 849–857 (1994)

34.85 D.D. Perovic, D.C. Houghton: The introduction of dislocations in low misfit epitaxial systems, Microsc. Semicond. Mater. 1995, Inst. Phys. Conf. Ser. **146**, 117–134 (1995)

34.86 E.A. Stach, R. Hull, R.M. Tromp, M.C. Reuter, M. Copel, F.K. LeGoues, J.C. Bean: Effect of the surface upon misfit dislocation velocities during the growth and annealing of SiGe/Si (001) heterostructures, J. Appl. Phys. **83**, 1931–1937 (1998)

34.87 D.C. Houghton, C.J. Gibbings, C.G. Tuppen, M.H. Lyons, M.A.G. Halliwell: The structural stability of uncapped versus buried $Si_{1-x}Ge_x$ strained layers through high temperature processing, Thin Solid Films **183**, 171–182 (1989)

34.88 D.C. Houghton: Strain relaxation kinetics in $Si_{1-x}Ge_x$/Si heterostructures, J. Appl. Phys. **70**, 2136–2151 (1991)

34.89 H. Hull, J.C. Bean, D. Bahnck, L.J. Peticolas, K.T. Short, F.C. Unterwald: Interpretation of dislocation propagation velocities in strained Ge_xSi_{1-x}/Si(100) heterostructures by the diffusive kink pair model, J. Appl. Phys. **70**, 2052–2065 (1991)

34.90 Y. Yamashita, K. Maeda, K. Fujita, N. Usami, K. Suzuki, S. Fukatsu, Y. Mera, Y. Shiraki: Dislocation glide motion in heteroepitaxial thin-films of $Si_{1-x}Ge_x$/Si(100), Philos. Mag. Lett. **67**, 165–171 (1993)

34.91 R. Hull, J.C. Bean: New insights into the microscopic motion of dislocations in covalently bonded semiconductors by in-situ transmission electron microscope observations of misfit dislocations in thin strained epitaxial layers, Phys. Status Solidi (a) **138**, 533–546 (1993)

34.92 W. Hagen, H. Strunk: New type of source generating misfit dislocations, Appl. Phys. **17**, 85–87 (1978)

34.93 R. Hull, J.C. Bean, D.J. Eaglesham, J.N. Bonara, C. Büscher: Strain relaxation phenomena in Ge_xSi_{1-x}/Si strained structures, Thin Solid Films **183**, 117–132 (1989)

34.94 A. Lefebvre, C. Herbeaux, J. Di Persio: Interactions of misfit dislocations in $In_xGa_{1-x}As$/GaAs interfaces, Philos. Mag. A **63**, 471–485 (1991)

34.95 F.K. LeGoues, B.S. Meyerson, J.F. Morar, P.D. Kirchner: Mechanism and conditions for anomalous strain relaxation in grade thin films and superlattices, J. Appl. Phys. **71**, 4230–4243 (1992)

34.96 J. Washburn, E.P. Kvam: Possible dislocation multiplication source in (001) semiconductor epitaxy, Appl. Phys. Lett. **57**, 1637–1639 (1990)

34.97 C.G. Tuppen, C.J. Gibbings, M. Hockly, S.G. Roberts: Misfit dislocation multiplication processes in $Si_{1-x}Ge_x$ alloys for $x < 0.15$, Appl. Phys. Lett. **56**, 54–56 (1990)

34.98 Y.J. Mii, Y.-H. Xie, E.A. Fitzgerald, F.B.E. Weir, L.C. Feldman: Extremely high electron mobility in Si/Ge_xSi_{1-x} structures grown by molecular beam epitaxy, Appl. Phys. Lett. **59**, 1611–1613 (1991)

34.99 J.J. Welser, J.L. Hoyt, S. Takagi, J.F. Gibbons: Strain dependence of the performance enhancement in strained-Si n-MOSFETs, Tech. Dig. Int. Electron Device Meet. (1994) pp. 373–376

34.100 K. Ismail, M. Arafa, K.L. Saenger, J.O. Chu, B.S. Meyerson: Extremely high electron mobility in Si/SiGe modulation-doped heterostructures, Appl. Phys. Lett. **66**, 1077–1079 (1995)

34.101 A.C. Churchill, D.J. Robbins, D.J. Wallis, N. Griffin, D.J. Paul, A.J. Pidduck: High-mobility two-dimensional electron gases in Si/SiGe heterostructures on relaxed SiGe layers grown at high temperature, Semicond. Sci. Technol. **12**, 943–946 (1997)

34.102 S. Bozzo, J.-L. Lazzari, C. Coudreau, A. Ronda, F. Arnaud d'Avitaya, J. Derrien, S. Mesters, B. Holländer, P. Gergaud, O. Thomas: Chemical vapor deposition of silicon-germanium heterostructures, J. Cryst. Growth **216**, 171–184 (2000)
34.103 K.H. Chang, R. Gibala, D.J. Srolovitz, P.K. Bhattacharya, J.F. Mansfield: Crosshatched surface morphology in strained III–V semiconductor films, J. Appl. Phys. **67**, 4093–4098 (1990)
34.104 A.J. Pidduck, D.J. Robbins, A.G. Cullis, W.Y. Leong, A.M. Pitt: Evolution of surface morphology and strain during SiGe epitaxy, Thin Solid Films **222**, 78–84 (1992)
34.105 J.W.P. Hsu, E.A. Fitzgerald, Y.-H. Xie, P.J. Silverman, M.J. Cardillo: Surface morphology of related Ge_xSi_{1-x} films, Appl. Phys. Lett. **61**, 1293–1295 (1992)
34.106 E.A. Fitzgerald, Y.-H. Xie, D. Monroe, P.J. Silverman, J.M. Kuo, A.R. Kortan, F.A. Thiel, B.E. Weir: Relaxed Ge_xSi_{1-x} structures for III–V integration with Si and high mobility two-dimensional electron gasses in Si, J. Vac. Sci. Technol. B **10**, 1807–1819 (1992)
34.107 S.Y. Shiryaev, F. Jensen, J.W. Petersen: On the nature of cross-hatch patterns on compositionally graded $Si_{1-x}Ge_x$ alloy layers, Appl. Phys. Lett. **64**, 3305–3307 (1994)
34.108 M.A. Lutz, R.M. Feenstra, F.K. LeGoues, P.M. Mooney, J.O. Chu: Influence of misfit dislocations on the surface morphology of $Si_{1-x}Ge_x$ films, Appl. Phys. Lett. **66**, 724–726 (1995)
34.109 A.M. Andrews, J.S. Speck, A.E. Romanov, M. Bobeth, W. Pompe: Modeling cross-hatch surface morphology in growing mismatched layers, J. Appl. Phys. **91**, 1933–1943 (2002)
34.110 E.A. Fitzgerald, S.B. Samavedam: Line, point and surface defect morphology of graded, relaxed GeSi alloys of Si substrates, Thin Solid Films **294**, 3–10 (1997)
34.111 C.W. Leitz, M.T. Currie, A.Y. Kim, J. Lai, E. Robbins, E.A. Fitzgerald, M.T. Bulsara: Dislocation glide and blocking kinetics in compositionally graded SiGe/Si, J. Appl. Phys. **90**, 2730–2736 (2001)
34.112 H. Chen, L.W. Guo, Q. Cui, Q. Hu, Q. Huang, J.M. Zhou: Low-temperature buffer layer for growth of a low-dislocation-density SiGe layer on Si by molecular-beam epitaxy, J. Appl. Phys. **79**, 1167–1169 (1996)
34.113 J.H. Li, C.S. Peng, Y. Wu, D.Y. Dai, J.M. Zhou, Z.H. Mai: Relaxed $Si_{0.7}Ge_{0.3}$ layers grown on low-temperature Si buffers with low threading dislocation density, Appl. Phys. Lett. **71**, 3132–3134 (1997)
34.114 T. Ueno, T. Irisawa, Y. Shiraki: p-type Ge channel modulation doped heterostructures with very high room-temperature mobilities, Physica E **7**, 790–794 (2000)
34.115 T. Ueno, T. Irisawa, Y. Shiraki, A. Uedono, S. Tanigawa, R. Suzuki, T. Ohdaira, T. Mikado: Characterization of low temperature grown Si layer for SiGe pseudo-substrates by positron annihilation spectroscopy, J. Cryst. Growth **227–228**, 761–765 (2001)
34.116 K. Sawano, K. Kawaguchi, T. Ueno, S. Koh, K. Nakagawa, Y. Shiraki: Surface smoothing of SiGe strain-relaxed buffer layers by chemical mechanical polishing, Mater. Sci. Eng. B **89**, 406–409 (2002)
34.117 K. Sawano, Y. Hirose, S. Koh, K. Nakagawa, T. Hattori, Y. Shiraki: Mobility enhancement in strained Si modulation-doped structures by chemical mechanical polishing, Appl. Phys. Lett. **82**, 412–414 (2003)
34.118 K. Sawano, K. Arimoto, Y. Hirose, S. Koh, N. Usami, K. Nakagawa, T. Hattori, Y. Shiraki: Planarization of SiGe virtual substrates by CMP and its application to strained Si modulation-doped structures, J. Cryst. Growth **251**, 693 (2003)
34.119 K. Sawano, K. Kawaguchi, S. Koh, Y. Hirose, T. Hattori, K. Nakagawa, Y. Shiraki: Surface planarization of strain-relaxed SiGe buffer layers by CMP and post cleaning, J. Electrochem. Soc. **150**, G376–G379 (2003)
34.120 K. Sawano, Y. Abe, H. Satoh, K. Nakagawa, Y. Shiraki: Mobility enhancement in strained Ge heterostructures by planarization of SiGe buffer layers grown on Si substrates, Jpn. J. Appl. Phys. **44**, L1320–L1322 (2005)
34.121 K. Sawano, N. Usami, K. Arimoto, S. Koh, K. Nakagawa, Y. Shiraki: Observation of strain field fluctuation in SiGe relaxed buffer layers and its influence on overgrown structures, Mater. Sci. Semicon. Process. **8**, 177–180 (2005)
34.122 R. Hull, J.C. Bean, J.M. Bonar, G.S. Higashi, K.T. Short, H. Temkin, A.E. White: Enhanced strain relaxation in $Si/Ge_xSi_{1-x}/Si$ heterostructures via point-defect concentrations introduced by ion implantation, Appl. Phys. Lett. **56**, 2445–2447 (1990)
34.123 B. Holländer, S. Mantl, R. Liedtke, S. Mesters, H.-J. Herzog, H. Kibbel, T. Hackbarth: Enhanced strain relaxation of epitaxial SiGe layers on Si(100) after H^+ ion implantation, Nucl. Instrum. Methods B **148**, 200–205 (1999)
34.124 H. Trinkaus, B. Holländer, S. Rongen, S. Mantl, H.-J. Herzog, J. Kuchenbecher, T. Hackbarth: Strain relaxation mechanism for hydrogen-implanted $Si_{1-x}Ge_x/Si(100)$ heterostructures, Appl. Phys. Lett. **76**, 3552–3554 (2000)
34.125 M. Luysberg, D. Kirch, H. Trinkaus, B. Holländer, S. Lenk, S. Mantl, H.-J. Herzog, T. Hackbarth, P.F.P. Fichtner: Effect of helium ion implantation and annealing on the relaxation behavior of pseudomorphic $Si_{1-x}Ge_x$ buffer layers on Si (100) substrates, J. Appl. Phys. **92**, 4290–4295 (2002)
34.126 L.-F. Zou, Z.G. Wang, D.Z. Sun, T.W. Fan, X.F. Liu, J.W. Zhang: Characterization of strain relaxation in As ion implanted $Si_{1-x}Ge_x$ epilayers grown by gas

source molecular beam epitaxy, Appl. Phys. Lett. **72**, 845–847 (1998)
34.127 K. Sawano, Y. Hirose, S. Koh, K. Nakagawa, T. Hattori, Y. Shiraki: Relaxation enhancement of SiGe thin layers by ion implantation into Si substrates, J. Cryst. Growth **251**, 685–688 (2003)
34.128 K. Sawano, Y. Hirose, Y. Ozawa, S. Koh, J. Yamanaka, K. Nakagawa, T. Hattori, Y. Shiraki: Enhancement of strain relaxation of SiGe thin layers by pre-ion-implantation into Si substrates, Jpn. J. Appl. Phys. **42**, L735–L737 (2003)
34.129 K. Sawano, Y. Hirose, S. Koh, K. Nakagawa, T. Hattori, Y. Shiraki: Formation of thin SiGe virtual substrates by ion implantation into Si substrates, Appl. Surf. Sci. **224**, 99–103 (2004)
34.130 K. Sawano, S. Koh, Y. Shiraki, Y. Ozawa, T. Hattori, J. Yamanaka, K. Suzuki, K. Arimoto, K. Nakagawa, N. Usami: Fabrication of high-quality strain-relaxed thin SiGe layers on ion-implanted Si substrates, Appl. Phys. Lett. **85**, 2514–2516 (2004)
34.131 K. Sawano, Y. Ozawa, A. Fukumoto, N. Usami, J. Yamanaka, K. Suzuki, K. Arimoto, K. Nakagawa, Y. Shiraki: Strain-field evaluation of strain-relaxed thin SiGe layers fabricated by ion implantation method, Jpn. J. Appl. Phys. **44**, L1316–L1319 (2005)
34.132 T. Tezuka, N. Sugiyama, S. Takagi: Fabrication of strained Si on an ultrathin SiGe-on-insulator virtual substrate with a high-Ge fraction, Appl. Phys. Lett. **79**, 1798–1800 (2001)
34.133 T. Mizuno, N. Sugiyama, T. Tezuka, S. Takagi: Relaxed SiGe-on-insulator substrates without thick SiGe buffer layers, Appl. Phys. Lett. **80**, 601–603 (2002)
34.134 T. Tezuka, N. Sugiyama, S. Takagi, T. Kawakubo: Dislocation-free formation of relaxed SiGe-on-insulator layers, Appl. Phys. Lett. **80**, 3560–3562 (2002)
34.135 N. Sugii, S. Yamaguchi, K. Washio: SiGe-on-insulator substrate fabricated by melt solidification for a strained-silicon complementary metal-oxide-semiconductor, J. Vac. Sci. Technol. B **20**, 1891–1896 (2002)
34.136 T. Mizuno, N. Sugiyama, T. Tezuka, T. Numata, S. Takagi: High performance CMOS operation of strained-SOI MOSFETs using thin film SiGe-on-insulator substrate, Dig. Tech. Pap. Symp. VLSI Technology 2002, pp. 106–107
34.137 S. Nakaharai, T. Tetsuka, N. Sugiyama, Y. Moriyama, S. Takagi: Characterization of 7-nm-thick strained Ge-on-insulator layer fabricated by Ge-condensation technique, Appl. Phys. Lett. **83**, 3516–3518 (2003)
34.138 T. Egawa, A. Sakai, T. Yamamoto, N. Taoka, O. Nakatsuka, S. Zaima, Y. Yasuda: Strain-relaxation mechanisms of SiGe layers formed by two-step growth on Si(001) substrates, Appl. Surf. Sci. **224**, 104–107 (2004)
34.139 A. Sakai, K. Sugimoto, T. Yamamoto, M. Okada, H. Ikeda, Y. Yasuda, S. Zaima: Reduction of threading dislocation density in SiGe layers on Si(001) using a two-step strain-relaxation procedure, Appl. Phys. Lett. **79**, 3398–3400 (2001)
34.140 A. Sakai, T. Tatsumi, K. Aoyama: Growth of strain-relaxed Ge films on Si(001) surfaces, Appl. Phys. Lett. **71**, 3510–3512 (1997)
34.141 N. Ikarashi, T. Tatsumi: Suppression of surface roughening on strained Si/SiGe layers by lowering surface stress, Jpn. J. Appl. Phys. **36**, L377–L379 (1997)
34.142 Y.H. Xie, S.B. Samavedam, M. Bulsara, T.A. Langdo, E.A. Fitzgerald: Relaxed template for fabricating regularly distributed quantum dot arrays, Appl. Phys. Lett. **71**, 3567–3568 (1997)
34.143 D.E. Jesson, S.J. Pennycook, J.-M. Baribeau, D.C. Houghton: Direct imaging of surface cusp evolution during strained-layer epitaxy and implications for strain relaxation, Phys. Rev. Lett. **71**, 1744–1747 (1993)
34.144 P.M. Mooney, F.K. LeGoues, J.O. Chu, S.F. Nelson: Strain relaxation and mosaic structure in relaxed SiGe layers, Appl. Phys. Lett. **62**, 3464–3466 (1993)
34.145 N. Taoka, A. Sakai, T. Egawa, O. Nakatsuka, S. Zaima, Y. Yasuda: Growth and characterization of strain-relaxed SiGe buffer layers on Si(001) substrates with pure-edge misfit dislocations, Mater. Sci. Semicond. Process. **8**, 131–135 (2005)
34.146 T. Yamamoto, A. Sakai, T. Egawa, N. Taoka, O. Nakatsuka, S. Zaima, Y. Yasuda: Dislocation structures and strain-relaxation in SiGe buffer layers on Si (001) substrates with an ultra-thin Ge interlayer, Appl. Surf. Sci. **224**, 108–112 (2004)
34.147 P.F. Fewster: *X-Ray Scattering from Semiconductors* (Imperial College Press, World Scientific, Singapore 2000)
34.148 S.S. Iyer, K. Eberl, M.S. Goorsky, F.K. LeGoues, J.C. Tsang, F. Cardone: Synthesis of $Si_{1-y}C_y$ alloys by molecular beam epitaxy, Appl. Phys. Lett. **60**, 356–358 (1992)
34.149 K. Eberl, S.S. Iyer, J.C. Tsang, M.S. Goorsky, F.K. LeGoues: The growth and characterization of $Si_{1-y}C_y$ alloys on Si(001) substrate, J. Vac. Sci. Technol. B **10**, 934–936 (1992)
34.150 K. Eberl, S.S. Iyer, S. Zollner, J.C. Tsang, F.K. LeGoues: Growth and strain compensation effects in the ternary $Si_{1-x-y}Ge_xC_y$ alloy system, Appl. Phys. Lett. **60**, 3033–3035 (1992)
34.151 B. Cordero, V. Gomez, A.E. Platero-Prats, M. Reves, J. Echeverria, E. Cremades, F. Barragan, S. Alvarez: Covalent radii revisited, J. Chem. Dalton Trans., 2832–2838 (2008)
34.152 H.-J. Osten, E. Bugiel, P. Zaumseil: Growth of an inverse tetragonal distorted SiGe layer on Si(001) by adding small amounts of carbon, Appl. Phys. Lett. **64**, 3440–3442 (1994)

34.153 P.C. Kelires: Monte Carlo studies of ternary semiconductor alloys: Application to the $Si_{1-x-y}Ge_xC_y$ system, Phys. Rev. Lett. **75**, 1114–1117 (1995)

34.154 P.C. Kelires: Short-range order, bulk moduli, and physical trends in c-$Si_{1-x}C_x$ alloys, Phys. Rev. B **55**, 8784–8787 (1997)

34.155 M. Berti, D. De Salvador, A.V. Drigo, F. Romanato, J. Stangl, S. Zerlauth, F. Schäffler, G. Bauer: Lattice parameter in $Si_{1-y}C_y$ epilayers: Deviation from Vegard's rule, Appl. Phys. Lett. **72**, 1602–1604 (1998)

34.156 H.-J. Osten, B. Heinemann, D. Knoll, G. Lippert, H. Rücker: Effects of carbon on boron diffusion in SiGe: Principles and impact on bipolar devices, J. Vac. Sci. Technol. B **16**, 1750–1753 (1998)

34.157 H. Rücker, B. Heinemann, K.D. Bolze, D. Knoll, D. Krüger, R. Kurps, H.-J. Osten, P. Schley, B. Tillack, P. Zaumseil: Dopant diffusion in C-doped Si and SiGe: Physical model and experimental verification, Tech. Dig. Int. Electron Device Meet. (1999) pp. 345–348

34.158 A. Biswas, P.K. Basu: Estimated effect of germanium and carbon on the early voltage of a $Si_{1-x-y}Ge_xC_y$ heterojunction bipolar transistor, Semicond. Sci. Technol. **16**, 947–953 (2001)

34.159 K. Oda, E. Ohue, I. Suzumura, R. Hayami, A. Kodama, H. Simamoto, K. Washio: High performance self-aligned SiGeC HBT with selectively grown $Si_{1-x-y}Ge_xC_y$ base by UHV/CVD, IEEE. Trans. Electron. Dev. **50**, 2213–2220 (2003)

34.160 P. Boucaud, C. Francis, F.H. Julien, J.-M. Lourtioz, D. Bouchier, S. Bodnar, B. Lambert, J.L. Regolini: Band-edge and deep level photoluminescence of pseudomorphic $Si_{1-x-y}Ge_xC_y$ alloys, Appl. Phys. Lett. **64**, 875–877 (1994)

34.161 K. Brunner, W. Winter, K. Eberl: Spatially indirect radiative recombination of carriers localized in $Si_{1-x-y}Ge_xC_y/Si_{1-y}C_y$ double quantum well structure on Si substrates, Appl. Phys. Lett. **69**, 1279–1281 (1996)

34.162 B.A. Orner, J. Olowolafe, K. Roe, J. Kolodzey, T. Laursen, J.W. Mayer, J. Spear: Band gap of Ge rich $Si_{1-x-y}Ge_xC_y$ alloys, Appl. Phys. Lett. **69**, 2557–2559 (1996)

34.163 B.A. Orner, J. Kolodzey: $Si_{1-x-y}Ge_xC_y$ alloy band structures by linear combination of atomic orbitals, J. Appl. Phys. **81**, 6773–6780 (1997)

34.164 B.L. Stein, E.T. Yu, E.T. Croke, A.T. Hunter, T. Laursen, A.E. Bair, J.W. Mayer, C.C. Ahn: Band offsets in $Si/Si_{1-x-y}Ge_xC_y$ heterojunctions measures by admittance spectroscopy, Appl. Phys. Lett. **70**, 3413–3415 (1997)

34.165 O.G. Schmidt, K. Eberl: Photoluminescence of tensile strained, exactly strain compensated, and compressively strained $Si_{1-x-y}Ge_xC_y$ layers in Si, Phys. Rev. Lett. **80**, 3396–3399 (1998)

34.166 D.V. Singh, K. Rim, T.O. Mitchell, J.L. Hoyt, J.F. Gibbons: Admittance spectroscopy analysis of the conduction band offsets in $Si/Si_{1-x-y}Ge_xC_y$ and $Si/Si_{1-y}C_y$ heterostructures, J. Appl. Phys. **85**, 985–993 (1999)

34.167 C.L. Chang, L.P. Rokhinson, J.C. Sturm: Direct optical measurement of the valence band offset of $p^+Si_{1-x-y}Ge_xC_y/p^-Si(100)$ by heterojunction internal photoemission, Appl. Phys. Lett. **73**, 3568–3570 (1998)

34.168 H.-J. Osten: Band-gap changes and band offsets for ternary $Si_{1-x-y}Ge_xC_y$ alloys on Si(001), J. Appl. Phys. **84**, 2716–2721 (1998)

34.169 K. Brunner, O.G. Schmidt, W. Winter, K. Eberl, M. Glück, U. König: SiGeC: Band gaps, band offsets, optical properties, and potential applications, J. Vac. Sci. Technol. B **16**, 1701–1706 (1998)

34.170 S. Galdin, P. Dollfus, V.A. Fortuna, P. Hesto, H.J. Osten: Band offset predictions for strained group IV alloys: $Si_{1-x-y}Ge_xC_y$ on Si(001) and $Si_{1-x}Ge_x$ on $Si_{1-z}Ge_z(001)$, Semicond. Sci. Technol. **15**, 565–572 (2000)

34.171 S. John, S.K. Ray, E. Quinones, S.K. Oswal, K. Banerjee: Heterostructure P-channel metal–oxide–semiconductor transistor utilizing a $Si_{1-x-y}Ge_xC_y$ channel, Appl. Phys. Lett. **74**, 847–849 (1999)

34.172 E. Cassan, P. Dollfus, S. Galdin: Effect of doping profile on the potential performance of buried channel SiGeC/Si heterostructure MOS devices, Physica E **13**, 957–960 (2002)

34.173 M. Glück, U. König, W. Winter, K. Brunner, K. Eberl: Modulation-doped $Si_{1-x-y}Ge_xC_y$ p-type hetero-FETs, Physica E **2**, 768–771 (1998)

34.174 R.A. Soref: Silicon-based group IV heterostructures for optoelectronic applications, J. Vac. Sci. Technol. A **14**, 913–918 (1996)

34.175 A.S. Amour, L.D. Lanzerotti, C.L. Chang, J.C. Sturm: Optical and electrical properties of $Si_{1-x-y}Ge_xC_y$ thin films and devices, Thin Solid Films **294**, 112–117 (1997)

34.176 O. Madelung (Ed.): *Semiconductors-Basic Data*, 2nd edn. (Springer, Berlin 1996) p. 22

34.177 M. Okinaka, Y. Hamana, T. Tokuda, J. Ohta, M. Nunoshita: MBE growth mode and C incorporation of GeC epilayers on Si(001) substrates using an arc plasma gun as a novel C source, J. Cryst. Growth **249**, 78–86 (2003)

34.178 M.W. Dashiell, L.V. Kulik, D.A. Hits, J. Kolodzey, G. Watson: Carbon incorporation in $Si_{1-y}C_y$ alloys grown by molecular beam epitaxy using a single silicon-graphite source, Appl. Phys. Lett. **72**, 833–835 (1998)

34.179 J.P. Liu, H.-J. Osten: Substitutional carbon incorporation during $Si_{1-x-y}Ge_xC_y$ growth on Si(100) by molecular-beam epitaxy: Dependence on germanium and carbon, Appl. Phys. Lett. **76**, 3546–3548 (2000)

34.180 Y. Kanzawa, K. Nozawa, T. Saitoh, M. Kubo: Dependence of substitutional C incorporation on Ge content for $Si_{1-x-y}Ge_xC_y$ crystals grown by ul-

trahigh vacuum chemical vapor deposition, Appl. Phys. Lett. **77**, 3962–3964 (2000)
34.181 V. LeThanh, C. Calmes, Y. Zheng, D. Bouchier, V. Fortuna, J.-C. Dupuy: In situ RHEED monitoring of carbon incorporation during SiGeC/Si(001) growth in a UHV-CVD system, Mater. Sci. Eng. B **89**, 246–251 (2002)
34.182 J. Mi, P. Warren, P. Letourneau, M. Judelewicz, M. Gailhanou, M. Dutoit, C. Dubois, J.C. Dupuy: High quality $Si_{1-x-y}Ge_xC_y$ epitaxial layers grown on (100) Si by rapid thermal chemical vapor deposition using methylsilane, Appl. Phys. Lett. **67**, 259–261 (1995)
34.183 W.K. Choi, J.H. Chen, L.K. Bera, W. Feng, K.L. Pey, J. Mi, C.Y. Yang, A. Ramam, S.J. Chua, J.S. Pan, A.T.S. Wee, R. Liu: Structural characterization of rapid thermal oxidized $Si_{1-x-y}Ge_xC_y$ alloy films grown by rapid thermal chemical vapor deposition, J . Appl. Phys. **87**, 192–197 (2000)
34.184 V. Loup, J.M. Hartmann, G. Rolland, P. Holliger, F. Laugier, M.N. Séméria: Growth temperature dependence of substitutional carbon incorporation in SiGeC/Si heterostructures, J. Vac. Sci. Technol. B **21**, 246–253 (2003)
34.185 J. Tersof: Enhanced solubility of impurities and enhanced diffusion near crystal surfaces, Phys. Rev. Lett. **74**, 5080–5083 (1995)
34.186 H. Rücker, M. Methfessel, E. Bugiel, H.-J. Osten: Strain-stabilized highly concentrated pseudomorphic $Si_{1-x}C_x$ layers in Si, Phys. Rev. Lett. **72**, 3578–3581 (1994)
34.187 P.C. Kelires: Theoretical investigation of the equilibrium surface structure of $Si_{1-x-y}Ge_xC_y$alloys, Surf. Sci. **418**, L62–L67 (1998)
34.188 H. Jacobson, J. Xiang, N. Herbots, S. Whaley, P. Ye, S. Hearne: Heteroepitaxial properties of $Si_{1-x-y}Ge_xC_y$ on Si(001) grown by combined ion- and molecular-beam deposition, J. Appl. Phys. **81**, 3081–3091 (1997)
34.189 A. Sakai, Y. Torige, M. Okada, H. Ikeda, Y. Yasuda, S. Zaima: Atomistic evolution of $Si_{1-x-y}Ge_xC_y$ thin films on Si(001) surfaces, Appl. Phys. Lett. **79**, 3242–3244 (2001)
34.190 O.G. Schmidt, C. Lange, K. Eberl, O. Kienzle, F. Ernst: Formation of carbon-induced germanium dots, Appl. Phys. Lett. **71**, 2340–2342 (1997)
34.191 O. Leifeld, A. Beyer, E. Müller, K. Kern, D. Grützmacher: Formation and ordering effects of C-induced Ge dots grown on Si(001) by molecular beam epitaxy, Mater. Sci. Eng. B **74**, 222–228 (1999)
34.192 R.I.G. Uhrberg, J.E. Northrup, D.K. Biegelsen, R.D. Bringans, L.-E. Swartz: Atomic structure of the metastable $c(4\times4)$ reconstruction of Si(100), Phys. Rev. B **46**, 10251–10256 (1992)
34.193 H. Nörenberg, G.A.D. Briggs: The Si(001) $c(4\times4)$ surface reconstruction: a comprehensive experimental study, Surf. Sci. **430**, 154–164 (1999)
34.194 O. Leifeld, D. Grützmacher, B. Müller, K. Kern, E. Kaxiras, P.C. Kelires: Dimer pairing on the C-alloyed Si(001) surface, Phys. Rev. Lett. **82**, 972–975 (1999)
34.195 S. Ariyoshi, S. Takeuchi, O. Nakatsuka, A. Sakai, S. Zaima, Y. Yasuda: Influence of $Si_{1-x}Ge_x$ interlayer on the initial growth of SiGeC on Si(100), Appl. Surf. Sci. **224**, 117–121 (2004)
34.196 S. Zaima, A. Sakai, Y. Yasuda: Control in the initial growth stage of heteroepitaxial $Si_{1-x-y}Ge_xC_y$ on Si(001) substrates, Appl. Surf. Sci. **212–213**, 184–192 (2003)
34.197 S. Takeuchi, O. Nakatsuka, Y. Wakazono, A. Sakai, S. Zaima, Y. Yasuda: Initial growth behaviors of SiGeC in SiGe and C alternate deposition, Mater. Sci. Semicond. Process. **6**, 5–9 (2005)
34.198 G. Ehrlich, F. Hudda: Atomic view of surface self-diffusion: Tungsten on tungsten, J. Chem. Phys. **44**, 1039–1049 (1966)
34.199 R.L. Schwoebel, E.J. Shipsey: Step motion on crystal surfaces, J. Appl. Phys. **37**, 3682–3686 (1966)
34.200 J. Myslveček, C. Schelling, G. Springholz, F. Schäffler, B. Voigtländer, P. Šmilauer: On the origin of the kinetic growth instability of homoepitaxy on Si(001), Mater. Sci. Eng. B **89**, 410–414 (2002)
34.201 S. Fukatsu, H. Yoshida, N. Usami, A. Fujiwara, T. Takahashi, Y. Shiraki, R. Ito: Quantum size effect of excitonic band-edge luminescence in strained $Si_{1-x}Ge_x$/Si single quantum well structures grown by gas-source Si molecular beam epitaxy, Jpn. J. Appl. Phys. **31**, L1319–L1321 (1992)
34.202 D.K. Nayak, N. Usami, S. Fukatsu, Y. Shiraki: Band-edge photoluminescence of SiGe/strained-Si/SiGe type-II quantum wells on Si(100), Appl. Phys. Lett. **63**, 3509–3511 (1993)
34.203 N. Usami, Y. Shiraki, S. Fukatsu: Spectroscopic study of Si-based quantum wells with neighbouring confinement structure, Semicond. Sci. Technol. **12**, 1596–1602 (1997)
34.204 S. Fukatsu, Y. Shiraki: Interwell coupling in strained $Si_{1-x}Ge_x$/Si quantum wells, Ext. Abs. 1993 Int. Conf. Solid State Devices Mater. (Makuhari, 1993) p. 895
34.205 S. Fukatsu: Luminescence investigation on strained $Si_{1-x}Ge_x$/Si modulated quantum wells, Solid-State Electron. **37**, 817–823 (1994)
34.206 K. Kawaguchi, Y. Shiraki, N. Usami, J. Zhang, N.J. Woods, G. Breton, G. Parry: Fabrication of strain-balanced Si/$Si_{1-x}Ge_x$ multiple quantum wells on $Si_{1-y}Ge_y$ virtual substrates and their optical properties, Appl. Phys. Lett. **79**, 344–346 (2001)
34.207 K. Kawaguchi, S. Koh, Y. Shiraki, J. Zhang: Fabrication of strain-balanced $Si_{0.73}Ge_{0.27}$/Si distributed Bragg reflectors on Si substrates, Appl. Phys. Lett. **79**, 476–478 (2001)
34.208 K. Kawaguchi, S. Koh, Y. Shiraki, J. Zhang: Fabrication of strain-balanced $Si_{0.73}Ge_{0.27}$/Si-distributed Bragg reflectors on Si substrates for optical device applications, Physica E **13**, 1051–1054 (2002)

34.209 K. Kawaguchi, K. Konishi, S. Koh, Y. Shiraki, Y. Kaneko, J. Zhang: Optical properties of strain-balanced $Si_{0.73}Ge_{0.27}$ planar microcavities on Si substrates, Jpn. J. Appl. Phys. **41**, 2664–2667 (2002)
34.210 K. Kawaguchi, M. Morooka, K. Konishi, S. Koh, Y. Shiraki: Optical properties of strain-balanced SiGe planar microcavities with Ge dots on Si substrates, Appl. Phys. Lett. **81**, 817–819 (2002)
34.211 N. Usami, T. Mine, S. Fukatsu, Y. Shiraki: Realization of crescent-shaped SiGe quantum wire structures on a V-groove patterned Si substrate by gas-source Si molecular beam epitaxy, Appl. Phys. Lett. **63**, 2789–2791 (1993)
34.212 N. Usami, T. Mine, S. Fukatsu, Y. Shiraki: Optical anisotropy in wire-geometry SiGe layers grown by gas-source selective epitaxial growth technique, Appl. Phys. Lett. **64**, 1126–1128 (1994)
34.213 B. Teichert: Self-organization of nanostructures in semiconductor heteroepitaxy, Phys. Rep. **365**, 335–432 (2002)
34.214 Z. Zhang, M.G. Lagally: *Morphological Organization in Epitaxial Growth and Removal* (World Scientific, Singapore 1998)
34.215 K. Brunner: Si/Ge nanostructures, Rep. Prog. Phys. **65**, 27–72 (2002)
34.216 J.-M. Baribeau, N.L. Rowell, D.J. Lockwood: Self-assembled $Si_{1-x}Ge_x$ dots and islands. In: *Self-organized Nanoscale Materials*, ed. by M. Adachi, D.J. Lockwood (Springer, New York 2006)
34.217 M. Miura: Studies of formation process and optical properties of self-assembled Ge/Si nanostructures. Ph.D. Thesis (University of Tokyo, Tokyo 2001)
34.218 M. Miura, J.M. Hartmann, J. Zhang, B. Joyce, Y. Shiraki: Formation process and ordering of self-assembled Ge islands, Thin Solid Films **369**, 104–107 (2000)
34.219 H. Takamiya, M. Miura, J. Mitsui, S. Koh, T. Hattori, Y. Shiraki: Size reduction of the Ge islands by utilizing the strain fields from the lower-temperature-grown hut-clusters buried in the Si matrix, Mater. Sci. Eng. B **89**, 58–61 (2002)
34.220 E.S. Kim, N. Usami, Y. Shiraki: Control of Ge dots in dimension and position by selective epitaxial growth and their optical properties, Appl. Phys. Lett. **72**, 1617–1619 (1998)
34.221 E.S. Kim, N. Usami, Y. Shiraki: Selective epitaxial growth of dot structures on patterned Si substrates by gas source molecular beam epitaxy, Semicond. Sci. Technol. **14**, 257–265 (1999)
34.222 H. Takamiya, M. Miura, N. Usami, Y. Shiraki: Drastic modification of the growth mode of Ge quantum dots on Si by using boron adlayer, Thin Solid Films **369**, 84–87 (2000)
34.223 S. Koh, K. Konishi, Y. Shiraki: Small and high-density GeSiC dots stacked on buried Ge hut-clusters in Si, Physica E **21**, 440–444 (2004)

35. Plasma Energetics in Pulsed Laser and Pulsed Electron Deposition

Mikhail D. Strikovski, Jeonggoo Kim, Solomon H. Kolagani

Surface bombardment by energetic particles strongly affects thin-film growth and allows surface processing under non-thermal-equilibrium conditions. Deposition techniques enabling energy control can effectively manipulate the microstructure of the film and tune the resulting mechanical, electrical, and optical properties. At the high power densities used for depositing stoichiometric films in the case of pulsed ablation techniques such as pulsed laser deposition (PLD) and pulsed electron deposition (PED), the initial energetics of the material flux are typically on the order of 100 eV, much higher than the optimal values ($\leq$ 10 eV) required for high-quality film growth. To overcome this problem and to facilitate particle energy transformation from the original as-ablated value to the optimal value for film growth, one needs to carefully select the ablation conditions, conditions for material flux propagation through a process gas, and location of the growth surface (substrate) within this flux. In this chapter, we discuss the evolution of the material particles energetics during the flux generation and propagation in PLD and PED, and identify critical control parameters that enable optimum thin-film growth. As an example, growth optimization of epitaxial GaN films is provided.

PED is complementary to PLD and exhibits an important ability to ablate materials that are transparent to laser wavelengths typically used in PLD. Some examples include wide-bandgap materials such as SiO_2, Al_2O_3, and MgO. Both PLD and PED can be integrated within a single deposition module. PLD–PED systems enable in situ deposition of a wide range of materials required for exploring the next generation of complex structures that incorporate metals, complex dielectrics, ferroelectrics, semiconductors, and glasses.

35.1 Energetic Condensation in Thin Film Deposition 1193
35.2 PLD and PED Techniques 1194
35.3 Transformations of Atomic Energy in PLD and PED 1195
35.3.1 Plasma Formation of Vaporized Material 1196
35.3.2 Plasma Formation in PED 1198
35.3.3 Expansion of Plasma and Particle Acceleration 1199
35.3.4 Deceleration of Plasma in Background Gas 1202
35.4 Optimization of Plasma Flux for Film Growth 1204
35.4.1 Ion Current of Plasma Propagating in Ambient Gas 1205
35.4.2 Optimization of Growth of GaN Films – A Materials Example 1206
35.5 Conclusions 1208
References 1209

35.1 Energetic Condensation in Thin Film Deposition

Thin-film deposition under energetic particle bombardment is well known. Originally suggested in 1938 for improving film density, the energetic process was further developed as *ion plating* in 1963 [35.1]. A variety of techniques to produce energetic conditions are currently in practice [35.2]. These include plasma-assisted

film deposition techniques [35.3], magnetron sputtering [35.4], and ion-beam-assisted deposition [35.5, 6]. A common factor accompanying all energetic condensation processes is that the film growth surface is under constant bombardment by an energetic particle flux as it receives the desired material for film growth.

Under the bombardment by energetic particles, surfaces and subsurface layers of crystals are in an unusual (nonequilibrium) state controlled by the particle flux parameters rather than by deposition conditions such as substrate temperature. In this state, the equilibrium dynamics are no longer applicable to the dynamics of the crystal surface modification processes, which include implantation of atoms, radiation-enhanced diffusion, point-defect interaction, sputtering, ion mixing, adsorption, adatom movement, and film growth. The modified layers retain a *memory* of the conditions of their origin and exhibit unusual properties. This can be beneficial for some applications of the modified layers, and undesirable for others. Control of the bombardment conditions is critical for optimizing a specific process. In general, the particle energy of interest is in the range of 10–100 eV, which is well above the common processing temperature of < 0.1 eV (≈ 1200 K). In some sense, the particle bombardment substitutes for the higher temperature that is problematic to achieve under equilibrium conditions.

Pulsed laser deposition (PLD) and pulsed electron deposition (PED) are relatively new energetic condensation techniques with unique plasma flux parameters. In PLD and PED, a pulsed laser or a pulsed electron beam rapidly vaporizes a thin section of a target material, providing a stream of energetic plasma flux under highly nonequilibrium conditions. This process is known as ablation. The film growth conditions are intricately linked to and controlled by the conditions of the flux generation and its propagation in an ambient process gas (or in vacuum). This chapter will consider and compare these two techniques, emphasizing the critical role of plasma energetics in thin-film formation. The objective is to follow the energy balance through the entire process sequence from flux generation at the target surface to flux arrival at the growth surface (substrate), and to draw conclusions on the possibilities of controlling and optimizing the deposition conditions in PLD and PED.

We limit our consideration to the most commonly used laser pulses of nanosecond duration. Also we leave beyond the scope of this chapter the subject of macrodefects (droplets) in deposited films.

35.2 PLD and PED Techniques

The idea of utilizing the exceptionally high power density of a laser beam for material processing has been tested contemporarily with the development of lasers since the 1960s. The effects of heating a material to temperatures well above an *evaporation temperature* and transforming the vapor into a state of plasma have been studied extensively. In their pioneering work, *Smith* and *Turner* [35.7] explored plasma condensation to form a film layer on a substrate placed in a vacuum. Since then, a variety of PLD processes have been considered and summarized in several excellent reviews [35.6, 8–10]. In the conventional PLD process, shown schematically in Fig. 35.1, a high-power pulsed laser beam is used as an external energy source to rapidly vaporize the target material. When a pulse of submicrosecond duration is used (accomplished in Q-switched solid-state or excimer lasers, for example), the general process is not limited to evaporation but is accompanied by strong absorption of the laser beam in the vapor and its transformation into plasma. The interaction of the laser beam with the target and the generated plasma is a very complex physical process, and depends on the properties of both the target and the laser beam.

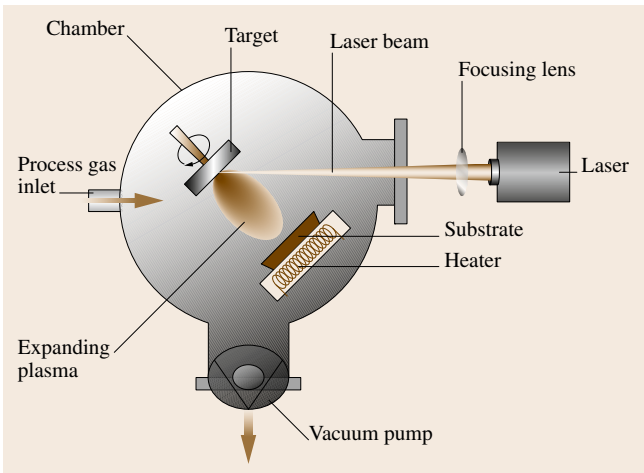


Fig. 35.1 Schematic representation of a conventional PLD process

In PED, the general scheme and ablation processes are very similar to in PLD, although a significant difference in the nature of the pulse energy source exists. In the PED process, the energy for ablation is delivered to the target by pulsed electrons rather than photons (as in PLD). The electron beam interacts differently with the target and the evaporant it generates. However, similarly dense plasma is formed at the target surface by both the pulsed electron and the pulsed laser beam. This dense plasma is the primary precursor source material for film deposition. The dense, high-temperature, strongly ionized plasma layer at the target surface propagating in a direction perpendicular to the target surface is known as the *plasma plume* in the published literature.

35.3 Transformations of Atomic Energy in PLD and PED

In energetic deposition processes, atoms from the target surface pass through several energy transformation phases before arriving at a substrate surface. Starting from their evaporation, the atoms are involved in a series of energy transfer processes controlled by different mechanisms. These atoms experience significant changes in their temperature, ionization state, as well as kinetic energy. To evaluate the possibilities of controlling the energetics in the film deposition process, it is critical to follow the dynamics of the energetics of the atoms. For every phase, there are key parameters enabling control of the transient processes. Ultimately, this enables control over the energy of the atoms arriving at the substrate, which is important for the film growth. PLD and PED processes, viewed as a sequence of several phases, are shown schematically in Fig. 35.2.

The pulse width of the laser or electron beam defines the duration of phase 1 in Fig. 35.2. During this period, most of the beam energy is transformed into the internal energy (enthalpy) of the atoms in the plasma. This includes energies of evaporation and ionization, the thermal energy of material atoms, as well as the energy of electrons in the plasma. Typically, the thickness of the plasma layer near the target surface is below 0.1 mm during this phase.

Under the action of high internal pressure, the plasma layer ejects itself in a direction perpendicular to the target surface (phase 2, Fig. 35.2). The plasma pressure is orders of magnitude larger than any practical gas pressure in the process chamber. Thus, the plasma ejection is not affected by the presence of process gases. The plasma experiences continuous acceleration by the pressure gradient supported through recombination of atoms and electrons. During this phase, the original energy spectrum of the plasma flux is formed. A typical plasma size in this phase is on the order of several times the spot size on the target. A three-dimensional expansion begins within this distance. In a typical PLD or PED process, this expansion is ≈ 10 mm in length. If deposition is carried out in a vacuum (10^{-6} Torr or below), the plasma expands and the original energy spectrum is *frozen*, and can be analyzed by an appropriate time-of-flight technique.

If deposition takes place in a background gas (phase 3, Fig. 35.2), the flux atoms collide with the gas atoms. As a result, the initial energetic spectrum of the atoms in the plasma evolves on a characteristic length scale related to the length scale of the mean free path at a given pressure. If the number of ablated atoms is sufficient to involve a comparable number of gas atoms in the mutual hydrodynamic movement, this *snowplow* effect drives most of the gas atoms towards the substrate, leaving behind a gas-rarefied region (shown pale brown in Fig. 35.2). Simultaneously, as the entire ensemble decelerates during this phase, the kinetic energy of both the target and gas atoms decreases.

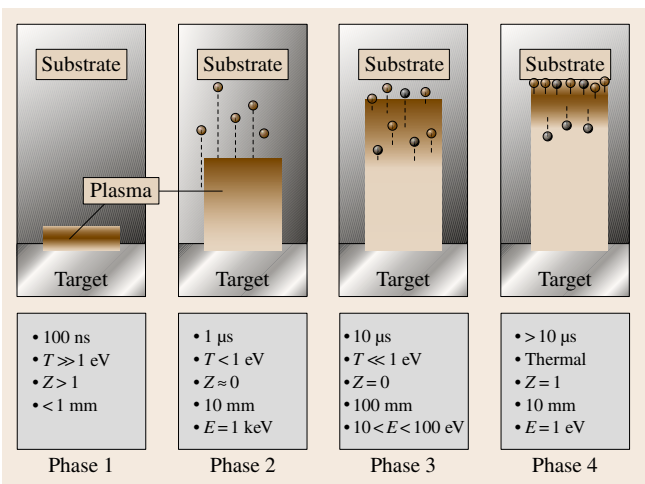


Fig. 35.2 Schematic representation of PLD and PED processes. Conventions: *brown circles*: fast atoms of target material, *black circles*: gas atoms, *pale brown area*: depleted gas volume behind the plasma front

In the next phase (phase 4, Fig. 35.2), the directional velocity of plasma flux becomes comparable to the thermal velocity of its particles (i.e., the flux transforms from supersonic to subsonic states). Plasma expansion nearly stops (seen as a range for its visible radiation *plume*), and becomes nearly isotropic. In most cases, an optimal substrate location for film growth can be found in the near vicinity of this range.

35.3.1 Plasma Formation of Vaporized Material

The dominant process in phase 1 is the evaporation and ionization of a small amount of material. First, at the typical beam power density used in ablation ($Q \approx 10^9\,\mathrm{W/cm^2}$), the target surface is rapidly heated, attaining temperatures well above the common *evaporation* temperature of all the elements in the target. The evaporation can be so intense that the vapor density at the surface can approach the limiting value of the atomic concentration in a solid ($n_\mathrm{L} \approx 5\times10^{22}\,\mathrm{cm^{-3}}$). Second, in the field of the intense laser beam, the vapor effectively ionizes (transforms into plasma) and absorbs the majority of the incoming energy from the laser, preventing further heating of the target. Several key parameters that control the process are, pulse duration τ, power density Q [$\mathrm{W/cm^2}$], and energy density Q_τ [$\mathrm{J/cm^2}$]. The laser wavelength λ in PLD and the energy of each electron [keV] in PED define the absorption depth of the beam with respect to the target material.

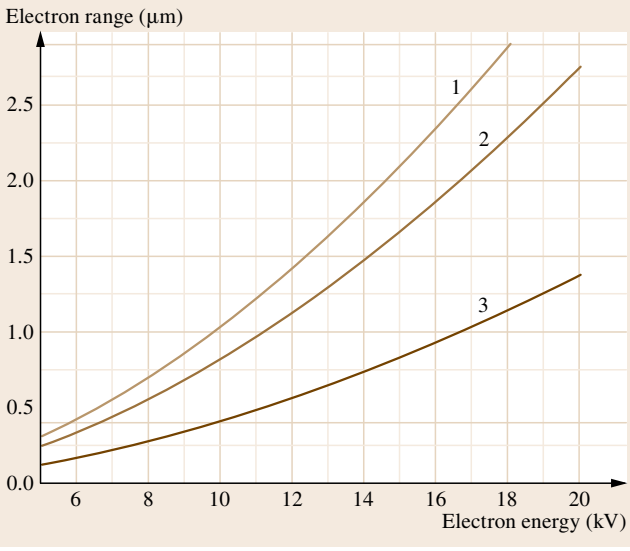


Fig. 35.3 Calculated penetration depth (range) of energetic electrons in Si (1), Al_2O_3 (2), and $YBa_2Cu_3O_x$ (3) (after [35.11])

The pulse energy is distributed in the target over a depth $(D + D_\mathrm{T})$, and the rate of surface temperature rise can be expressed as

$$\frac{\mathrm{d}T}{\mathrm{d}t} \sim \frac{Q}{D + D_\mathrm{T}}\,, \tag{35.1}$$

where D is the absorption length (photon or electron range) and D_T is the thermal diffusion length. D_T can be expressed as $D_\mathrm{T} = 2(at)^{1/2}$, where a is the thermal diffusivity and t is the time. D_T is also the characteristic depth of exponential decay of the temperature distribution, $T(x,t) \propto \exp(-x^2/4at)$ for one-dimensional (x) heat diffusion from a surface source [35.12].

For $\tau = 10\,\mathrm{ns}$, a pulse width typical in PLD, D_T is about 0.2 μm for most common dielectrics. This value is less than or comparable to the photon absorption depth D, and D_T can be neglected in (35.1). Thus the heating rate is controlled by the target optical properties at a given wavelength, and increases *proportionally* with the beam power.

This is not the case in PED. The penetration depth (or absorption range) of an electron in a solid depends on its energy. In PED, both the electron beam power Q and the electron range depend on the electron source voltage V. Since the number of electrons in the pulse is proportional to the total beam current density I, and since the average energy of each electron ($\sim$ eV) is directly related to the pulsed electron beam source (PEBS) voltage V, the power density carried by the beam is $Q \approx IV$, where $I = I(V)$. The $I(V)$ relation characterizes a PEBS design. For 10–100 kV electrons, the range D changes with voltage as $D \propto V^2$, as shown in Fig. 35.3. If $D > D_\mathrm{T}$, the $D \propto V^2$ dependence controls the denominator in (35.1), and thus strongly affects the heating rate in PED. In some sense, changing the voltage in the PEBS is equivalent to changing the laser wavelength (and absorption depth) in PLD. The interplay of $Q(V)$ and $D(V)$ controls the surface heating rate in PED. The expected heating rate for PED has been examined elsewhere [35.13]. Experiments show that the beam power is not able to increase as fast as the beam penetration length at $> 15\,\mathrm{kV}$, which leads to a maximum in the heating rate at $\approx 15\,\mathrm{kV}$ (Fig. 35.4). Therefore, it is not always beneficial to use the source at higher voltage (and accordingly with higher energy electrons). A more efficient way to improve the heating rate in PED is via a larger current *density* at the target surface by minimizing the beam cross-section. In order to increase the current density, the target has to

be as close to the tube output as is practically possible since the electron beam diverges with distance from the output channel tube.

In both PLD and PED, once vapor with a sufficient density is formed at the target surface, the beam ionizes it. PLD differs from PED in the target surface heating dynamics and the mechanism of ionization. In PLD, free electrons gain energy in the oscillating electric field of a laser and ionize atoms by impact. This avalanche-like process is similar to the *optical breakdown* of gases by intense laser beams. The breakdown physically means that, once transformed into a strongly ionized plasma state, the plasma becomes nontransparent to the laser beam. The plasma shields the target material from the beam, and absorbs the majority of the pulse energy. This plasma shielding strongly affects the energy balance in PLD and PED, the primary topic of interest in this chapter. Well above the ablation threshold, the fraction of the beam energy used for target heating is relatively small. The majority of the beam energy is preserved within the plasma [35.14]. The net energy balance of the entire ablation process can be roughly understood as

$$Qt = E_1 N \,, \tag{35.2}$$

where E_1 is the *total* average energy per evaporated particle (ion or atom) and N is the total number of atoms or ions per unit area. The energy E_1 eventually transforms into kinetic energy $E \le E_1$ of atoms ejected from the target. The balance indicates that higher particle energy E_1 is expected in an ablation process in which a smaller number (N) of ablated particles enables a strong shielding effect.

The total number N of atoms absorbing the beam energy can be estimated as $N \approx nL$, where n is the atomic concentration in the shielding plasma and L is the thickness of the plasma layer. Some understanding of $n(\lambda)$ scaling laws for different laser wavelength λ (frequency $\omega = c/\lambda$) can be gained by considering the conditions for optical breakdown of gases. Under a crude approximation, the gas concentration most favorable for developing the electron avalanche corresponds to the condition where the frequency Ω of electron collisions with atoms is on the same order as the beam frequency ω. The critical concentration $n = n_c$ can be found from the condition $\Omega(n_c) \approx \omega$. The physical reason for the optimum is that, for a given ω of the laser, the rate of electron energy gain in the electric field of the beam has a maximum at this value of collision frequency [35.15]. The gas ionization develops via inverse bremsstrahlung absorption. The collision frequency relates directly to the gas (vapor) concentration as $\Omega = \alpha n$. As an example, for normal conditions of 760 Torr ($n = 2.47 \times 10^{19}\,\mathrm{cm}^{-3}$) of Ar gas ($\alpha = 1.6 \times 10^{-7}\,\mathrm{cm}^3\mathrm{s}^{-1}$), one finds $\Omega = 4 \times 10^{12}\,\mathrm{s}^{-1}$, and $\lambda = 75\,\mu\mathrm{m}$ radiation would be optimal to create plasma.

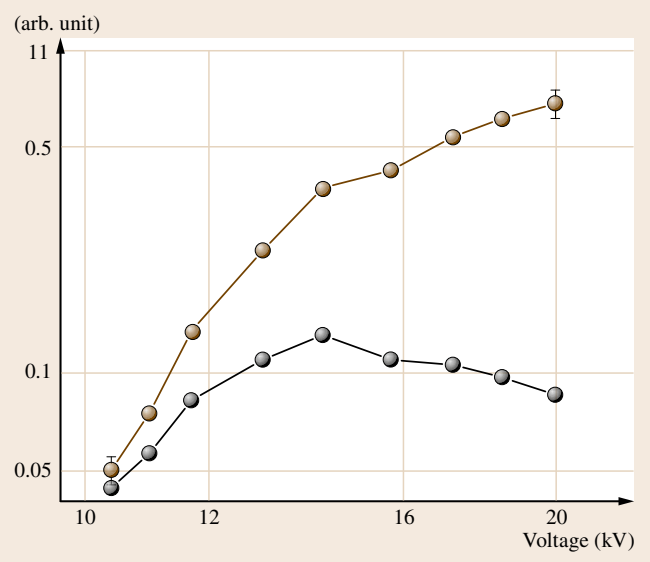


Fig. 35.4 Estimated surface heating rate (*black circles*) and measured beam power (*brown circles*) as functions of the electron energy

As a wide range of atom concentration in the vapor is available in PLD at the prebreakdown stage, the above condition can be satisfied at least for λ down to $\approx 1\,\mu\mathrm{m}$.

The condition gives an estimate for the initial atomic concentration to develop the shielding plasma. It is inversely proportional to the laser wavelength, $n_c \propto \lambda^{-1}$. The relations above describe proper conditions to initiate plasma at microwave frequencies, where the pressure of ≈ 2 Torr is optimal for 2.8 GHz ($\lambda = 1.0$ mm). It is also suitable for CO_2 laser with $\lambda = 10.6\,\mu\mathrm{m}$. The calculated value of 15 atm ($n_c = 3.8 \times 10^{20}\,\mathrm{cm}^{-3}$) as the optimal pressure for Xe gas plasma formation is close to the experimental value of 20 atm [35.15]. The breakdown threshold power density Q^* of the CO_2 pulsed laser is as low as $\approx 5 \times 10^7\,\mathrm{W/cm}^2$. For radiation of an Nd:yttrium aluminum garnet (YAG) laser ($\lambda = 1.06\,\mu\mathrm{m}$), it can be estimated that the optimal vapor concentration is $n_c = 3.8 \times 10^{21}\,\mathrm{cm}^{-3}$. For a shorter-wavelength laser, the plasma starts to develop in a higher atomic concentration in the vapor. For the PLD process, this suggests that more material has to be evaporated before shielding occurs.

For much shorter wavelengths such as those of ultraviolet (UV) excimer lasers, the vapor breakdown condition and mechanism are more complex. Theoretical estimation from the above relations implies that the estimated optimal atomic concentration in a vapor of $n_c = 1.6\times10^{22}\,\mathrm{cm}^{-3}$ becomes unrealistically high at $\lambda = 248\,\mathrm{nm}$ (KrF laser), for example. The vapor breakdown threshold intensity of the beam, according to the classical theory, should also increase sharply as $Q^* \propto \lambda^{-2}$ [35.16]. The scaling from 10.6 μm leads to $Q^* \approx 9\times10^{10}\,\mathrm{W/cm^2}$. These estimations, however, have not been supported by experimental results. Experimental values as low as $\approx 1\times10^{8}\,\mathrm{W/cm^2}$ for Al and Cu targets has been found for 248 nm [35.17]. The suggested explanation of the reduced breakdown threshold for UV lasers is that a multiphoton photoelectric effect takes over from the atomic ionization process as the high energy of UV quantum becomes comparable to the atomic excitation and ionization potentials [35.15, 18]. The threshold can actually *decrease* with wavelength at $\lambda < 1\,\mu\mathrm{m}$. The low thresholds indicate that a very dense vapor is not needed to form UV laser plasma. Thus, the atomic concentration on the order of $n \approx 1\times10^{21}\,\mathrm{cm}^{-3}$ in the vapor may be a reasonable estimate for typical breakdown conditions for wavelengths in the range of $\lambda = 1.0$–$0.25\,\mu\mathrm{m}$ in PLD. However, the concentration does change with λ. Data for the laser wavelength dependence of the ablated material mass [35.19] indicate that the plasma can still be created in vapor with *higher* density for shorter-wavelength lasers. Following the considerations mentioned above, in this case the energy balance for the PLD process is shifted towards a smaller number of ablated atoms and a larger energy of ejected atoms/ions. This trend is indeed seen in the energy spectrum of ejected ions [35.20, 21].

Experiments suggest that, at $3\times10^{8}\,\mathrm{W/cm^2}$ ($\lambda = 248\,\mathrm{nm}$), breakdown starts at the beginning of a 30 ns pulse and takes just ∼ 5 ns to develop a plasma absorbing ≈ 85% of the beam intensity [35.17]. The majority of the beam energy is loaded into the vapor. The plasma thickness L can be estimated as $v_T t \approx 10^{-3}\,\mathrm{cm}$ where $v_T \approx 10^{5}\,\mathrm{cm/s}$ is the thermal velocity of vapor expanding from the target surface. The plasma contains $nL \approx 1\times10^{18}\,/\mathrm{cm^2}$ atoms, which would originally occupy a $\approx 2\times10^{-5}\,\mathrm{cm}$ deep layer in a solid target. Note that, for typical cohesive energy of ≈ 4 eV/atom in solids, the evaporation of this number of atoms requires $\approx 0.8\,\mathrm{J/cm^2}$. This is typically a small fraction of the beam energy in intensive ablation conditions ($> 3\,\mathrm{J/cm^2}$). Most of the pulse energy is loaded into the total energy (enthalpy) of plasma, including the thermal energy of ions and electrons as well as the ionization energy.

With increasing laser intensity Q, both N and E increase, as confirmed by numerous data on the Q dependence of evaporated materials, deposited film thickness, and ion/neutral atom kinetic energy. Although most of these data were obtained in conditions of one-dimensional plasma expansion (plasma layer thickness smaller than the focal spot size of the laser beam), the dependencies generally follow the scaling of the model [35.22] developed for the regime with partial target shielding by the flowing plasma. Such a self-regulating regime (*optical plasmatron*) is established once the laser pulse duration is much longer than the time of the plasma expansion to the characteristic scale of the ablation spot diameter, as has been observed in ≥ 1 μs long-pulse (transversely excited atmospheric pressure (TEA) CO_2 laser, [35.23, 24]) or small-spot nanosecond-pulse ablation [35.12, p. 243].

Based on mass and energy conservation laws and based on the assumption of a three-body plasma recombination process, the *steady-state* ablation model suggests that the ablation rate can be expressed as

$$\frac{\mathrm{d}N}{\mathrm{d}t} \propto \lambda^{-4/9} Q^{5/9} \,. \tag{35.3}$$

Accordingly, to satisfy the $E_1\,\mathrm{d}N/\mathrm{d}t = Q = \text{const.}$ condition, the energy per atom can be expressed by

$$E_1 \propto \lambda^{4/9} Q^{4/9} \,. \tag{35.4}$$

These relations again point to the tendency of a shorter-wavelength laser beam to evaporate more material from a target, providing each atom with a lower energy. As a crude approximation, one can conclude that the average kinetic energy E of particles produced by laser ablation scales as $E \propto \lambda$ for 1.06–10.6 μm wavelength lasers and as $E \propto \lambda^{1/2}$ for 0.2–1.0 μm wavelength lasers. Choice of the laser is one of the possibilities to control the energy spectrum of the plasma flux in PLD.

35.3.2 Plasma Formation in PED

In PED, the electron beam energy of 5–15 keV is well above the threshold potential for atomic excitations and ionizations. Fast electrons collide directly with the deep-level electrons of target material atoms, experience broad-angle scattering, and eventually stop at a certain depth from the target surface (electron range). This stopping at a certain depth requires a certain number of collisions to happen. This number is practically independent of the specific state of the atoms, with atoms in the solid lattice scattering

electrons as effectively as atoms in a gas phase (vapor). An important consequence of this fact is that, for the electron beam, there is no transition to the vapor plasma shielding effect that is characteristic for the laser ablation. Throughout the entire pulse length, the electron beam continues to interact with the same number of atoms as contained within the beam absorption range D in the target. As $D \propto 1/\rho$ (ρ [g/cm^3] is the target material's density), the total mass ablated by the intense beam is expected to be independent of materials ablated, $m = \rho D = \text{const.}$ [35.11]. As mentioned above, the electron range is strongly dependent on the electron energy ε [kV]. The absorption range can be expressed as

$$D\,[\mu\text{m}] \approx 5.37\varepsilon \left(1 - \frac{0.9815}{1 + 0.0031\varepsilon}\right) \rho^{-1} .$$

Figure 35.3 shows the range of 0.5–2.0 μm for a typical PED electron beam energy of 10–15 keV. Even for dense materials (such as YBCO with a density of ≈ 6 g/cm^3), the range of ≈ 1.0 μm is greater than the typical depth of ≈ 0.2 μm for laser ablation (using a 248 nm KrF laser at 5 J/cm^2) [35.12]. If all the material within the penetration depth is PED-ablated, it follows, from (35.2), that the average kinetic energy of particles in the plasma flux can be smaller in PED than in PLD. For lower-density materials, a higher PED ablation rate and accordingly smaller energy per ablated atom could be expected. This is actually the case. However, the particle energy spectrum in PED is broader than in PLD due to the presence of a small amount of highly energetic ions, up to several keV. Some specific mechanism of ion acceleration could be responsible for the fast ion generation in PED.

35.3.3 Expansion of Plasma and Particle Acceleration

A dense layer of highly ionized high-pressure plasma is formed at the target surface by the focused laser pulse. The plasma expands and accelerates primarily in a direction normal to the target surface (Fig. 35.2, phase 2). During this process, the energy E_1 (the total energy per atom, i. e., the enthalpy per atom) is transformed into an asymptotic kinetic energy E of atoms [35.25], which is the energy of interest in the present discussion, as this flux of energetic particles arrives at the substrate for film formation (when PLD is carried out in vacuum). Gradient of plasma pressure is the driving force for the acceleration of particles. Note that, even when PLD is carried out in a typical pressure background gas, due to the large difference between the plasma pressure at the ablation spot and the gas pressure, the initial energy spectrum of E is established at a distance of a few millimeters (≈ 3 times the spot diameter) from the target and is not affected by the presence of the process gas.

For a vacuum process (or at a low gas pressure), the ablation particle energy spectrum established soon after leaving the spot region remains mostly unchanged during the following geometrical expansion. The ion condition is *frozen* due to the low thermal energy of the plasma and the lack of efficient collisions. Laser-produced plasma retains a significant portion of the atoms in the ionized state. Film growth in vacuum is accompanied by the impact of particles of the largest possible energy. This feature also makes the laser flux an efficient source of ions for analytical laser plasma spectroscopy or for further acceleration of ions in special applications.

Specific conditions of the laser-produced plasma facilitate kinetic energy of particles well above the plasma temperature. At the end of the laser pulse, the total internal energy per plasma atom, $I + (1 + Z)T$, is stored in the thermal energy of the ion, Z electrons, and the total potential energy I of the atomic ionization. During the plasma expansion, electrons contribute to the acceleration pressure. The plasma temperature is also supported by release of the potential energy by three-body recombination. Thus the resulting kinetic energy of the flux particles can greatly exceed the original thermal energy of the plasma atom $\sim T$ [eV].

The model for the steady plasma flow regime (pulse length > 1000 ns) [35.22] predicts for the asymptotic particle energy

$$E = 5(1 + Z)T . \qquad (35.5)$$

The functional relation $Z(I, T)$ can be used for atoms of a specific target material. As the beam energy stored per a plasma atom is ≈ 3 times greater than the target atom cohesion energy ≈ 4 eV (see before), a plasma temperature of > 10 eV can be estimated for a typical ablation with 3 J/cm^2. The potential energy of the first ionization I_1 is usually within the range 7–15 eV, and the $Z = 1$ ionization is already significant at temperature $T \approx I_1/5 = 1.4$–3 eV [35.16]. At $T \approx 2$–3 eV, the generation of $Z = 2$ ions is already possible. $Z = 2$ ions have been routinely observed under PLD ablation conditions. Equation (35.5) estimates $E = 30$–45 eV for $Z = 2$ ions according to the model [35.22] for a steady flow ablation regime.

The steady regime is not usually realized in a practical excimer-laser-based PLD configuration with spot

sizes of ≈ 1 mm, where the ≈ 30 ns laser pulse is too short in comparison with the time to establish the steady regime. However, this regime has been realized in small-spot nanosecond-pulse ablation [35.12, p. 243], and ablation with a long TEA CO_2 laser ($\lambda = 10.6\,\mu m$) pulse of ≈ 2 μs duration [35.23,24]. Its specific features are an increased average ablation rate, and narrower energy spectrum $dN/dE = f(E)$ relative to the unsteady pulsed regime.

The narrow ion energy spectrum $dN/dE = f(E)$ observed [35.23, 24] can be explained by the fact that the majority of the registered ions have been generated under nearly similar, quasistationary evaporation–absorption–acceleration conditions during most of the pulse duration. The average particles energy is laser wavelength dependent. For a $\lambda = 10.6\,\mu m$ CO_2 laser with $Q = 10^9\,W/cm^2$, a narrow spectrum with a maximum at ≈ 400 eV was measured. This energy is about a factor of ≈ 2.5 larger than that observed for the $\lambda = 1.06\,\mu m$ laser, and ≈ 5 times larger than that for 0.248 μm radiation, in accordance with the prediction of (35.4). Another feature of the steady regime is that the average energy of neutrals is nearly as high as that of ions [35.23, 24]. This reflects the fact that both the neutrals and the ions experience the same acceleration conditions.

Thus, at the same beam energy density of a 11 ns duration excimer laser pulse, the ablation rate (μm/pulse) increases by several times if the spot size is reduced from 83 to 24 μm [35.12]. This can be explained by a transition to a steady regime with a plasma more transparent to the laser beam than in the case of one-dimensional expansion (large spot). In accordance with the energy balance, one can expect a lower average kinetic energy of ablated particles to result.

If the laser beam power density is high enough, such an important feature of PLD as congruent evaporation of complex materials is preserved in the quasistationary ablation regime. Indeed, the basic requirement for stoichiometric material removal is that the volume of ablated material is much greater than the volume of remaining heated material. That can be expressed as $v\tau \gg a/v$, where the ablation rate v is the velocity of the ablation surface propagation into a target, and a/v is the thickness of the heated material in front of it. As mentioned above, for a given power density, v is even higher in stationary than in pulsed ablation, and the condition is satisfied. Experimentally, for example, stoichiometric ablation in the quasistationary regime has been demonstrated by fabricating high-quality YBCO films using 2 μs long pulses from a TEA CO_2 laser [35.26].

In an *unsteady* pulsed regime, the energy spectrum broadens in both directions compared with the energy spectrum for the steady flow regime. The leading ions on the front of the plasma layer are in favorable conditions to experience more effective acceleration, and attain greater energy. On the other hand, at the end of the laser pulse, the plasma pressure build up is terminated, and the ions at the surface created in this time are not able to accelerate. Thus, the leading ions with higher energy are present on the front of the expanding plasma stream. The effect is well known in classical gas dynamics in unsteady (abruptly started) gas flow in vacuum [35.16]. The energy of the leading particles E_{max} is $\sim 2/(\gamma - 1)$ times larger than the energy from the steady flow. The additional kinetic energy is acquired at the expense of heat energy in the neighboring particles. For plasma at the ablation conditions, the exponent γ can be approximated to 1.3–1.4. The maximal energy of the accelerated particles is $E_{max} = 10T(1 + Z)/(\gamma - 1) = 100$–$200\,eV$ at $Z = 2$ and $T = 2$–$3\,eV$.

The velocity or energy of neutral components of the flux is generally somewhat lower, but still close to that of the ions. A possible explanation is that a significant number of the neutrals is in effect the former ions fraction, which has recombined. This fact makes it possible to apply the ion time-of-flight technique to probe the energetics of the laser-producing plasma fluxes. Data of the time-of-flight mass spectrometry and time-resolved emission spectroscopy applicable to trace the neutrals dynamics [35.27, 28] show a neutral's velocity of ≈ 0.6–0.75 for ions (energy difference about a factor of 2–3). Still, the ion probe measurements are representative of the general flux energetics and represent a good estimate of the upper energy limit of the particles.

Films grow in vacuum under the impact of the particles of largest possible energy. The ion probe (Faraday cap) is a convenient and frequently used technique to characterize the dynamics of the plasma arriving at a point a distance L from the target. In this technique, a slightly negatively biased probe registers the current of the ions in the flux. For ablation in vacuum (or at low gas pressure), due to a large mean free path, the ions travel the distance ballistically, and the delay time t for ions arriving at the probe directly relates to their speed $v = L/t$ and kinetic energy $mV^2/2$. Thus, the ionic energy spectrum can be obtained as $dN/dE = f(E)$. The spectrum for $f(E)$ can be calculated from the ion probe current density $j = eZ(dN/dt)$ using the relation $dN/dE = (dN/dt)(dE/dt)^{-1}$, where $E = M(L/t)^2/2$

is the kinetic energy of an ion with mass M arriving at time t.

For ablation in an ambient background gas pressure, collisions with the gas alter the velocity of flux species and the probe signal. Since the relation $v = L/t$ is not valid any more due to the velocity change with distance, the probe signal cannot be used to extract the arriving ion's energy spectrum. The relation overestimates the ion's kinetic energy at distance L, as the original velocity is greater than L/t and as the velocity at arrival is smaller than L/t. A proper particle deceleration model can be used to describe the special evolution of their average energy.

A typical energy spectrum of Si ions is given in Fig. 35.5 for the plasma flux created by KrF laser pulse (wavelength 248 nm, 20 ns pulse, 2.5 J/cm^2 energy density) in vacuum. The insert shows the original ion current signal obtained at a distance of 11 cm. The median energy is ≈ 60 eV, large enough not only to significantly change the conditions of the surface movement of the arriving atoms but also to introduce property changes of the previously deposited layers of the growing film [35.29]. Characteristically, the spectrum exhibits an extended lip towards energies up to ≈ 200 eV. As discussed above, the fastest ions are those that are on the front of the expanding plasma cloud, where the largest pressure gradient exists for an extended time. Two additional factors can also contribute to the increased ion signal amplitude at the higher end of the spectrum and extending from the median energy value. First, multiple-ionized ions create a larger probe signal by a factor of Z (ion charge) than single-charged ones. Second, the energy of leading ions can be additionally increased by the effect of electrostatic acceleration of ions in a double layer of space charge existing at the front of the plasma [35.30].

Figure 35.6 compares the energy spectra of CeO_2 plasma flux ions obtained by PLD and PED at a distance of 14.5 cm at a similar energy density. The relatively small oxygen gas pressure of 1 mTorr does not alter the spectra significantly, at this distance, in comparison with the vacuum conditions. The spectra are normalized to probe signal maximum. A somewhat surprising result of this comparison is that the energy of the fastest ions in the PED-generated plasma flux is larger than in PLD. Possibly, some additional mechanism of ion acceleration exists in PED, the origin of which is not fully understood at the present time. One possible explanation is that the negative charging of the plasma cloud leads to a stronger expulsion of the electrons and to a stronger effect of electrostatic acceleration

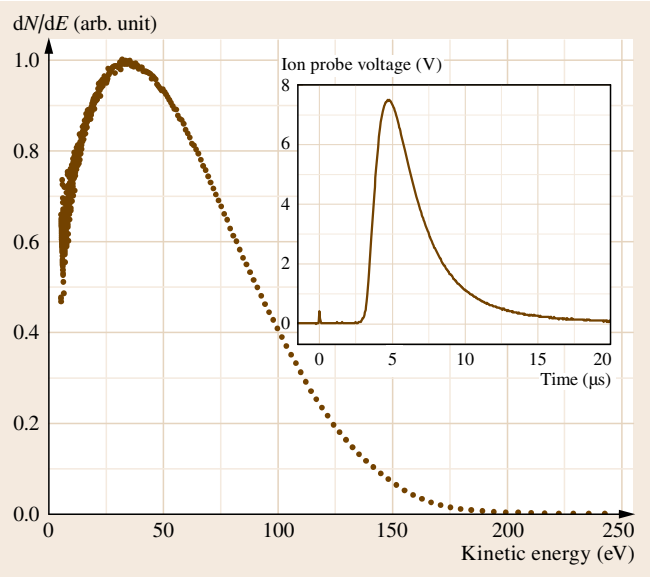


Fig. 35.5 Typical energy spectrum of Si ions generated by a KrF (248 nm) pulsed laser in vacuum. The *inset* shows the original ion probe signal at a distance of 11 cm from the Si target

of ions on the plasma front. At the same time, the PED spectrum is broader, and relatively more atoms with low energy are in ionized state. The average energy of the PED flux particles can be smaller than the

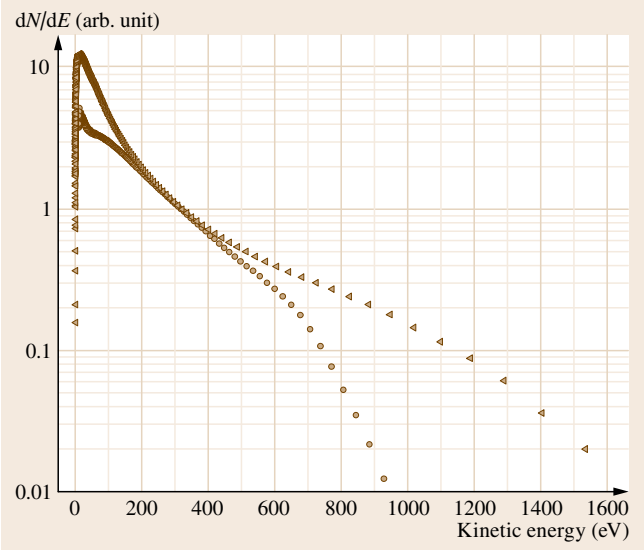


Fig. 35.6 Energy spectra of ions produced by pulsed laser (*closed circles*) and pulsed electron beam (*open triangles*) in oxygen pressure of 1 mTorr

energy in PLD. More studies are needed to fully understand the dynamics of PED-generated plasmas and their properties.

35.3.4 Deceleration of Plasma in Background Gas

In PED and PLD, the presence of background gases during ablation can significantly change both the average energy of the particles arriving at a substrate and the general attributes of the energy spectrum. When the mass of the evaporated atoms and of the incorporated gas atoms become comparable, hydrodynamic models can be used to describe deceleration of this ensemble and its thermalization. These are the phases 3 and 4 shown in Fig. 35.2. Introduction of a reactive or inert buffer gas into the deposition chamber is an effective tool to control the energy of species during film growth. Unlike vacuum PLD, plasma flux parameters change drastically with distance L from the target, which is critical for film growth. At a characteristic distance L_0, the flux loses its unidirectional velocity, as it undergoes scattering, thermalization, and finally deceleration. It has been experimentally discovered that optimum conditions for film growth exist in the vicinity of this distance [35.31–33].

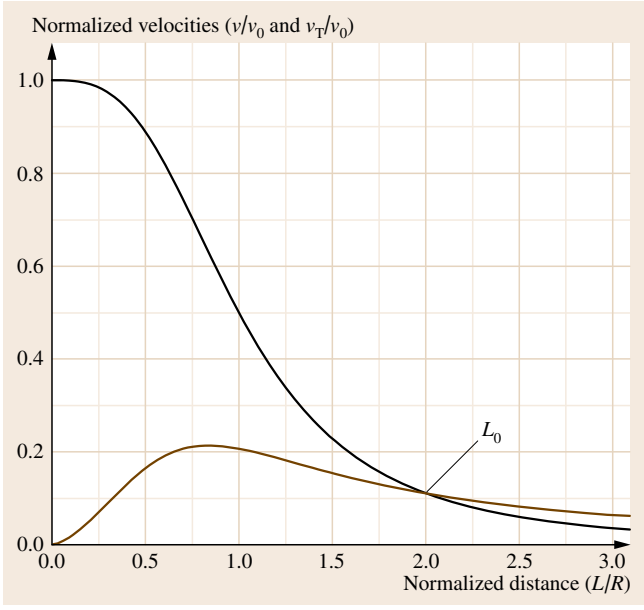


Fig. 35.7 Evolution of the directional v and thermal v_T velocities of plasma flux particles as a function of normalized distance L/R from the target. Velocities are normalized to the initial velocity v_0 of particles. L_0 indicates the location of the plasma range, where $v(L_0) = v_T(L_0)$

In the presence of background gases, the ion probe signal is not useful to calculate the arriving particles' energy as their velocity changes with distance. If calculated as the ratio L/t, the velocity (and the energy) would be overestimated. A simple model has been suggested to describe the dynamics of plasma deceleration, the average particle energy at a distance L_0, and the scaling of L_0 as a function of ablation and gas parameters [35.34]. In this model, the ensemble of initially ablated N_0 atoms, with an average velocity v_0, is considered as a piston, incorporating and driving the gas atoms as it propagates. The essential processes here are energy transfer from the initial kinetic energy of the N_0 atoms to the thermal energy of the atom and the gas cloud (i. e., deceleration and thermalization). From the conservation of momentum and energy in an adiabatic process, the following dependencies are obtained for the velocity v of the directed movement of the ensemble and the characteristic thermal velocity $v_T = (k_B T/M)^{1/2}$ of the particles.

$$\frac{v}{v_0} = (1+x^3)^{-1} ,$$

$$\frac{v_T}{v_0} = x^{3/2}[3(1+x^3)(1+\mu x^3)]^{-1/2} ,$$

where $\mu = M/m$ is the atomic mass ratio between the target material and the gas, and $x = L/R$ is the normalized distance. R has a clear physical meaning: after the ensemble covers the distance R, the total mass of the gas atoms becomes equal to the mass of ejected N_0 atoms.

Comparison of the velocities $v(x)$ and $v_T(x)$ (Fig. 35.7) shows that there is a characteristic distance L_0 where the thermal expansion rate of the ensemble exceeds its forward movement rate. It is reasonable to view the distance L_0 as the *range* of the flux defined according to the condition

$$v(L_0) = v_T(L_0) .$$

By the definition of v_T, half of the ensemble atoms are not moving towards the substrate at a distance L_0. The optimal position of the substrate for growth of high-quality films is usually found experimentally to be at around the end of the luminous flux area (*plume*). It is reasonable to associate this position with the characteristic range L_0 of the plasma flux. Let us consider three issues regarding this location important for film growth: the average particle energy $E(L_0) = Mv^2(L_0)/2$ at the distance L_0, the dependence of the distance L_0 on the gas pressure and intensity of ablation N_0, and the deposition rate $h(L_0)$ [nm/pulse] at this location.

Note that the flux range L_0 does not depend on the initial velocity v_0. However, an increase in the initial energy $E_0 = (Mv_0^2)/2$ of ejected atoms makes the process faster and increases the kinetic and thermal energies of particles at distance L_0. As can be seen from Fig. 35.7, the ensemble velocity in the vicinity of the flux range drops to ≈ 10–20% of the initial velocity, i.e., $E(L_0) \approx (0.01$–$0.04)E_0$ at this point. Thus, for a flux with an average initial energy of particles of ≈ 200 eV, the expected energy of particles arriving at the substrate surface at the distance L_0 is 2–8 eV. Particles with this energy are acceptable for film growth as they are able to nonthermally activate the film surface without damaging it.

The distance L_0 itself is controlled by the ratio of the ablated material amount to the background gas concentration n_g (gas pressure $P \propto n_g$)

$$L_0 \propto \left(\frac{N_0}{n_g}\right)^{1/3} .$$

To maintain the substrate in the optimal position, the ejected material amount N_0 should be changed proportionally with the changing pressure. The parameter N_0 characterizes a specific PLD system and can be easily measured. To do so, it is sufficient to measure the thickness (per pulse) h of the film deposited with the system (at some target-to-substrate distance d) in a vacuum or at a pressure low enough to assure that all the ejected atoms reach the substrate. By definition, N_0 is directly related to h as $N_0 = hn_L d^2$, where n_L is the concentration of deposited atoms ($n_L \approx 5 \times 10^{22}$ cm^{-3} for most solids). Numerically, $N_0 = 5 \times 10^{16}$ for a system delivering $h = 0.1$ nm/pulse at $d = 10$ cm. For typical PLD and PED of dielectric materials, the deposition rates at 5 cm distance are within the range 0.01–0.1 nm/pulse. The dependence $L_0(n_g)$ of the range on gas pressures is shown in Fig. 35.8. The process pressure for PED (in the range 1–10 mTorr) is about two orders of magnitude lower than that for PLD in the case of several dielectrics. This leads to larger target-to-substrate distances that are expected to be optimal in PED relative to PLD.

The deposition rate at the optimum distance $h(L_0)$ turns out to be only weakly dependent on the ablation intensity N_0

$$h(L_0) \propto n_g^{2/3} N_0^{1/3} . \quad (35.6)$$

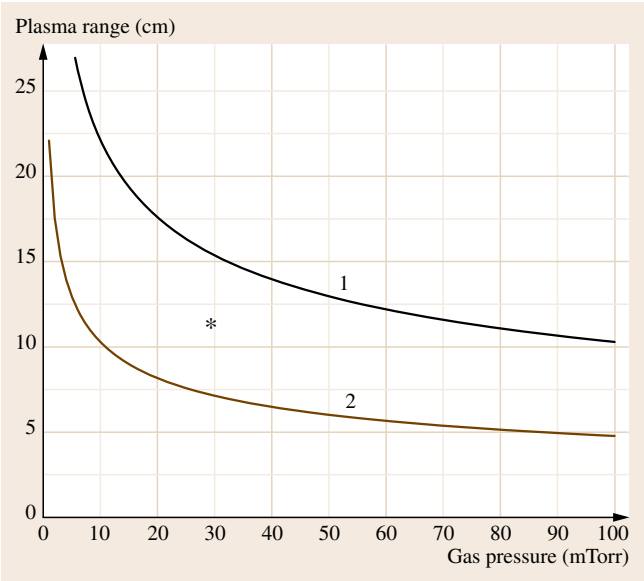


Fig. 35.8 Plasma propagation range L_0 as a function of gas pressure for ablation intensities N_0 of 0.1 nm/pulse (1) and 0.01 nm/pulse (2) at 10 cm distance. The *asterisk* represents an experimental optimal condition for PED growth of YBCO films

This result is remarkable, since it explains why different PLD systems (with different N_0 values) exhibit quite similar optimal deposition rates. Indeed, the reported deposition rates of YBCO films were close to 0.1 nm/pulse in a number of publications [35.35–39]. Thus, that deposition rate could be considered typical for PLD in a background gas pressure of 100–200 mTorr.

On the other hand, as follows from (35.6), the deposition rate at an optimal distance scales rather strongly with the gas pressure. The optimal deposition rates in PED (at 5 mTorr) are expected to be about ten times lower than in PLD (at 150 mTorr). Note here that all these differences follow from the difference in operational pressures. The dynamics of the deceleration of plasmas produced with pulsed lasers and pulsed electron beams is the same. Lower-pressure operation generally leads to a longer flux range, larger optimal film deposition distance, and larger deposition area. The overall *productivity* of the film fabrication (grams of material per second) is controlled by the average power of the beam (pulsed laser or pulsed electron).

35.4 Optimization of Plasma Flux for Film Growth

As the models predict, and experimental data confirm, the particle energy generated by lasers (not considering exotic high-power, subnanosecond pulse lasers) and intensive electron beams can cover a wide energy spectrum range of 3–3000 eV. Depending on the particle energy, their interaction with crystal surface can have very different results. Due to the significant dispersion of the spectrum, it is difficult to predict the cumulative integrated effect of the energetic particle flux on a crystal (substrate) surface. However, to fit the spectrum and the desired result for a particular application, an optimal type of laser (λ, τ), and beam intensity Q can be chosen accordingly to the discussions presented above.

High-energy (> 1000 eV) particles cause shallow implantation on a crystal surface, resulting in generally improved coating adhesion through ion mixing. Nonequilibrium generation of lattice vacancies promotes radiation-enhanced diffusion into the subsurface layer. Crystal surface compensation or doping by the laser-ablated materials can be realized [35.40].

When the incoming flux has a large fraction of energetic ions ≥ 200 eV, this can strongly affect the average deposition rate of metallic films where self-sputtering becomes significant. The effect is especially strong for large-atomic number metals such as Cu, Ag or Au [35.41]. For these materials, the self-sputtering coefficient of unity is reached at the ion energy as low as 100 eV [35.42, 43]. When the high-energy fraction in the deposition flux is large enough, no film can be formed. In some instances, much thinner film (or no film at all) is obtained on the axis of the deposition flux (where most of the higher-energy particles are concentrated) than in off-axis directions. Attempt to reduce the self-sputtering effect by adding a background gas leads to a maximum in the deposition rate (at ≈ 100 mTorr) dependence on pressure [35.44]. The effect can be understood as due to the interplay between two kinds of rate reductions. One is due to scattering or stopping of the flux material at higher pressures while the other is due to partial self-sputtering of films by the energetic portion of plasma flux ions at lower pressures. It is interesting that the maximum deposition rate was obtained in He background gas, which was able to decelerate fast ions without causing their effective large-angle scattering. The film stress also showed strong gas pressure dependence.

The concurrent bombardment of growing film promotes the growth of dense films. The increase in film density is a major factor in modifying film properties such as hardness, electrical resistivity, optical properties, and corrosion resistance. At the same time, it can introduce high compressive stresses. Especially strong is the effect of stress level on the electrical properties of conducting oxide films where their charge carrier concentration and mobility critically depend on the film lattice parameters.

Enhancement of film growth and film properties by the energetic particle interaction with film surfaces is a well-known phenomenon. Low-energy (≈ 5 eV) bombardment promotes the surface mobility of adatoms that is used for epitaxial growth [35.45]. Properties of films grown from wide-energy-spectrum plasma fluxes are the integrated result of all the effects above. Especially important process parameters are the average energy of arriving atoms and the ratio of high- and low-energy particles. According to the well-known Thornton model, the energy per deposited atom should be about 20 eV to complete the disruption of the columnar morphology of the growing film for the maximum density [35.46, 47]. A variety of techniques have been developed to provide additional nonthermal activation at the film growth surface. In the low-pressure process of ion-beam-assisted deposition (IBAD), an auxiliary source of energetic ions is used to concurrently bombard the surface of the growing film. The effectiveness of IBAD combined with PLD has also been demonstrated elsewhere [35.48]. In plasma-based techniques such as direct-current (DC) or radiofrequency (RF) sputtering deposition, substrate biasing is used to extract and accelerate the energetic ions directly from the deposition plasma [35.29].

The above considerations show several means that can be used to control the energetics of the PLD- and PED-generated plasma flux for thin-film deposition. Accepting particles with energy of ≈ 10 eV as the most favorable for nonthermally activated film growth, one can see that most of the plasmas produced by laser or electron beam pulses are excessively energetic, especially at the higher-energetic end. The reason for this higher-energy tail is that, especially in the common nonsteady pulsed ablation regime, much larger energy/atom than the target's atom cohesive energy of ≈ 5–10 eV is loaded into the target. This is necessary to facilitate reproducible stoichiometric ablation, far from its threshold. The pulsed nature of the ablation leads to a broad energy spectrum that includes a significant portion of the extra-energetic particles generated at the front of generated plasma flux.

Film deposition with both the PLD and PED techniques includes two processes that are very different in their requirements, and yet both have to be facilitated. First is the ablation of the target material, which is necessarily a highly nonequilibrium process. Second is the film condensation process, which is desirable to be close to equilibrium (i. e., providing the surface with nonthermal excitation without causing surface damage). Thus, the general objective in optimizing the film deposition process is to transform the original energetic ablation plasma flux into a flux most suitable for the film growth.

As discussed above, some control over the initial particles' energy can be accomplished at the level of the energy source. Minimization of the particles energy leads to the choice of a shorter-wavelength laser (248 nm KrF excimer, for example) and energy density not too far above the ablation threshold. Another possibility is to thermalize the flux with ambient gases. Proper choices of gas pressure and target-to-substrate distance can also drastically improve the resultant film quality.

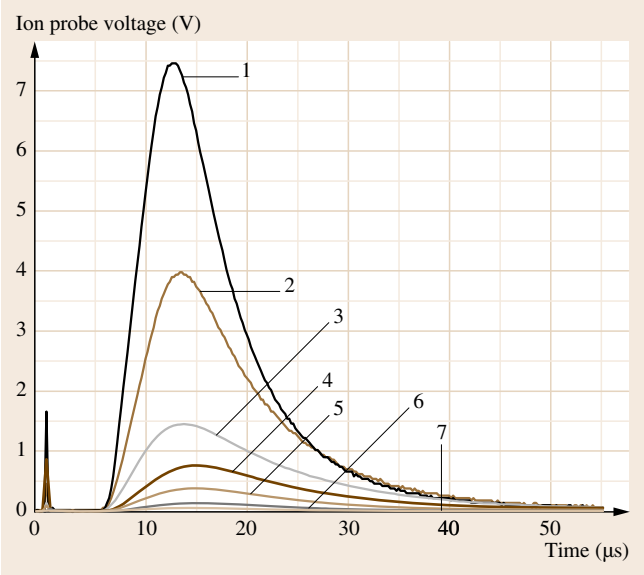


Fig. 35.9 Evolution of ion probe signal of ions propagating in a background gas with different pressure of 6 (1), 10 (2), 15 (3), 17 (4), 20 (5), 25 (6), and 30 mTorr (7) of N_2 (GaN target, KrF 248 nm laser, 17.5 cm target-to-probe distance)

35.4.1 Ion Current of Plasma Propagating in Ambient Gas

The ion probe signal reflects the dynamics of the ionized component arriving at a substrate surface, regardless of whether it is in a vacuum or an ambient gas. Typically, the maximum intensity of the ion current is created by the relatively small amount of faster-moving ions on the front of the plasma flux. Therefore, the maximum probe signal represents the dynamics of this flux component rather than that of the bulk plasma. The probe signal from the relatively slow but numerous ions with a lower energy is small, especially for ablation in a background gas. Still, those particles contribute to most of the film formation. The average energy of the arriving particles cannot be obtained directly from the probe signal when measured in a background gas pressure. However, the dependence of the probe signal on the gas pressure provides useful information on the degree of interaction between the particles of the plasma flux and the background gas atoms.

A typical evolution of the probe signal is given in Fig. 35.9 for GaN film formation with a KrF laser at a distance of 17.5 cm from the target. The signal amplitude decreases with pressure, as expected. However, the energy of the faster ions (taken as the flight time of the ions contributing to the signal at half-maximum) is nearly independent of the pressure. This can be understood by considering that the number of the fastest ions is small, and they propagate individually, unable to create a hydrodynamic ensemble with gas atoms. The fraction of ions passing a distance L without a collision is $F \propto \exp(-\alpha PL)$, where the constant α is directly related to an effective cross-section of ion collisions with gas atoms. Evidently, the *surviving* ions arrive at the probe with their original energy. However their number exponentially decreases with pressure. The $F(P)$ dependence can be used to estimate the cross-section. This portion of ballistically moving ions is also seen as *fast components* in the emission studies of the plasma dynamics [35.49,50]. A single collision of a fast ion significantly reduces the energy, and most likely changes its direction away from the aperture of the ion probe.

From the viewpoint of film growth, this implies that the fast ions cannot be completely removed from the plasma flux arriving at the substrate. Their number can only be exponentially decreased by increasing the gas pressure and/or distance from the plasma source. At some degree of filtering out the fast ions by the ambient gas, their fraction in the film growth flux can become acceptably small.

Propagation of the bulk plasma becomes hydrodynamic as it incorporates a gas mass comparable to the plasma mass. Dynamics of the deceleration in this slow flux component can be seen in time-resolved emission

measurements [35.49]. Due to the much lower ion velocity (ion current), the probe signal generated by the bulk plasma is smaller than the first peak. However, at some conditions, the decelerated ions can be seen as a second peak in the probe signal (plume splitting) [35.51, 52]. The amount of ion energy lost per collision depends on the mass ratio of the ion and the gas atom. For heavier gases, a single collision can significantly drive down the ion energy and it appears in the slower, second peak. The bulk plasma propagation distance increases with time as $\sim t^{2/5}$ in a point-blast model [35.16,53] or $\sim t^{1/4}$ in the simple model [35.34].

PLD and PED share the common feature of energetic particles in the as-generated plasma flux. Thus, both techniques require some *dumping* of the energy in a process gas of suitable pressure to adjust it for optimal film growth conditions. The main difference between the methods is in the dynamic range of the gas pressure, and the ability to ablate different materials. In contrast to PLD, with its very broad range of allowed gas pressures, PED operates in a rather narrow pressure interval of 5–10 mTorr. Accordingly, it requires more attention to optimizing target-to-substrate distance.

The main advantageous feature of PED is its ability to ablate materials based on wide-bandgap (SiO_2, Al_2O_3, MgO, etc.) dielectrics, which is difficult for PLD due to their transparency at (excimer) laser wavelengths. Due to strong absorption of electrons and relatively low thermal conductivity, PED ablation of these dielectrics results in films of higher quality and with fewer droplets. Furthermore, PED easily creates plasmas of polymers.

Thus, PLD and PED can be considered as complementary techniques, sharing a similar arrangement, and able to be explored in a unified deposition chamber. A PLD–PED deposition system enable in situ deposition of a wider range of materials, which is important the for exploration of complex structures, incorporating metals, complex dielectrics, ferroelectrics, semiconductors, glasses, etc.

35.4.2 Optimization of Growth of GaN Films – A Materials Example

Wide-bandgap GaN thin-film growth has been intensively researched due to this material's excellent blue-light emitting characteristics. Solid-state lighting-based applications have fueled a large amount of research in this area. A variety of deposition techniques have been used, including chemical vapor deposition (CVD) [35.54], molecular-beam epitaxy (MBE) [35.55], vapor-phase epitaxy (VPE) [35.56], pulsed laser deposition (PLD) [35.57], and sputtering (SP) [35.58]. PLD-produced GaN films have been reported with extremely high crystalline quality, but only a few groups have been successful in demonstrating photoluminescence in these films, usually at low temperature (≈ 12 K) only [35.57], or in films with an associated broadband *yellow* emissions indicative of impurity states in these films [35.59].

Table 35.1 PLD process parameters for AlN and GaN films on *c*-axis Al_2O_3 substrate

Parameter	AlN	GaN
Temperature	930 °C	950 °C
Target distance	20 cm	20 cm
Deposition rate	0.1 Å per pulse	0.08 Å per pulse
Energy density	2.5 J/cm^2	3.1 J/m^2
Gas pressure	0.5 mTorr NH_3	1–50 mTorr N_2

While PLD has the advantage of producing a very energetic plasma [35.60], the kinetic energy of particles in the plasma flux can be too high in the as-generated plasmas and could potentially deteriorate the electronic properties of the deposited films. To demonstrate the importance of controlling the plasma energetics in tailoring the electronic properties, a material example is presented in this section. The process optimization for the fabrication of GaN films, performed at Neocera, is presented. The plasma energy was optimized by a systematic variation of gas pressure at fixed ablation intensity and target-to-substrate distance. Room-temperature photoluminescence accompanied by excellent crystalline quality indicate that careful optimization of plasma energetics during film growth is critical for realizing optimum electronic and structural properties during PLD film growth.

GaN films by PLD were fabricated on single-crystalline *c*-axis-oriented Al_2O_3 substrates with an AlN buffer layer. A background of NH_3 and N_2 was used during depositions. For process optimization, the plasma energy was tailored to an optimal level by gas-phase collisions with the ambient gas. Excellent photoluminescence was observed at room temperature for GaN films fabricated under optimized conditions. The general PLD process conditions are presented in Table 35.1.

During optimization of film growth, increasing the nitrogen gas pressure during film growth systematically reduced the plasma energy. The best GaN films were obtained at nitrogen gas pressure of 30 mTorr. Both the AlN buffer layers and GaN layers were grown epi-

Table 35.2 Impact of gas pressure on GaN film parameters

Process pressure (mTorr)	ω-Scan width (deg)	Room-temperature photoluminescence
1	0.47	no
15	0.45	no
30	0.38	yes
50	1.78	no

taxially on the c-axis-oriented single-crystalline Al_2O_3 substrates. For 450 nm thick GaN films with AlN buffer layers, the film crystal structures were measured by four-circle x-ray diffractometer; the θ–2θ scan is illustrated in Fig. 35.10.

c-Axis-oriented GaN(0002) and AlN(0002) peaks were clearly visible at 2θ of 34.65 and 36.18°, respectively. Comparing with the Al_2O_3 single-crystal peak ($2\theta = 41.75°$), the GaN and AlN peaks were strongly intensified, suggesting high quality of the GaN and AlN layers. In the optimized films deposited at 30 mTorr N_2, full-width at half-maximum (FWHM) of 0.38 and 0.48° was obtained from ω-scans of GaN(0002) and AlN(0002) respectively, indicating highly c-axis-oriented layers. In addition, Φ-scans were performed for GaN($10\bar{1}1$), AlN($10\bar{1}1$), and Al_2O_3($11\bar{2}3$) peaks, as illustrated in Fig. 35.11. The repeated peak-to-peak 60° distance indicates the hexagonal structure of the film layers, and the clear matching peak positions of GaN, AlN, and Al_2O_3 confirm that GaN and AlN layers are epitaxial with regards to the (0001)Al_2O_3 substrate.

Photoluminescence (PL) was measured at room temperature and is illustrated in Fig. 35.12. A nitrogen laser with wavelength of 340 nm was used as the excitation source, and room-temperature PL intensity was plotted with respect to wavelength. For GaN films deposited at pressures lower than 30 mTorr excellent crystallinity was exhibited, but no PL was observed at room temperature. Room-temperature PL was only observed when the chamber pressure for GaN deposition was 30 mTorr. Blue luminescence was detected at 370 nm with significant peak intensity. The yellow emission at 545 nm was very weak, which was mainly attributed to structural imperfections such as grain boundaries or dislocations [35.59, 61]. The peak-to-peak ratio of blue emission to yellow emission is superior to other results from GaN films processed by CVD [35.61], MBE [35.55], VPE [35.62], SP [35.63], and in several nonoptimal PLD films [35.57, 59].

Table 35.2 summarizes the influence of ambient gas pressures on the GaN film quality as measured by FWHM of ω-scans and room-temperature photoluminescence (RTPL). RTPL appears to be extremely sensitive to the energy of plasma particles, being observed only from films deposited at pressure

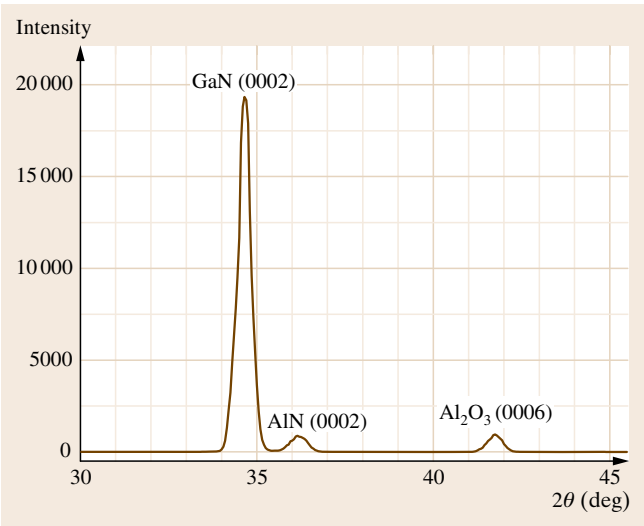


Fig. 35.10 θ–2θ scan of GaN grown under the conditions of Table 35.1 on (0001)Al_2O_3 substrate with AlN buffer layer

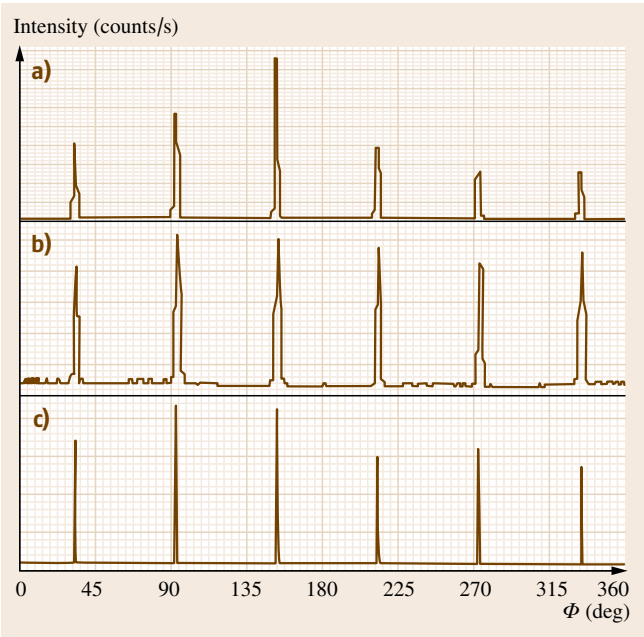


Fig. 35.11a–c Φ-scans for GaN/AlN/Al_2O_3 structure: GaN($10\bar{1}1$) peak **(a)**, AlN($10\bar{1}1$) peak **(b)**, Al_2O_3($11\bar{2}3$) peak **(c)**

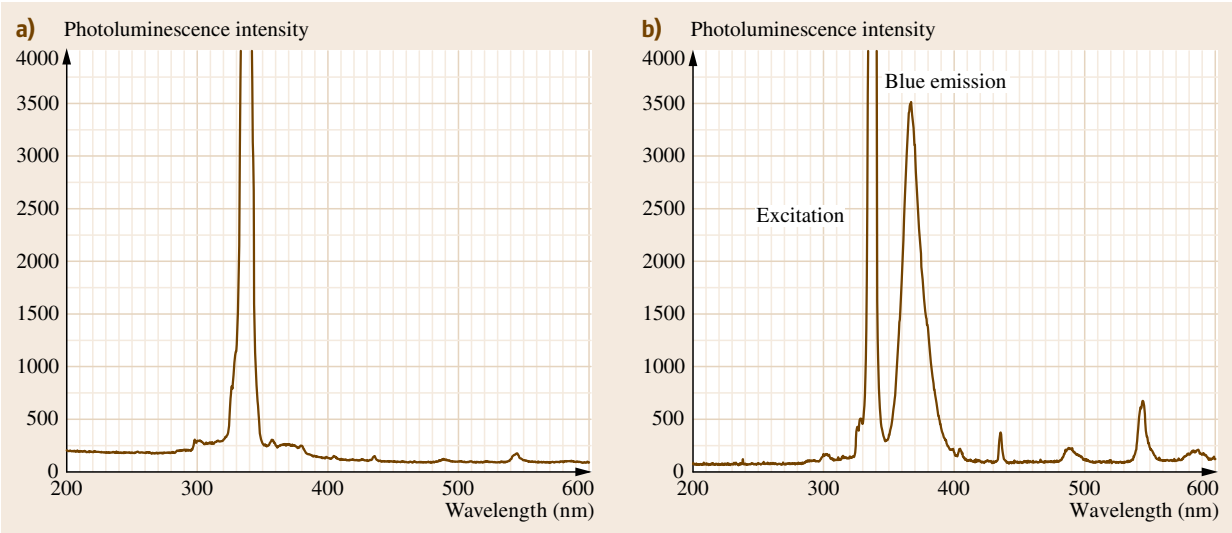


Fig. 35.12a,b Photoluminescence spectrum of PLD GaN films deposited at 14 mTorr **(a)**, and 30 mTorr **(b)**

≈ 30 mTorr. There was no RTPL in GaN deposited at lower pressures. This can be related to the excessive energy of arriving particles. However, remarkably, room-temperature PL was not observed in film deposited at 50 mTorr. Thus, some degree of nonthermal activation (present in the 30 mTorr process) is indeed necessary to obtain optimal film properties, such as photoluminescence at room temperature. ω-Scan width is another parameter exhibiting the pressure dependence, with the best performance in the 30 mTorr process. As seen from the Table 35.2, ω-scan has the minimal FWHM at this pressure, implying that GaN crystallinity was significantly degraded at higher pressure of 50 mTorr.

Comparison of the optimal conditions with the expected range (Fig. 35.8, asterisk) of the plasma flux (Table 35.1) in 30 mTorr nitrogen shows that the substrate position of 20 cm corresponds to a distance from the target greater than the expected ≈ 12 cm. Thus, the optimal average energy of the arriving particles can be estimated as lower than expected from the model [35.34]. The initial mean energy of the Ga ions is ≈ 50 eV (from Fig. 35.9 data at low pressure). The mean energy at the 12 cm is (0.02–0.04) × 50 eV ≅ 1–2 eV, according to the model. The estimated mean energy at the optimal 20 cm distance is lower but roughly agrees with that predicted by the model.

35.5 Conclusions

Thin-film deposition by condensation of energetic plasma combines two events. First, material from a target is ablated in a highly nonequilibrium process, resulting in the generation of particles with kinetic energies orders of magnitude above those observed in equilibrium evaporation. This rapid (pulsed) evaporation of material is critically needed to achieve a congruent, stoichiometric flux from a multicomponent target. The second event is the film formation on a substrate that requires critical control of the plasma energetics leading to an optimal material flux arriving at the substrate. The optimal films can be deposited at energetic conditions that are not too far from equilibrium. Lower deposition rates and particle energies below the point-defect generation threshold seem to aid the formation of optimal films. In a vacuum process, these two contradicting requirements (high-energy ablation plasma near the target surface and a lower-energy materials flux at the substrate surface) can only be satisfied partially by choosing a shorter-wavelength laser (in the case of PLD) and lower beam intensity near the ablation threshold. A smaller material flux and particle energy

can be achieved in this way. An appropriate process gas in a deposition chamber can serve as a more effective tool for converting the high-density high-energy ablation plasma flux into the required low-density low-energy condensation flux. By controlling the deposition conditions, low-temperature film growth activated by particles with optimal energy can be achieved. Careful optimization of the entire deposition system design and process parameters facilitates growth of high-quality single- and multicomponent films by both PLD and PED. The main advantageous feature of PED is its ability to ablate materials based on wide-bandgap (SiO_2, Al_2O_3, MgO, etc.) dielectrics, which are difficult for PLD due to their transparency at (excimer) laser wavelengths. Thus, PLD and PED can be considered as complementary techniques, sharing a similar arrangement, and able to be explored in a unified deposition chamber. PLD–PED deposition systems enable in situ deposition of a wider range of materials, which is important for the exploration of complex structures that incorporate metals, complex dielectrics, ferroelectrics, semiconductors, and glasses.

References

35.1 D.M. Mattox, J.E. McDonald: Interface formation during thin film deposition, J. Appl. Phys. **34**, 2493–2496 (1963)

35.2 J.S. Colligon: Energetic condensation: processes, properties, and products, J. Vac. Sci. Technol. A **13**, 1649–1657 (1995)

35.3 H. Fehler: Vacuum **45**, 997–1000 (1994)

35.4 W.D. Sproul: Mater. Sci. Eng. A **163**, 187–190 (1993)

35.5 J.M.E. Harper, J.J. Kuomo, H.R. Kaufman: J. Vac. Sci. Technol. **21**, 737 (1982)

35.6 D.B. Chrisey, G. Hubler (Eds.): *Pulsed Laser Deposition of Thin Films* (Wiley, New York 1994)

35.7 H.M. Smith, A.F. Turner: Vacuum deposited thin films using a ruby laser, Appl. Opt. **4**, 147–148 (1965)

35.8 K.L. Saenger: Pulsed laser deposition, Part 1, Process. Adv. Mater. **2**, 1–24 (1993)

35.9 K.L. Saenger: Pulsed laser deposition, Part 2, Process. Adv. Mater. **3**, 63–82 (1993)

35.10 J. Schou: Laser beam-solid interactions: fundamental aspects. In: *Materials Surface Processing by Direct Energy Techniques*, ed. by Y. Pauleau (Elsevier, Oxford 2006) pp. 35–66

35.11 G. Müller, M. Konijnenberg, G. Krafft, C. Schultheiss: Thin film deposition by means of pulsed electron beam ablation. In: *Science and Technology of Thin Films*, ed. by F.C. Matacotta, G. Ottaviani (World Scientific, Singapore 1995) pp. 89–119

35.12 D. Bäuerle: *Laser Processing and Chemistry* (Springer, Berlin, Heidelberg 2000)

35.13 M. Strikovski, K.S. Harshavardhan: Parameters that control pulsed electron beam ablation of materials and film deposition processes, Appl. Phys. Lett. **82**, 853–855 (2003)

35.14 A. Bogaerts, Z. Chen: Effect of laser parameters on laser ablation and laser-induced plasma formation, Spectrochim. Acta A **60**, 1280–1307 (2005)

35.15 Y.P. Raizer: *Gas Discharge Physics* (Springer, Berlin, Heidelberg 1997)

35.16 Y.B. Zel'dovich, Y.P. Raizer: *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena* (Academic, New York 1966)

35.17 H. Schittenhelm, G. Callies, P. Berger, H. Hügel: Investigation of extinction coefficient during excimer laser ablation and their interpretation in terms of Rayleigh scattering, J. Phys. D Appl. Phys. **29**, 1564–1575 (1996)

35.18 V.I. Mazhukin, V.V. Nossov, M.G. Nickiforov: Optical breakdown in aluminum vapor induced by ultraviolet laser radiation, J. Appl. Phys. **93**, 56–66 (2003)

35.19 C. Geertsen, P. Mauchien: Optical spectrometry coupled with laser ablation for analytical applications on solids. In: *Application of Beams in Materials Technology*, ed. by P. Misaelides (Kluwer, Dordrecht 1995) pp. 237–258

35.20 S. Metev: Process characteristics and film properties in pulsed laser deposition. In: *Pulsed Laser Deposition of Thin Films*, ed. by D.B. Chrisey, G. Hubler (Wiley, New York 1994) pp. 255–264

35.21 L. Torrisi, S. Gammino, L. Ando, V. Nassisi, D. Doria, A. Pedone: Comparison of nanosecond laser ablation at 1064 and 308 nm wavelength, Appl. Surf. Sci. **210**, 262–273 (2003)

35.22 H. Puell, H.J. Neusser, W. Kaiser: Heating of laser plasma generated at plane solid targets, Z. Naturforsch. A **25**, 1807–1815 (1970)

35.23 S.V. Gaponov, M.D. Strikovski: Formation of plasma during vaporisation of materials by the radiation of a CO_2 TEA laser, Sov. Phys. Tech. Phys. **27**(9), 1127–1130 (1982)

35.24 N. Arnold, J. Gruber, J. Heitz: Spherical expansion of the vapor into ambient gas: an analytical model, Appl. Phys. A **69**, s87–s93 (1999), (suppl.)

35.25 J. Stevefelt, C.B. Collins: Modelling of a laser plasma source of amorphic diamond, J. Phys. D Appl. Phys. **24**, 2149–2153 (1991)

35.26 M. Strikovski: unpublished (1987)

35.27 R. Teghil, L. D'Alessio, A. Santagata, M. Zaccagnino, D. Ferro, D.J. Sordelet: Picosecond and femtosecond pulsed laser ablation and deposition of quasicrystals, Appl. Surf. Sci. **210**, 307–317 (2003)

35.28 L. D'Alessio, A. Galasso, A. Santagata, R. Teghil, A.R. Villani, P. Villani, M. Zaccagnino: Plume dynamics in TiC laser ablation, Appl. Surf. Sci. **208/209**, 113–118 (2003)

35.29 D.M. Mattox: *Handbook of Physical Vapor Deposition (PVD) Processing* (Noyes, Westwood 1998)

35.30 N.M. Bulgakova, A.V. Bulgakov, O.F. Bobrenok: Double layer effects in laser-ablation plasma plumes, Phys. Rev. E **62**, 5624–5635 (2000)

35.31 H.S. Kim, H.S. Kwok: Correlation between target substrate distance and oxygen pressure in pulsed laser deposition of $YBa_2Cu_3O_7$, Appl. Phys. Lett. **61**, 2234–2236 (1992)

35.32 P.E. Dyer, A. Issa, P.H. Key: An investigation of laser ablation and deposition of Y-Ba-Cu-O in an oxygen environment, Appl. Surf. Sci. **46**, 89–95 (1990)

35.33 H.S. Kwok, H.S. Kim, D.H. Kim, W.P. Chen, X.W. Sun, R.F. Xiao: Correlation between plasma dynamics and thin film properties in pulsed laser deposition, Appl. Surf. Sci. **109/110**, 595–600 (1997)

35.34 M. Strikovski, J. Miller: Pulsed laser deposition of oxides: Why the optimum rate is about 1 Å per pulse, Appl. Phys. Lett. **73**, 1733–1735 (1998)

35.35 C.C. Chang, X.D. Wu, R. Ramesh, X.X. Xi, T.S. Ravi, T. Venkatesan, D.M. Hwang, R.E. Muenchausen, S. Foltyn, N.S. Nogar: Origin of surface roughness for c-axis oriented Y-Ba-Cu-O superconducting films, Appl. Phys. Lett. **57**, 1814–1816 (1990)

35.36 A. Gupta, B.W. Hussey: Laser deposition of $YBa_2Cu_3O_{7-x}$ films using a pulsed oxygen source, Appl. Phys. Lett. **58**, 1211–1213 (1991)

35.37 S.J. Pennycook, M.F. Chisholm, D.E. Jesson, R. Feenstra, S. Zhu, X.Y. Zheng, D.J. Lowndes: Growth and relaxation mechanisms of $YBa_2Cu_3O_{7-x}$ films, Physica C **202**, 1–11 (1992)

35.38 A.T. Findikoglu, C. Doughty, S.M. Anlage, Q. Li, X.X. Xi, T. Venkatesan: DC electric field effect on the microwave properties of $YBa_2Cu_3O_7/SrTiO_3$ layered structures, J. Appl. Phys. **76**, 2937–2944 (1994)

35.39 W. Zhang, I.W. Boyd, M. Elliott, W. Herrenden-Harkerand: Transport properties and giant magneto resistance behavior in La-Nd-Sr-Mn-O films, Appl. Phys. Lett. **69**, 1154–1156 (1996)

35.40 Y.A. Bityurin, S.V. Gaponov, E.B. Klyuenkov, M.D. Strikovsky: GaAs compensation by intense fluxes of low energy particles, Solid State Commun. **45**, 997–1000 (1983)

35.41 S. Fähler, K. Sturm, H.U. Krebs: Resputtering during the growth of pulsed-laser-deposited metallic films in vacuum and in ambient gas, Appl. Phys. Lett. **75**, 3766–3768 (1999)

35.42 A. Anders: Observation of self-sputtering in energetic condensation of metal ions, Appl. Phys. Lett. **85**, 6137–6139 (2004)

35.43 Oak Ridge National Laboratory: *Atomic Data for Fusion*, Vol. 3 (1985), http://www-cfadc.phy.ornl.gov/redbooks/three/a/3a18.html

35.44 T. Scharf, J. Faupel, K. Sturm, H.-U. Krebs: Pulsed laser deposition of metals in various inert gas atmospheres, Appl. Phys. A **79**, 1587–1589 (2004)

35.45 T. Ohmi, T. Shibata: Advanced scientific semiconductor processing based on high-precision controlled low-energy ion bombardment, Thin Solid Films **241**, 159–162 (1993)

35.46 J.A. Thornton: The influence of bias sputter parameters on thick cupper coatings deposited using a hollow cathode, Thin Solid Films **40**, 335–340 (1977)

35.47 D.R. Brighton, G.K. Hubler: Binary collision cascade prediction of critical ion-to-atom arrival ratio in the production of thin films with reduced intrinsic stress, Nucl. Instrum. Methods Phys. Res. B **28**, 527–530 (1987)

35.48 K.S. Harshavardhan, H.M. Christen, S.D. Silliman, V.V. Talanov, S.M. Anlage, M. Rajeswari, J. Claasen: Low-loss $YBa_2Cu_3O_7$ films on flexible, polycrystalline-yttria-stabilized zirconia tapes for cryoelectronic applications, Appl. Phys. Lett. **78**, 1888–1890 (2001)

35.49 S.S. Harilal, C.V. Bindhu, M.S. Tillack, F. Najmabadi, A.C. Gaeris: Internal structure and expansion dynamics of laser ablation plumes into ambient gases, J. Appl. Phys. **93**, 2380–2388 (2003)

35.50 S.S. Harilal, B. O'Shay, Y. Tao, M.S. Tillack: Ambient gas effects on the dynamics of laser-produced tin plume expansion, J. Appl. Phys. **99**, 083303–1–083303–10 (2006)

35.51 R.F. Wood, J.N. Leboeuf, K.R. Chen, D.B. Geohegan, A.A. Puretzky: Dynamics of plume propagation, splitting, and nano-particle formation during pulsed-laser ablation, Appl. Surf. Sci. **127–129**, 151–158 (1998)

35.52 S. Amoruso, B. Toftman, J. Schou: Broadening and attenuation of UV laser ablation plumes in background gases, Appl. Surf. Sci. **248**, 323–328 (2005)

35.53 N. Arnold, J. Gruber, J. Heitz: Spherical expansion of the vapor into ambient gas: An analytical model, Proc. COLA'99, 5th Int. Conf. Laser Ablation, Göttingen (Springer, Berlin, Heidelberg 1999)

35.54 J. Han, M.H. Crawford, R.J. Shul, J.J. Figiel, M. Banas, L. Zhang, Y.K. Song, H. Zhou, A.V. Nurmikko: AlGaN/GaN quantum well ultraviolet light emitting diodes, Appl. Phys. Lett. **73**, 1688–1690 (1998)

35.55 D. Doppalapudi, E. Iliopoulos, S.N. Basu, T.D. Moustakas: Epitaxial growth of gallium nitride thin films on a-Plane sapphire by molecular beam epitaxy, J. Appl. Phys. **85**, 3582–3589 (1999)

35.56 T. Nishida, H. Saito, N. Kobayashi: Efficient and high-power AlGaN-based ultraviolet light-emitting diode grown on bulk GaN, Appl. Phys. Lett. **79**, 711–712 (2001)

35.57 M. Cazzanelli, D. Cole, J.F. Donegan, J.G. Lunney, P.G. Middleton, K.P. O'Donnell, C. Vinegoni, L. Pavesi: Photoluminescence of localized excitons in pulsed-laser-deposited GaN, Appl. Phys. Lett. **73**, 3390–3392 (1998)

35.58 M.P. Chowdhury, R.K. Roy, S.R. Bhattacharyya, A.K. Pal: Stress in polycrystalline GaN films prepared by R.F. Sputtering, Eur. Phys. J. B **48**, 47–53 (2005)

35.59 T. Venkatesan, K.S. Harshavardhan, M. Strikovski, J. Kim: Recent advances in the deposition of multi-component oxide films by pulsed energy deposition. In: *Thin Films and Heterostructures for Oxide Electronics*, ed. by S.B. Ogale (Springer, New York 2005) pp. 385–413

35.60 S. Ito, H. Fusioka, J. Ohta, H. Takahshi, M. Oshima: Effect of AlN. Buffer Layers on GaN/MnO Structure, Phys. Status Solidi (c) **0**, 192–195 (2002)

35.61 A.N. Red'kin, V.I. Tatsii, Z.I. Makovei, A.N. Gruzintsev, E.E. Yakimov: Chemical vapor deposition of GaN from gallium and ammonium chloride, Inorg. Mater. **40**, 1049–1053 (2004)

35.62 P.R. Tavernier, P.M. Verghese, D.R. Clarke: Photoluminescence from laser assisted debonded epitaxial GaN and ZnO films, Appl. Phys. Lett. **74**, 2678–2680 (1999)

35.63 T. Miyazaki, K. Takada, S. Adachi, K. Ohtsuka: Properties of radio-frequency-sputter-deposited GaN films in a nitrogen/hydrogen mixed gas, J. Appl. Phys. **97**, 093516–093518 (2005)

Part F Modeling

Part F Modeling in Crystal Growth and Defects

36 Convection and Control in Melt Growth of Bulk Crystals
Chung-Wen Lan, Taipei, Taiwan

37 Vapor Growth of III Nitrides
Dang Cai, Ronkonkoma, USA
Lili Zheng, Beijing, China
Hui Zhang, Beijing, China

38 Continuum-Scale Quantitative Defect Dynamics in Growing Czochralski Silicon Crystals
Milind S. Kulkarni, St. Peters, USA

39 Models for Stress and Dislocation Generation in Melt Based Compound Crystal Growth
Vishwanath (Vish) Prasad, Denton, USA
Srinivas Pendurti, Cincinnati, USA

40 Mass and Heat Transport in BS and EFG Systems
Thomas F. George, St. Louis, USA
Stefan Balint, Timisoara, Romania
Liliana Braescu, Timisoara, Romania

36. Convection and Control in Melt Growth of Bulk Crystals

Chung-Wen Lan

During melt growth of bulk crystals, convection in the melt plays a critical role in the quality of the grown crystal. Convection in the melt can be induced by buoyancy force, rotation, surface tension gradients, etc., and these usually coexist and interact with one another. The dominant convection mode is also different for different growth configurations and operation conditions. Due to the complexity of the hydrodynamics, the control of melt convection is nontrivial and requires a better understanding of the melt flow structures. Finding a proper growth condition for optimum melt flow is difficult and the operation window is often narrow. Therefore, to control the convection effectively, external forces, such as magnetic fields and accelerated rotation, are used in practice. In this chapter, we will first discuss the convections and their effects on the interface morphology and segregation for some melt growth configurations. The control of the flows by external forces will also be discussed through some experimental and simulation results.

36.1 **Physical Laws for Transport Processes** 1217
36.1.1 Conservation Equations 1217
36.1.2 Boundary Conditions 1218
36.2 **Flow Structures in the Melt** 1219
36.2.1 ZM Configuration 1219
36.2.2 Bridgman Configuration 1225
36.3 **Flow Control by External Forces** 1228
36.3.1 Steady Magnetic Field 1229
36.3.2 Rotation 1233
36.3.3 Vibration 1237
36.4 **Outlook** 1238
References 1238

Bulk crystals used in electronic and optoelectronic devices require low defects and good composition uniformity, and melt convection during growth of bulk crystals plays a crucial role. The growth interface, which is critical to defect formation, is significantly affected by heat flow. Dopant incorporation and distribution on both macroscopic and microscopic scales are affected by melt convection as well. Therefore, the control of melt flow is necessary for crystal growth in practice. This chapter aims to provide a basic understanding of melt convection and its control through a few heuristic experimental and simulation examples, focusing on the effects of melt convection on the growth interface and composition uniformity.

Various melt growth techniques have been used for the growth of semiconductor and oxide bulk crystals. They can be grouped into three categories: namely, the Czochralski, Bridgman, and zone-melting configurations, as illustrated in Fig. 36.1. The Czochralski (CZ) configuration shown in Fig. 36.1a (*left*) is the most popular growth process, especially for silicon and compound semiconductors. Dislocation-free single crystals up to 16 inch in diameter have been grown, while 8–12 inch-diameter silicon has been produced routinely for integrated-circuit applications. Due to its practical importance, research into melt convections and their control has been very extensive [36.1–4]. The melt flow in the CZ melt is dauntingly complicated. The rising buoyancy flow from the heated crucial wall turns inwards radially and joins the thermocapillary (Marangoni) flow at the melt surface. On the other hand, the rotating crystal sucks up the melt axially at the center of the melt, spinning it up in a thin Ekman layer. The radially outward flow meets the buoyancy and Marangoni flows, leading to a complicated flow structure [36.5]. The rotation of the crucible further

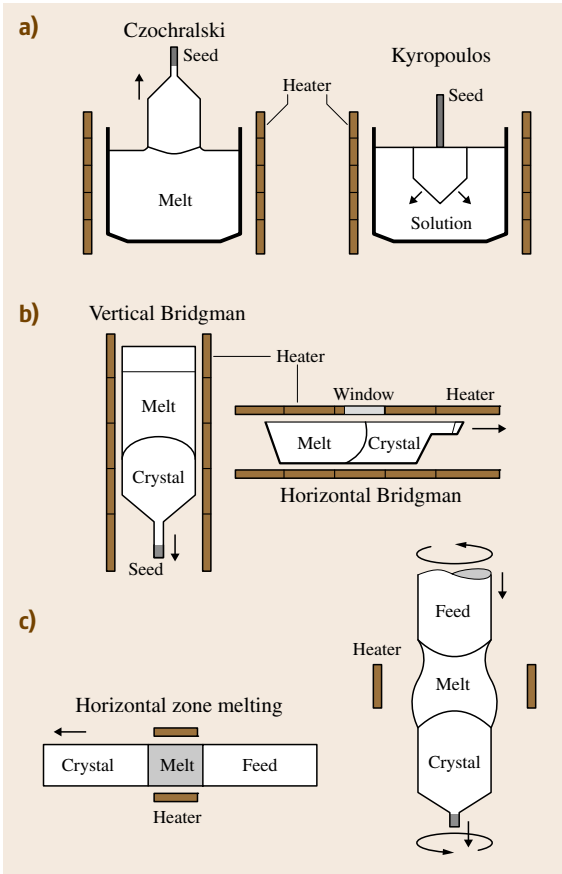


Fig. 36.1a–c Schematic of some configurations for bulk crystal growth from the melt: **(a)** Czochralski (*left*) and Kyropolous configurations; **(b)** vertical (*left*) and horizontal Bridgman configurations; **(c)** zone melting (*left*) and floating zone configurations

complicates the flow, and a Taylor–Proudman column appears under the crystal, while outside this column the melt rotates with the crucible. Such a flow easily becomes nonaxisymmetric [36.6]. Furthermore, in commercial large-scale silicon growth, the melt flow is usually turbulent. With the advance of computing power, by using the large-eddy simulation (LES), *Lukanin* et al. [36.7] successfully simulated 300 mm-diameter CZ silicon growth. The calculated thermal fields and the growth front shape are in good agreement with the measured ones. The Kyropolous method is a variant of the CZ configuration, but its crystal growth is carried out by slow cooling, usually without crystal pulling. The Bridgman and zone-melting (ZM) configurations, as depicted in Fig. 36.1b,c, respectively, are also important in applications, but the crystal size that can be grown, as compared with that from the CZ configuration, is much smaller. Nevertheless, 6 inch-diameter GaAs ingots grown by the vertical Bridgman (VB) configuration are now commercially available. Beside the general impression, their applications in compound semiconductors and oxides have increased rapidly in recent years because of the ease of operation and the lower cost in mass production. Furthermore, due to the use of the ampoule, especially for VB, very low thermal gradients are allowed and thus the thermal stress in the grown crystal is greatly reduced. Due to the thermally stable configuration, the buoyancy convection in the melt for VB growth is much weaker and simpler. The buoyancy convection is usually generated by the radial thermal gradients caused by the interface deformation or the radial heating from the furnace. Some reviews of earlier research can be found elsewhere [36.8–12]. In particular, the convections in the melt are discussed in detail by [36.10].

For the Bridgman and ZM configurations, the growth orientation can be vertical or horizontal. For the horizontal configuration [on the right of Fig. 36.1b and on the left of Fig. 36.1c for the horizontal Bridgman (HB) and horizontal ZM (HZM) configurations, respectively], an open boat is usually used due to the ease of observation. As we will discuss shortly, due to gravity being perpendicular to the thermal gradient, buoyancy convection, as well as the Marangoni convection due to the free surface, is much stronger as compared with that in the vertical configuration [36.13, 14]. Similarly, due to the radial heating, the buoyancy convection in the ZM configuration (Fig. 36.1c) is also strong. For both horizontal configurations, the growth interface is often significantly affected by the melt convection [36.14,15]. Another variation for the ZM configuration is the so-called floating-zone (FZ) process, as illustrated on the right of Fig. 36.1c. In the FZ configuration, because the molten zone is supported by its surface tension, contamination from the ampoule material can be avoided. However, due to the limitation of the zone stability, the grown crystal size is usually small, being less than 1 cm for most materials. Nevertheless, FZ growth has been a popular process for research [36.16–21]. Moreover, visualization is easy, and power consumption and material usage are small. Meanwhile, the heat flow phenomena in FZ growth are complicated and interesting. The interplay of the heat flow and the interfaces, as well as the crystal quality, can be investigated effectively. Therefore, much attention has been paid to the flow instability, especially the Marangoni flow, for the

past 20 years [36.17,20]. For the growth of silicon using induction-heated FZ method, the needle-eye configuration is used. Due to the stabilization by the magnetic force, the grown silicon crystal diameter can be up to 8 inch. The convection in the molten zone has also been investigated in detail [36.22,23].

In general, to grow crystal with good quality, careful control of convections is necessary, so that heat and mass transport are favored for dopant segregation and the shape of the growth interface. For example, a flat or slightly convex growth front toward the melt is desired to minimize parasitic nucleation. Dopant uniformity, both radial and axial, is also a major concern to ensure product quality. Therefore, the basic understanding of the role of the convections in the growth interface and the segregation is important for crystal growth. In addition, these flows affect the growth in many different ways. For the VB configuration, the convection in the melt is usually quite weak and stable. Therefore, the interface shape is mainly determined by global heat transfer and solidification [36.9,12]. Although the convection is weak, due to the low diffusivity of dopants added to the melt, dopant uniformity is influenced dramatically by the flow. Therefore, manipulating the melt flow to obtain a desired dopant distribution has become a key issue in VB growth. On the contrary, for the ZM (both vertical and horizontal) or HB configurations, the buoyancy convection or the Marangoni flow is usually strong, so that dopant can be well mixed, but the interface shape is strongly affected.

Besides the understanding of melt convections, finding an effective strategy for better growth control is also important. Tuning the temperature profile through furnace design to achieve a desired condition is typical, but this approach is laborious, rather inflexible, and not always very successful. Thus, the use of external forces to assist in controlling crystal growth is becoming popular. Examples include magnetic fields [36.24–33], accelerated crucible rotation (ACRT) [36.34–38], vibration [36.39–47], centrifugation [36.48–58], and reduced gravity [36.59]. Again, we will briefly discuss the results for achieving better growth control, in terms of both interface and dopant distribution, through the use of external forces. Although the introduction is not extensive, we shall illustrate some important concepts that help improve understanding and control of melt convections. Particularly, we shall discuss the origin of the convections through the basic conservation equations and boundary conditions for mass, momentum, energy, and species first. From the basic equations, the mechanisms for flow control by external forces can be easily understood. As discussed by *Müller* and *Ostrogorsky* [36.10], for buoyancy convection, the growth configurations can be characterized by the orientation of the body force with respect to the thermal gradients, which mainly depend on the way of heating. The crystal or crucible rotation and the Marangoni flow are imposed from the boundaries. Therefore, the governing equations and boundary conditions are indeed necessary in understanding and further controlling the convections.

In the next section, the basic governing equations and some boundary conditions are summarized. Since these equations govern the physical laws, the mechanisms and driving forces for melt convections can be easily realized; the control of melt flows by external forces can be better realized as well. In Sect. 36.2, we discuss the basic flow structures of the crystal growth processes through flow visualization experiments using a transparent material and simulation results. In Sect. 36.3, how to control the melt convections through external forces is discussed. The discussion will be focused on the Bridgman and ZM configurations. Due to space limitation, the CZ configuration is not discussed. An outlook is given in Sect. 36.4, where some recent research in melt convection and its control is briefly discussed.

36.1 Physical Laws for Transport Processes

36.1.1 Conservation Equations

The heat flow and dopant transport in crystal growth from the melt are governed by the conservation of momentum, energy, and mass. In the melt, one can consider in general an incompressible heat problem in a rotational field having a constant angular speed Ω and a uniform magnetic field $\boldsymbol{B}$; the crystal is rotating at a speed Ω_c. In a dimensionless form, the conservation laws for mass, momentum, energy, and dopant based on the Boussinesq approximation for laminar flow can be expressed by the following equations in a rotating frame [36.51,56]:

$$\nabla v = 0\,, \tag{36.1}$$

$$\frac{\partial v}{\partial \tau} + v\nabla v = -\nabla P + \mathrm{Pr}\nabla^2 v + \boldsymbol{F}\,, \tag{36.2}$$

$$\frac{\partial T}{\partial \tau} + v\nabla T = \nabla^2 T + q \,, \tag{36.3}$$

$$\frac{\partial C}{\partial \tau} + v\nabla C = \frac{\mathrm{Pr}}{\mathrm{Sc}}\nabla^2 C \,, \tag{36.4}$$

where the dimensionless body force $\boldsymbol{F}$ is given as

$$\begin{aligned}\boldsymbol{F} = &-\mathrm{Pr}\left[\mathrm{Ra_T}\left(T - T_{\mathrm{ref}}\right) + \mathrm{Ra_S}\left(C - C_{\mathrm{ref}}\right)\right] \\ &\times\left[e_{\mathrm{g}} - \mathrm{Fr}e_{\mathrm{c}}\right] + \mathrm{Ha}^2\mathrm{Pr}\left(-\nabla\Phi + v\boldsymbol{e}_{\mathrm{B}}\right) \\ &\times\boldsymbol{e}_{\mathrm{B}} - \mathrm{Ta}^{1/2}\mathrm{Pr}e_{\Omega}\times v \,,\end{aligned}$$

where v, τ, P, T, and C are dimensionless velocity, time, pressure, temperature, and dopant concentration, respectively; $\boldsymbol{e}_{\mathrm{g}}$, $\boldsymbol{e}_{\mathrm{c}}$, $\boldsymbol{e}_{\Omega}$, and $\boldsymbol{e}_{\mathrm{B}}$ are unit vectors in the gravitational, centrifugal, angular rotational, and magnetic field directions, respectively; and T_{ref} and C_{ref} are the dimensionless reference temperature and dopant concentrations, respectively. The dimensionless thermal Rayleigh $\mathrm{Ra_T}$, Hartmann Ha, Taylor Ta, and Froude Fr numbers and their physical meaning are summarized as

$$\mathrm{Ra_T} = \frac{\beta_{\mathrm{T}} g \Delta T L^3}{\nu_{\mathrm{m}}\alpha_{\mathrm{m}}} = \frac{\text{Buoyancy force}}{\text{Viscous force}} \,,$$

$$\mathrm{Ta} = 4\Omega^2 L^4/\nu_{\mathrm{m}}^2 = \frac{\text{Coriolis force}}{\text{Viscous force}} \,,$$

$$\mathrm{Ha} = |\boldsymbol{B}|L(\sigma/\mu_{\mathrm{m}})^{1/2} = \frac{\text{Lorentz force}}{\text{Viscous force}} \,,$$

$$\mathrm{Fr} = \Omega^2 L/g = \frac{\text{Centrifugal force}}{\text{Gravitatonal force}} \,,$$

where L is the characteristic length, σ is the electrical conductivity, μ_{m} is the melt viscosity, and g is the gravitational acceleration. Additional dimensionless parameters are the Prandtl number $\mathrm{Pr} = \nu_{\mathrm{m}}/\alpha_{\mathrm{m}}$ and the Schmidt number $\mathrm{Sc} = \nu_{\mathrm{m}}/D$, where ν_{m} is the kinematic viscosity and α_{m} and D are the thermal and dopant diffusivities, respectively. Also, $\mathrm{Ra_S}$ is the solutal Rayleigh number, similar to the thermal one but where the driving force is based on $\beta_{\mathrm{S}}\Delta C$, where β_{S} is the solutal expansion coefficient, and ΔC the concentration difference. For molten oxides, Pr is about unity and the thermal field can be affected easily by convection. However, for molten semiconductors or metals, $\mathrm{Pr} \ll 1$ and the effect of flow on the heat transfer and the interface shape is much less. On the other hand, for most dopants in the melts, $\mathrm{Sc} \gg 1$, indicating that the dopant field and thus segregation are dominated by convection rather than by molecular diffusion.

36.1.2 Boundary Conditions

To solve the above equations, a set of proper boundary conditions is necessary [36.55]. For crystal growth, the energy and solute balances at the interfaces are particularly important and cannot be ignored. Taking the Bridgman configuration as an example, if the dimensionless ampoule moving speed is v_{amp}, the balances for energy (ignoring internal radiation) and dopant (ignoring solid-state diffusion) at the growth front are

$$\begin{aligned}0 = &\, Q_{\mathrm{c}} - Q_{\mathrm{m}} + \gamma\mathrm{St} \\ &\times\left[\left(v_{\mathrm{amp}} - \frac{\mathrm{d}h_{\mathrm{c}}}{\mathrm{d}\tau}\right)\boldsymbol{n}\boldsymbol{e}_{\mathrm{z}} + r\Omega_{\mathrm{c}}\boldsymbol{n}\boldsymbol{e}_{\phi}\right] \,,\end{aligned} \tag{36.5}$$

$$\begin{aligned}0 = &\, \boldsymbol{n}\nabla C|_{\mathrm{m}} + \frac{\mathrm{Sc}}{\mathrm{Pr}}(1-K)C \\ &\times\left[\left(v_{\mathrm{amp}} - \frac{\mathrm{d}h_{\mathrm{c}}}{\mathrm{d}\tau}\right)(\boldsymbol{n}\boldsymbol{e}_{\mathrm{z}}) + r\Omega_{\mathrm{c}}^{*}(\boldsymbol{n}\boldsymbol{e}_{\phi})\right] \,,\end{aligned} \tag{36.6}$$

respectively, where Q_{c} and Q_{m} are the normal heat fluxes in the crystal and melt sides, respectively, γ_{c} is the density ratio of the crystal and the melt, St is the Stefan number scaling the heat of fusion by the sensible heat of the melt ($\mathrm{St} = \Delta H/Cp_{\mathrm{m}}\Delta T$), h_{c} is the dimensionless interface position, $\boldsymbol{e}_z$ is the unit vector in the axial direction, K is the segregation coefficient of the dopant, $\boldsymbol{e}_{\phi}$ is the unit vector in the azimuthal direction, and $\boldsymbol{n}$ is the unit normal vector. If the interface shape is not axisymmetric ($\boldsymbol{n}\boldsymbol{e}_{\phi} \neq 0$), the freezing rate may oscillate and melt-back may even occur ($\boldsymbol{n}\boldsymbol{e}_{\phi} > 0$) during rotation if the growth rate is very small [36.15].

For the FZ and HB configurations, the free surface requires normal and tangential shear stress balances for the momentum equations. At the free surface, the shear stress balance is imposed ([36.60]

$$\tau : \boldsymbol{n}\boldsymbol{s} = \mathrm{Ma}\,\partial T/\partial s \,, \tag{36.7}$$

where $\tau : \boldsymbol{n}\boldsymbol{s}$ is the shear stress at the n–s plane of the free surface; $\boldsymbol{n}$ and $\boldsymbol{s}$ are the unit normal and tangential vectors at the free surface, respectively. Also, Ma is the Marangoni number, which is defined as

$$\mathrm{Ma} = \frac{|\partial\gamma/\partial T|T_{\mathrm{m}}L}{\rho_{\mathrm{m}}\nu_{\mathrm{m}}\alpha_{\mathrm{m}}} \,,$$

where $\partial\gamma/\partial T$ is the surface tension–temperature coefficient of the melt. Two tangential directions need to be considered for the stress balance. In addition, the kinematic condition ($\boldsymbol{n}\boldsymbol{v} = 0$) at the free surface and the normal stress balance (the Young–Laplace equation) are also satisfied, i. e.,

$$\tau : \boldsymbol{n}\boldsymbol{n} = (2H)\mathrm{Bo} + \lambda_0 \,, \tag{36.8}$$

where $2H$ is the mean curvature scaled by $1/L$ and $\mathrm{Bo} = \gamma/(\rho_m g_0 L^2)$ is the static Bond number, where γ is the surface tension of the melt. The detailed procedure for calculating the mean curvature can be found elsewhere [36.61]. Also, λ_0 is a reference pressure head that needs to be determined to satisfy the growth angle constraint for a steady growth.

The governing equations and the associated boundary conditions can be solved by numerical methods. For example, *Lan* and *Liang* [36.62] have developed a multiblock three-dimensional (3-D) finite-volume method with multigrid acceleration to solve the above equations. In each block, a structure mesh is adopted. During iterations, the interface is found by locating the melting temperature. The iterations continue until all the variables converge.

In the following sections, we will discuss first the flow structures and their effect on crystal growth through a few examples, focusing on the Bridgman and ZM configurations. These examples provide a heuristic introduction of melt flow structure and its effects on interface shape and segregation. Much more complicated cases exist for large-scale Czochralski growth of silicon, where the melt flow is usually turbulent [36.1, 7, 63]. However, due to space limitation, we do not attempt to touch on this topic. This introduction is not extensive, and the interested reader can find further discussion in the related references [36.5, 9, 10].

36.2 Flow Structures in the Melt

Flow structures in the melt have significant influence on interface morphology and dopant segregation. Oscillatory melt flows can also cause growth striations, as a result of microscopic composition nonuniformity [36.5]. Therefore, understanding the basic flow structures in the melt is important for growth control. Furthermore, beside the flow structures, the heat of fusion released during crystal growth (the Stefan effect) could significantly affect the interface shape. Therefore, the interaction of the heat flow and solidification is the key factor for the interface shape. In this section, typical convection structures in the ZM and Bridgman configurations will be discussed. The buoyancy, Marangoni, and forced convections are illustrated. Solutal effects on the melt convection will be discussed as well.

36.2.1 ZM Configuration

For the ZM configuration, if the molten zone generated by a resistance heater is confined in the ampoule, the buoyancy convection is caused by radial heating. The typical buoyancy flow is like that in Fig. 36.2a, observed during the ZM growth of a 11 mm-diameter sodium nitrate ($NaNO_3$) crystal [36.64]. As shown, due to radial heating, the hotter and lighter melt near the ampoule floats upward, while the cooler and heavier melt near the centerline sinks, thus producing the flow loop. Because $NaNO_3$ is a high-Prandtl-number material ($\mathrm{Pr} = 9.2$), the heat transfer in the molten zone is significantly enhanced. As a result, the growth interface is also affected, having a gull-wing shape.

On the other hand, with the presence of the free surface for the FZ configuration, the flow structure is dramatically changed due to the Marangoni flow. The Marangoni flow is induced by the nonuniform surface tension due to temperature variation at the free surface. For small-scale growth, the Marangoni flow could be dominant in the molten zone. Figure 36.2b shows a visualized Marangoni flow in a floating molten zone of a 4 mm-diameter $NaNO_3$ rod [36.65]. As shown, there are four flow cells near the free surface, induced by the surface-tension difference along the free surface; the surface tension is higher near the solid side with a lower temperature. Due to the flow, the heat absorbed from the heater is brought toward to the melt–solid interfaces, leading to very convex interfaces. Compared with that in Fig. 36.2a, the interface convexity is significantly larger than the buoyancy-flow induced one. The effects of buoyancy and Marangoni convections can be better understood from the simulation results shown

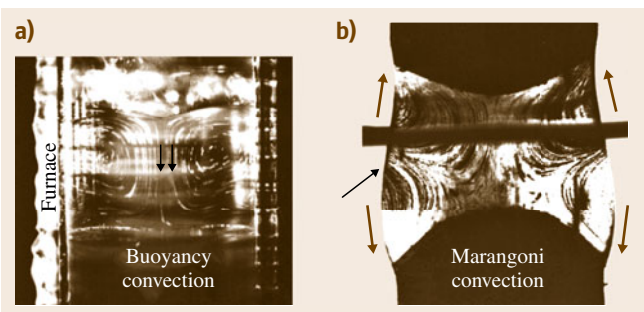


Fig. 36.2a,b Observed flow structures and zone shapes in **(a)** zone-melting growth of $NaNO_3$ (11 mm diameter, growth rate of 1 cm/h); **(b)** observed flow patterns in a stationary floating zone of $NaNO_3$ (4 mm diameter)

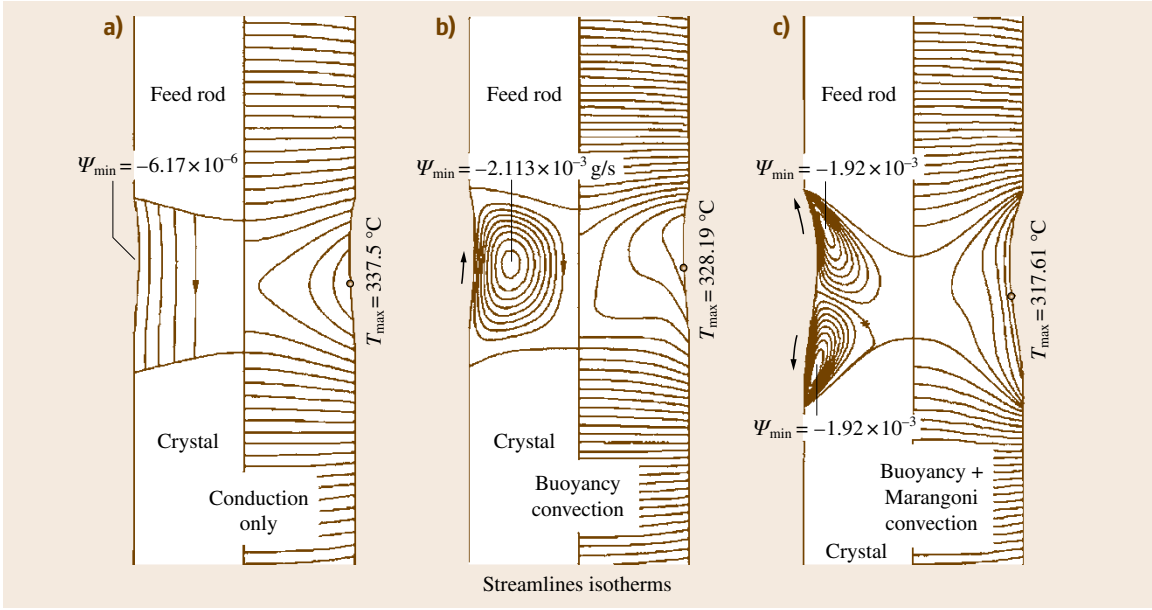


Fig. 36.3a–c Effect of convection modes on the interface shapes for FZ growth of a 4 mm-diameter $NaNO_3$ crystal: **(a)** conduction mode; **(b)** buoyancy convection; **(c)** buoyancy and Marangoni convections; the *left-hand side* shbows the flow patterns and the *right-hand side* the isotherms

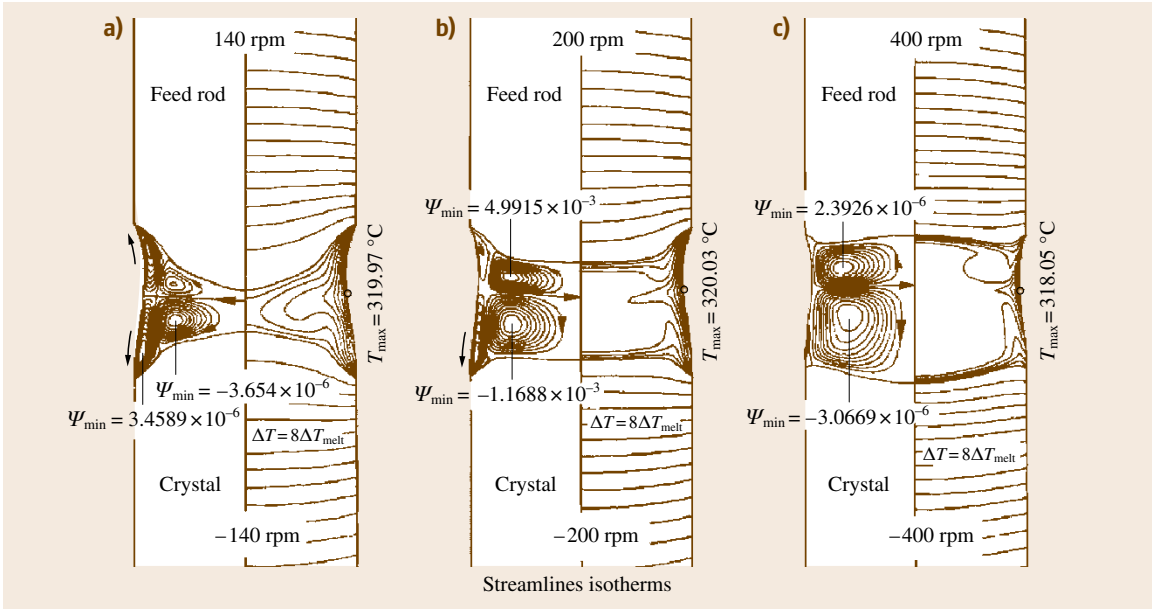


Fig. 36.4a–c Effect of counter-rotation on the flow structures and interface shapes for FZ growth of a 6 mm-diameter $NaNO_3$ crystal: **(a)** 140 rpm; **(b)** 200 rpm; **(c)** 400 rpm

in Fig. 36.3 [36.66]. In Fig. 36.3a, the streamlines are due to the rod feeding and crystal pulling (downwards). With the buoyancy flow in Fig. 36.3b, the flow structure and the resulted interface shape are similar to the ones

in Fig. 36.2a. As the Marangoni effect is considered, the interface shape becomes very convex, consistent with that observed in Fig. 36.2b.

Forced convection can also be introduced for FZ growth through rotation. In general, counter-rotating the feed and crystal rods is often used, but usually below 30 rpm. *Lan* and *Kou* [36.66] proposed an extreme situation by using high-speed rotation, say 200 rpm, for zone shape control. Figure 36.4 shows the effects of counter-rotation for a FZ growth of a 6 mm-diameter $NaNO_3$ crystal. As shown, the flow structures due to the forced convection are also very different from the previous ones. The counter-rotating rods act as a centrifugal fan, sucking the melt axially toward the solid and spinning outward radially in the thin Ekman boundary layer near the solid. Thus, the induced convection cells are near the center core of the molten zone. Because their flow directions are different from the Marangoni cells, the Marangoni flow is suppressed with increasing rotation speed. The photograph of the visualized flow at 200 rpm is shown in Fig. 36.5a. Due to the lens effect of the melt as a result of its different reflective index from the air, the Marangoni flow cannot be seen. From this side view, it can be imagined that the heat absorbed from the heater can be brought into the zone center effectively. As a result, the interfaces become flat and the zone can be kept very short by using a smaller heating power. This idea has been applied to the growth of 1 cm-diameter $NaNO_3$ crystal with a very short and stable zone [36.67]. A photograph of the growth is shown in Fig. 36.5b. Computer simulation has also been carried out, and good agreement with experimental observation has been obtained.

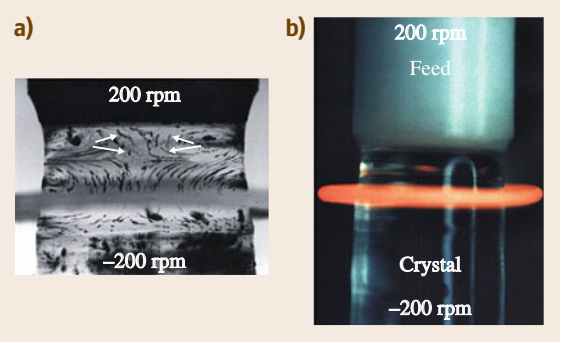


Fig. 36.5 **(a)** Observed flow patterns in FZ growth (6 mm-diameter rod) under 200 rpm counter-rotation; **(b)** FZ growth of a 1 cm-diameter $NaNO_3$ crystal under 200 rpm counter-rotation

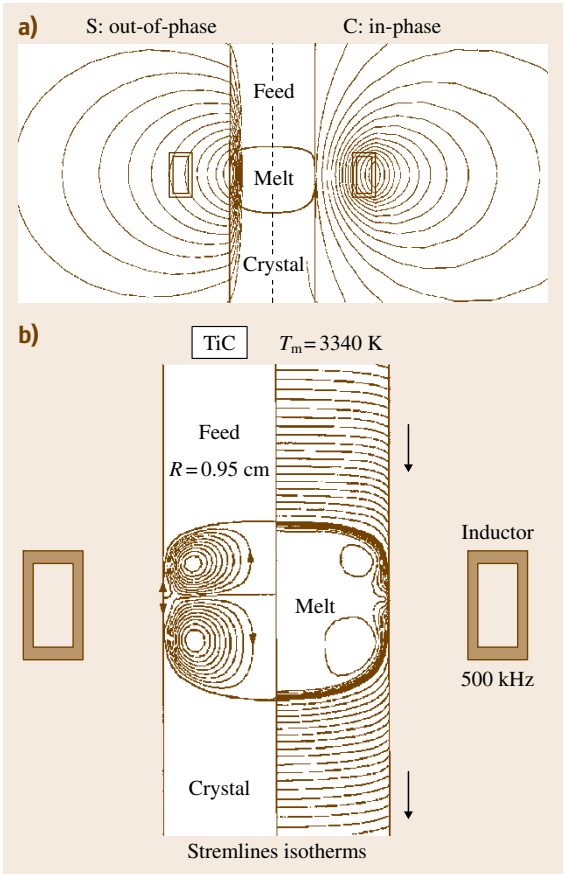


Fig. 36.6a,b Effect of induction heating for FZ growth of 2 cm-diameter TiC: **(a)** calculated in-phase (*right*) and out-of-phase (*left*) magnetic components; **(b)** calculated flow patterns (*left*) and isotherms (*right*) and the zone shape

Furthermore, for electrically conductive melt, the molten zone can also be produced by an induction coil; the heat is generated by the Joule heating of the induced eddy current. In addition, the Lorentz force can induce significant magnetic stirring and this leads to a severe distortion of the molten zone. Especially, if no insulation is used for the zone, the significant radiative heat loss from the molten zone causes a very concave growth front. Figure 36.6a shows calculated magnetic fields for FZ growth of a 2 cm-diameter TiC rod, and Fig. 36.6b shows the calculated flow (*right*) and thermal fields (*left*) as well as the zone shape. From the distortion of the isotherms, the magnetic stirring inside the molten zone is quite vigorous. The Marangoni flow in this case is weak as compared with the flow generated by the magnetic stirring. Increasing the coil frequency

reduces the skin depth and magnetic stirring. The interface concavity can be reduced as well. However, a more effective way for interface control is to reduce the radiation heat loss, for which the use of insulation or a secondary heater is helpful [36.68]. In addition, the power (induction current) required and also the magnetic stirring can be significantly reduced.

The flow structures in the previous examples are axisymmetric. However, in reality, several factors can cause symmetry breaking. Imperfect heating or alignment can cause asymmetric zone as well. Inherent nonlinear bifurcation is also typical. *Lan* and *Liang* [36.69] have demonstrated through numerical simulation the nonlinear symmetry breaking leading to three-dimensional (3-D) flows for a ZM growth for a 2 cm-diameter GaAs crystal in a quartz ampoule (wall thickness 2 mm). The bifurcation diagrams for the zone length and surface temperature differences using the heater temperature (T_p) as the parameter are illustrated in Fig. 36.7a and b, respectively. Some plots for thermal and flow fields indicated in Fig. 36.7 are further illustrated in Fig. 36.8. First, we start the solution at 1500 °C, as shown in Figs. 36.7 and 36.8a, the solution is axisymmetric and stable. On further increasing the heater temperature, we can observe a subcritical bifurcation at about 1513 °C. However, the new mode that bifurcates from the axisymmetric ($m0$) mode is an $m2$ mode instead of an $m1$ mode. Two stable solutions b and c at 1510 °C are shown in Fig. 36.8b,b′, respectively. The $m2$ mode c has a twofold symmetry, which can be clearly seen from the cut of the isotherms at the middle section of the molten zone ($z = L/2$). As shown at the bottom of Fig. 36.8b′, although they are both symmetric with respect to the centerline, the results are quite different. By comparing the interface shapes from the two cuts, one may also get a 3-D view of the interface shapes. Further increasing the heater temperature, as shown in Fig. 36.7, we encounter a secondary bifurcation, and the $m2$ mode becomes unstable. The new mode branching from $m2$ is the $m1$ mode again. This bifurcation is subcritical as well. Two stable solutions c and c′ are further illustrated in Fig. 36.8c,c′, respectively.

It is believed that the existence of the $m2$ mode at lower T_p is due to the much smaller aspect ratio of the molten zone. As the zone length is increased by increasing the heating power, the $m1$ mode becomes dominant. Furthermore, the subcriticality is due to the existence of the deformable interfaces. The bifurcation becomes supercritical if we fix the interfaces. Such an observation is similar to that in the two-phase Rayleigh–Bernard

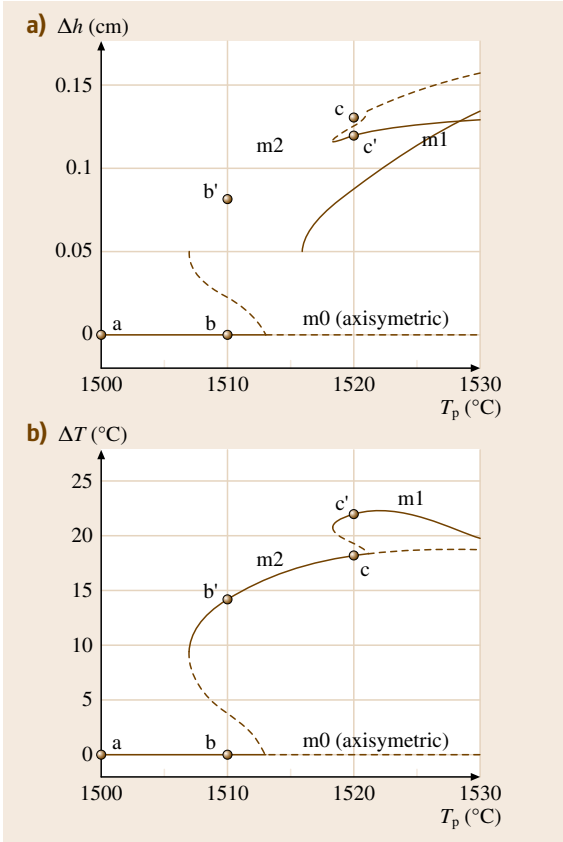


Fig. 36.7a,b Bifurcation diagrams of ZM growth of 1 inch-diameter GaAs using the heater temperature as the parameter of: **(a)** the zone length deviation and **(b)** the surface temperature deviation

problem discussed by *Lan* and *Wang* [36.70]. Again, although Ra_w [$Ra_w = (\beta_T g_0 \Delta T L^3)/(\nu_m \alpha_m)$; $\Delta T = T_w - T_m$, where T_w is the wall temperature] is up to 5.78×10^5 ($\Delta T = 63.33$ K is the maximum temperature in the molten zone) for the $m1$ mode at $T_p = 1530$ °C, the solution is still stable. Similar observations were found by *Baumgartl* et al. [36.71].

Besides the symmetry breaking, the bifurcation to unsteady state is often encountered, and this time-dependent flow is often related to growth striation; for example, in ZM growth of GaAs single crystals at $Ra_w \approx 4 \times 10^5$, 3-D striation patterns were found [36.71]. In the previous experiment, a steady axisymmetric mode bifurcated to a 3-D toroidal mode first at $Ra_w \approx 3 \times 10^4$ and then to a time-dependent mode at a higher Ra_w ($\approx 4 \times 10^4$) for a parabolic thermal profile at the melt surface. A steady one-roll

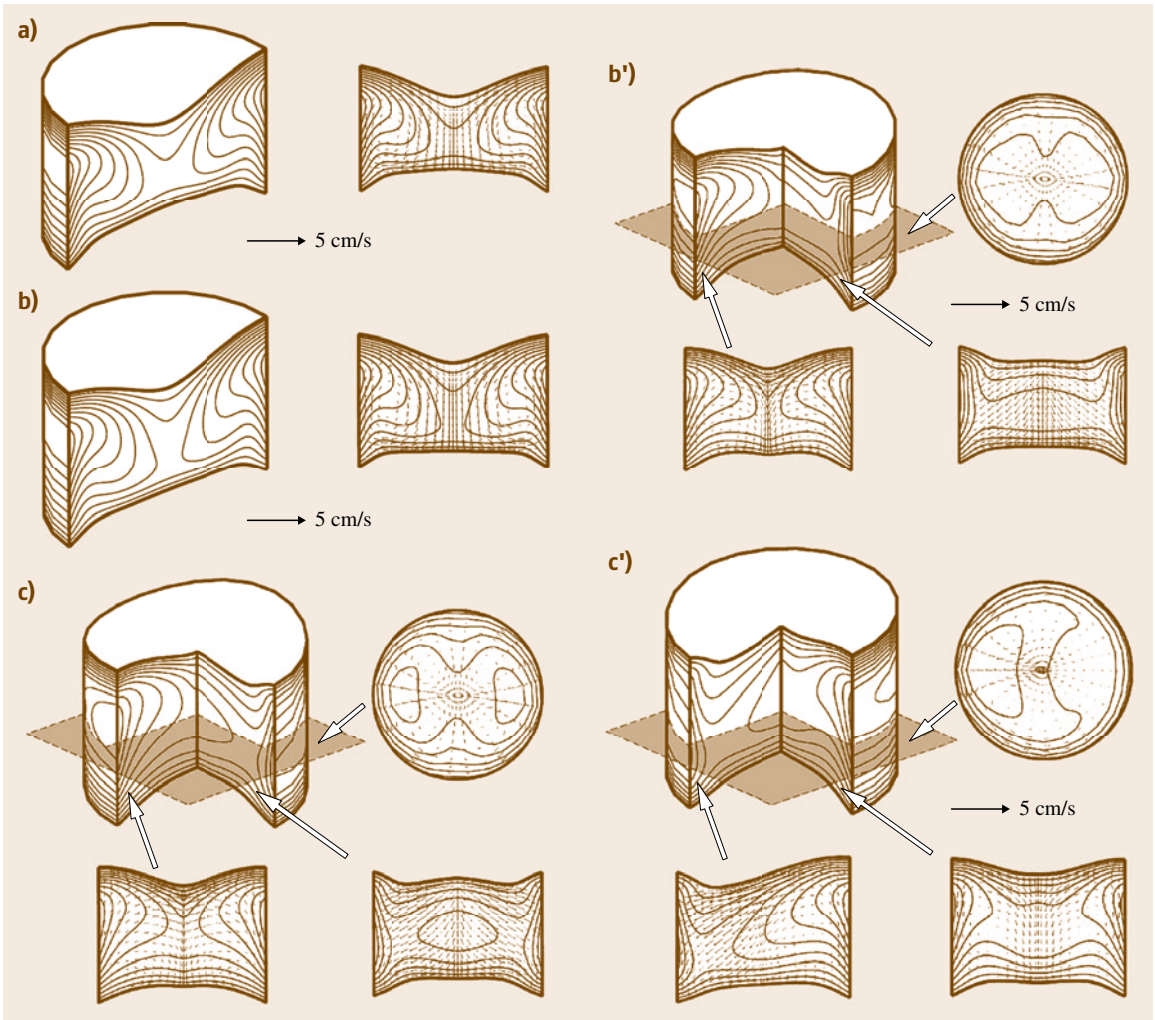


Fig. 36.8a–c Some solutions of Fig. 36.7: **(a)** 2-D solution at $T_p = 1500\,°C$; **(b)** 2-D solution at $T_p = 1510\,°C$; **(b′)** 3-D $m0$ solution at $T_p = 1510\,°C$; **(c)** 3-D $m2$ solution at $T_p = 1520\,°C$; **(c′)** 3-D $m1$ solution at $T_p = 1520\,°C$

($m1$) mode was found for 8×10^4 ($Ra_w \approx 1.85\times10^5$). Beyond $Ra_w \approx 1.85\times10^5$, the $m1$ flow became time dependent.

Similar to the nonlinear bifurcations in the ZM configuration in an ampoule, the FZ configuration also has a similar behavior. Indeed, the flow bifurcation is caused by Marangoni flow, and extensive studies have been made of this subject [36.17, 20, 21, 56].

In addition to the vertical configuration, the horizontal ZM (HZM) configuration is also used in bulk crystal growth and material purification [36.72]. The observed flow structure for a horizontal molten zone of succinonitrile (SCN) is shown in Fig. 36.9a. Due to the buoyancy force, there are two flow cells in the melt. The melt flows upwards at the center, and downwards near the interfaces. The upward melt (lighter) is also heated at the top wall and then flows to the two sides to melt back the interfaces. As a result, the upper part of the zone becomes wider. The downward melt (heavier) from the interfaces is warmed up again from the bottom wall at the center and then flows upwards. Although the picture shown in Fig. 36.9a appears to be two dimensional, the flow structure is fully 3-D. However, the flow patterns at other light-cut planes are much more difficult to observe. The trapezoidal zone shape in Fig. 36.9a is caused by the buoyancy flow and is consistent with

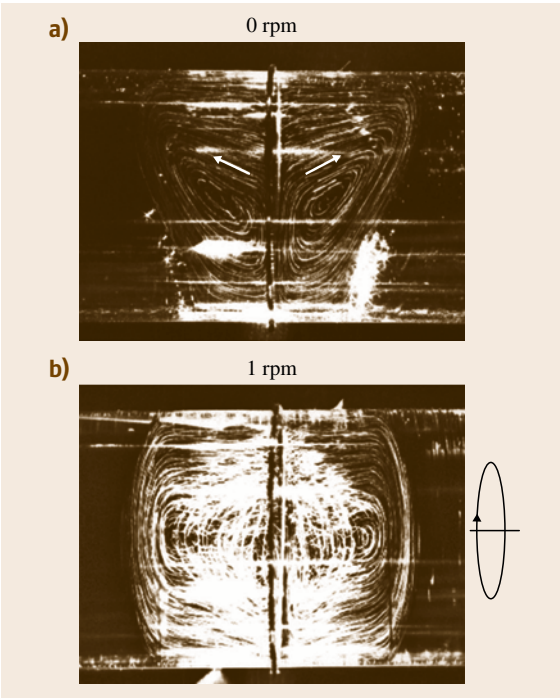


Fig. 36.9a,b Flow patterns in horizontal zone-melting growth of SCN: **(a)** without rotation; **(b)** with 1 rpm rotation

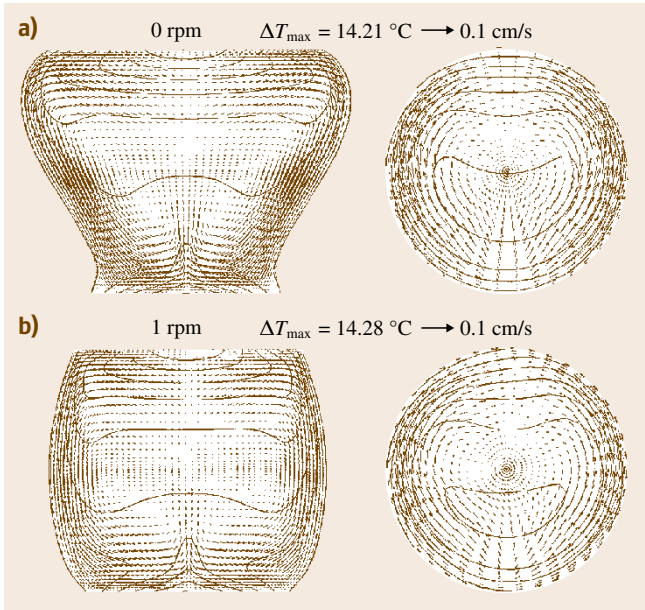


Fig. 36.10a,b Simulated flow and thermal fields: **(a)** 0 rpm; **(b)** 1 rpm rotation

that reported [36.72]. Interestingly, if we impose only 1 rpm ampoule rotation, a significant change of the zone shape is observed, as shown in Fig. 36.9b. As shown, the zone shape becomes much more symmetric, and the upper and lower zone lengths are about the same. As compared with Fig. 36.9a, the rotation reduces the upper zone length, while widens the lower part; the zone length seems to be averaged by the rotation. The flow structure is also changed quite significantly. In fact, due to rotation, the flow visualization became much more difficult. The tracer particles (aluminum powder) are heavier than the melt. As a result, larger particles tended to settle down, as shown by the bright area at the melt bottom in Fig. 36.10a.

Computer simulation has also been conducted for the observed flows and interface shape [36.73], as shown in Fig. 36.10. Indeed, the simulation captures the key features of the experiments shown in Fig. 36.9: both the flow structures and the interface shapes. For the case without rotation, as shown in Fig. 36.10a, the flow patterns and interface shapes are in good agreement with Fig. 36.9a. Due to the buoyancy force, the hotter (lighter) melt at the bottom floats upwards. It is warmed up again at the top wall, and flows to the two sides, leading to a significant back melting of the interfaces at the top. The downward melt near the interfaces is cooler and heavier, but it is heated up again by the heater at the bottom wall and floats up at the center. One can also get a better picture from the distorted isotherms, which are consistent with the flow directions. The right-hand side view of Fig. 36.10a also illustrates the buoyancy convection; the melt near the wall is heated and flows upward. One cannot see a downward flow from this plane, because the major downward flows appear near the interfaces. With 1 rpm rotation, as shown in Fig. 36.10b, the zone shape becomes much more axisymmetric, which agrees well with the observation in Fig. 36.9b. The flow structure shown in Fig. 36.10b is also similar to that in Fig. 36.9b as well. One shall pay more attention on the flow shapes and vortices. The right figure of Fig. 36.10b also indicates that, although the effect of rotation becomes more obvious, the change of the flow and isotherms is small.

The success of the full model calculation in Fig. 36.10b is due to consideration of the interface energy balance. As shown in (36.5), for steady rotation without ampoule translation, the term $\gamma_c \mathrm{St}(r\Omega_c \boldsymbol{n}\boldsymbol{e}_\phi)$ plays a crucial role. During rotation, solidification at the interface occurs at $\gamma_c \mathrm{St}(r\Omega_c n\boldsymbol{e}_\phi) < 0$, while melting occurs at $\gamma_c \mathrm{St}(r\Omega_c \boldsymbol{n}\boldsymbol{e}_\phi) > 0$; for example, for the left interface, at the front side, $\gamma_c \mathrm{St}(r\Omega_c \boldsymbol{n}\boldsymbol{e}_\phi) < 0$ and solid-

ification dominates. The heat of fusion is released to the melt and it distorts the interface toward the rotational direction. Meanwhile, the melting occurs at the back side, and the heat of fusion is extracted from the melt, making the solid interface intrude toward the melt in the rotational direction there. As a result, the zone length is averaged out with the period of rotation. This mechanism (the Stefan effect) is dominant for the interface shapes here and cannot be ignored in the simulation.

36.2.2 Bridgman Configuration

The Bridgman configuration is also widely adopted for melt growth of bulk crystals. Unlike the ZM configuration, in which significant superheating is necessary to form a stable zone, the Bridgman configuration is suitable for growth that requires low superheating or thermal gradients. For the HB configuration using an open boat, due to the free surface, the Marangoni convection is also important. Nevertheless, due to the much larger melt volume, the buoyancy flow is also strong. To understand its basic flow structure, we take the HB growth of $NaNO_3$ as an example. The top view of the growth in a transparent furnace is shown in Fig. 36.11a [36.53]. During growth, the interface is always concave, due to the release of heat of fusion and the effect of melt convection, although the latter factor is much more significant. As will be discussed shortly, the Marangoni flow is responsible for the sharp contact angle between the interface and the crucible [36.13]. Such an acute angle often induces parasitic nucleation from the crucible wall, leading to a polycrystalline growth. The visualized flow is shown in Fig. 36.11b. As shown, the buoyancy and Marangoni flows are in the same direction, and the hot melt near the surface flows from the left to the right, pushing the interface to the crystal side. Meanwhile, because of the no-slip boundary condition at the solid wall, the isotherms at the solid wall are nearly pinned by the energy balance there with the conduction in the ampoule. Accordingly, the interface becomes very concave. The calculated flow patterns and isotherms, shown in Fig. 36.11b and c, respectively, further explain the situation.

Further understanding of the flow structures can be obtained from the numerical simulation of HB growth of GaAs by *Lan* and *Liang* [36.61]. Figure 36.12 shows the effects of convection modes on the thermal and velocity fields (on the planes at $z = 0$ and $y = 0$) and the growth front (crucible puling speed is 5 mm/h). As shown in Fig. 36.12a, without any driving forces for convection, the growth front is only slightly concave;

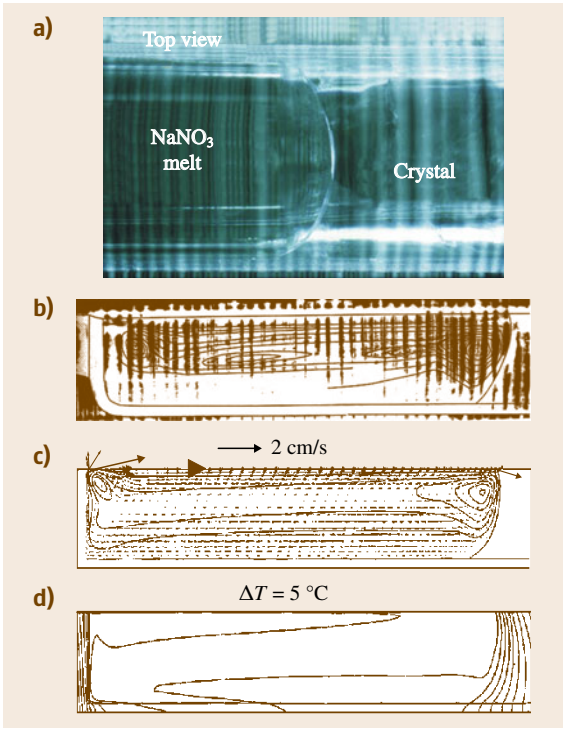


Fig. 36.11a–d Effect of buoyancy and Marangoni flows on the HB growth of $NaNO_3$ crystals: **(a)** observed interface seen from the top; **(b)** observed flow patterns; **(c)** calculated flow patterns; **(d)** calculated isotherms and interface shape

it is quit flat viewed from the top. When buoyancy force is considered in Fig. 36.12b, as shown in the side view of the flow fields, a clockwise natural convection is induced. Although the maximum melt velocity is less than 1.5 cm/s, the isotherms in the melt are highly distorted by the convection. As a result, the interface shape becomes highly distorted and very concave. The interface position at the bottom is not affected much, but the upper interface is melted backward significantly. A 3-D view of the growth front shape is illustrated on the right-hand side for comparison. Clearly, the effect of buoyancy convection on the growth front shape is significant.

When the Marangoni effect is included, as shown in Fig. 36.12c, the flow structures are not changed much; the flow in the side view is still clockwise in direction. However, the melt velocity near the free surface and the growth interface becomes much higher; the maximum velocity is about 5 cm/s. The isotherms near the free surface are thus further affected, even though the overall view of the isotherms is not changed much. As

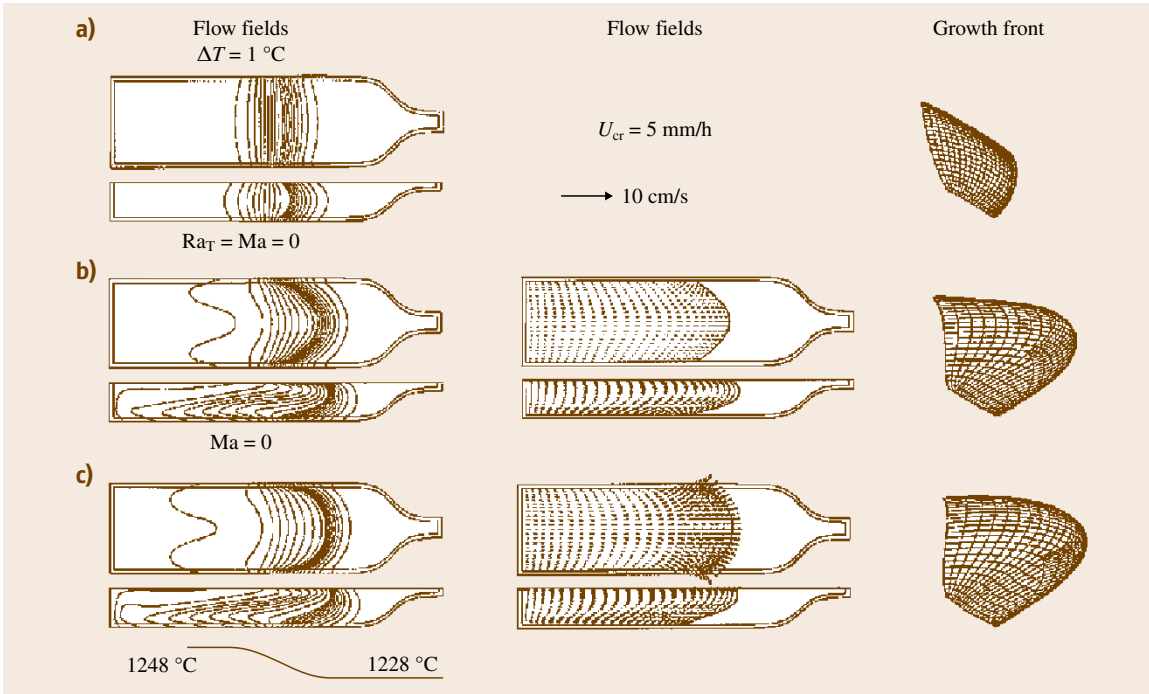


Fig. 36.12a–c Effects of buoyancy and Marangoni flows on the flow, thermal fields, and the interface shape for the HB growth of GaAs single crystal: **(a)** conduction mode; **(b)** buoyancy convection; **(c)** buoyancy and Marangoni convection

a result, the growth front near the top surface is melted back more. As shown in the side view, the growth interface also becomes sharper. Because the driving force for the Marangoni convection is proportional to local thermal gradients at the free surface, the flow near the interface is thus stronger, where the thermal gradients are higher. In addition, the flow direction there becomes more perpendicular to the isotherms, leading to a diverged flow near the interface, as shown in the top view of the flow fields. This diverged flow also melts back the interface toward the crucible and results in a rounder interface shape at the free surface when viewed from the top. More importantly, the interface–crucible contact angle becomes much smaller as well, which from the crystal growth point of view, is more likely to induce parasitic nucleation and the formation of polycrystals. This is consistent with the observation in Fig. 36.11a.

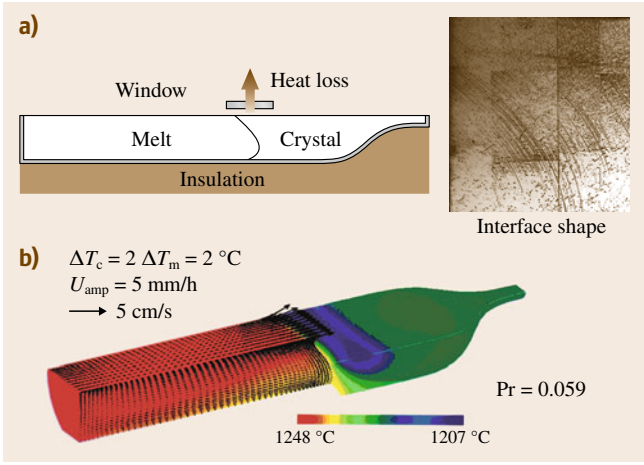


Fig. 36.13 **(a)** Schematic of HB growth of GaAs with a viewing window for local radiation cooling; on the right is the observed interface shape (cross section); **(b)** simulated flow and thermal fields

In practice, control of the growth front shape is important for crystal quality. The results in Fig. 36.12 clearly illustrate the significance of the buoyancy and Marangoni convections for the growth interface. However, since the direction of driving forces for both convection modes is not likely to be changed, the room for modulating the convection and thus controlling the interface is not great. The convections can be reduced with smaller thermal gradients, but the smaller thermal gradients also cause other problems. With a fixed growth rate, the heat of fusion will cause a more de-

formed interface shape for lower thermal gradients. Constitutional supercooling [36.73] is a concern as well.

For this configuration, an effective approach is to provide local cooling at the top interface. This concept has been adopted for the growth of GaAs using a view window above the interface, as shown on the left of Fig. 36.13a [36.61]; the observed interface shape [36.74] is shown on the right of Fig. 36.13a. This window provides local radiative cooling and thus controls the interface shape. In addition to the top radiative cooling, booster heating (through the center zone) at the bottom of the interface is also useful, and this further reduces the interface concavity of the lower interface. Computer simulation based on these ideas has been conducted [36.61] and the result is shown in Fig. 36.13b. As shown, the temperature is greatly reduced near the window and the temperature gradients are also increased significantly there; the isotherm spacing in the crystal is increased to 2 °C, which is twice that in the melt. Because the quartz crucible has a higher emissivity than GaAs, the lowest temperature beneath the window is located at the crucible surface. Interestingly, although the interface is still slightly concave, its shape is reversed; the upper interface becomes closer to the hot zone. The growth interface is now more perpendicular to the crucible wall. A similar interface shape was also observed in crystal growth experiments [36.75]. A flatter interface is possible by reducing the heat loss through the window. Local cooling using an air jet was applied to the HB growth of $NaNO_3$, and a convex interface was successfully obtained [36.14].

Unlike the previous examples with large radial thermal gradients, the vertical Bridgman (VB) configuration is thermally stable and the radial thermal gradients are small. Accordingly, the thermally driven buoyancy convection is significantly weaker; for example, if the furnace thermal profile is linear, the buoyancy flow is only induced by the interface deformation. When stationary, the interface deflection is mainly due to the difference in the thermal conductivity between the melt and the crystal. If the thermal conductivity of the melt is larger than that of the crystal, the interface is concave. The concave interface induces radial thermal gradients and thus buoyancy flow. The concavity increases with the crystal growth speed due to the release of the heat of fusion. The buoyancy convection due to the interface concavity appears near the growth interface and affects the radial segregation significantly [36.76–78].

Figure 36.14a shows the simulated buoyancy flow and solute fields in VB growth of gallium-doped germanium in a small graphite ampoule at $Ra_T = 10^7$ [36.78]; $Ra_T = 2.489 \times 10^8$ for normal gravity at a thermal gradient of 50 K/cm. Because the ampoule is pulled downward, the streamlines in the bulk melt indicate the material flow due to the ampoule translation. Also, the interface concavity is here is due to the larger thermal conductivity of the melt than that of the crystal. In addition to the thermal convection, the solutal effect can also affect the flow; for example, if a solute is added to germanium, solute segregation causes density variation in the melt, leading to solutal convection. Figure 36.14b,c show the effect of solute (silicon), in terms of solutal Rayleigh number Ra_S, on the flow. Because silicon dissolves more in the germanium solid than in the melt, silicon is depleted near the interface during crystal growth. Due to the less and lighter silicon near the interface, the melt density near the interface increases and this suppresses the thermal convection. As shown in Fig. 36.14c, the buoyancy convection is almost suppressed due to the solutal effect.

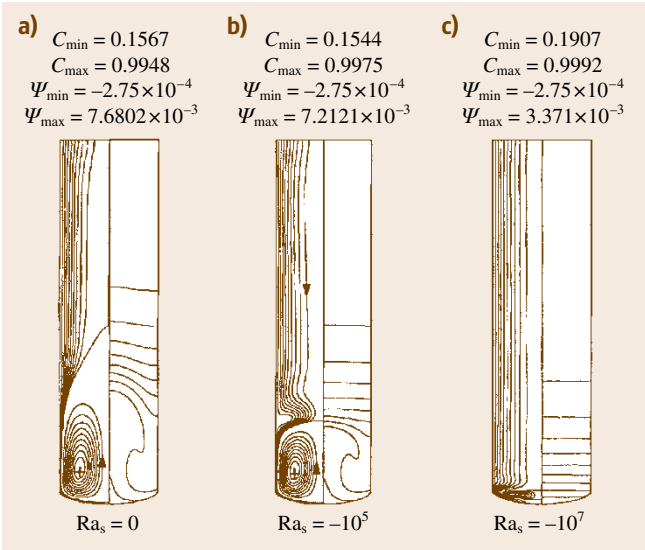


Fig. 36.14a–c Solutal effect on the buoyancy convection for VG growth of gallium-doped germanium: **(a)** $Ra_S = 0$; **(b)** $Ra_S = -10^5$; **(c)** $Ra_S = -10^7$; in each figure on the *left-hand side* of the simulated results is the flow fields and on the *right-hand side* the solutal fields

Through the visualization experiments of SCN containing acetone (lighter) or salol (heavier), the thermal–solutal convection can be better understood. Figure 36.15a,b shows the interface shape evolution during Bridgman growth of SCN containing 0.064 wt % acetone and 0.15 wt % salol, respectively; the pulling speed is 1.6 μm/s [36.47]. The simulated results are

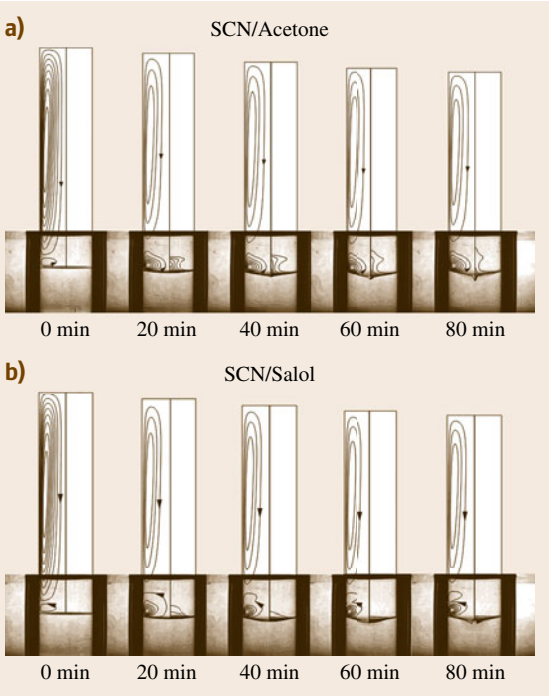


Fig. 36.15a,b Double diffusive convections in VB crystal growth of SCN: **(a)** SCN/acetone; **(b)** SCN/salol; in each figure on the *left-hand side* of the simulated results is the flow fields and on the *right-hand side* the solutal fields

also shown for comparison. In each simulated result, the left-hand side shows the streamlines, and the right-hand side the acetone fields. As shown in Fig. 36.15a for SCN/acetone stationary, the interface is flat and the convection near the interface is extremely weak. The upper cell is caused by the radial heating from the hot zone. As the solidification starts, the interface becomes concave and the flow cell near the interface is enhanced. As a result, the acetone rejected during solidification is redistributed by the flow, with increasing acetone concentration toward the center of the interface. Because acetone is lighter than SCN, the radial acetone gradients also enhance the flow, leading to a highly localized solute distribution at the center of the interface. The local acetone accumulation further causes a depression of the interface there, which becomes obvious at about 40 min. As the supercooling is built up, morphological breakdown can occur when the supercooling overcomes the interfacial energy. The simulated results at 60 and 80 min show a deep depression (pit) at the center of the interface, where high constitutional supercooling exists. The convection in the bulk melt remains about the same during crystal growth.

The evolution of interface morphology for SCN/Salol shows a similar behavior. However, as shown in Fig. 36.15b, the depression shape is wider. The morphological breakdown occurs at about the same time, i.e., about 40 min. The bottom of the breakdown area is much flatter than that for SCN/acetone. The convection near the interface is much weaker than that for SCN/acetone, as shown in Fig. 36.15b. Such a weaker flow is due to the heavier solute, which suppresses the flow. In other words, the radial density due to thermal gradients is counterbalanced by the solutal gradients; one can observe this contribution from the source term of (36.1). In addition, the flow cell near the interface is closer to the ampoule wall as compared with that in Fig. 36.15a. More importantly, the concentration profile is quite uniform near the interface, as shown by the much flatter isoconcentration lines. This also indicates that the convective effect on the solute transport is much weaker.

Because the convection in the VB configuration in the previous examples is rather weak, nonlinear flow bifurcation to symmetry breaking or oscillatory flows is not likely. Nevertheless, in reality, a perfect axisymmetric growth condition is hard to achieve. An asymmetric furnace thermal profile or slight ampoule tilting can lead to significantly 3-D convections [36.79]. As a result, the segregation behavior predicted by an axisymmetric numerical model is often erroneous. Nevertheless, since the convection is rather weak, flow suppression by using rotation or magnetic field could be quite effective. We will thus briefly discuss its flow control by using external forces.

36.3 Flow Control by External Forces

So far, we have illustrated the convections in the ZM and Bridgman configurations. As just discussed, for the FZ configuration, crystal rotation is an effective way to introduce forced convection. However, in many cases, the suppression of unstable flows to avoid growth striations and the manipulation of local flow to reduce segregation are necessary. The use of external forces is particularly effective for such purposes. Various external forces have been considered for use in the melt growth of bulk crystals. The static magnetic field is one

of the most popular ones, and has been used widely for flow suppression [36.24, 27, 31], but it is restricted to electrically conductive melt. Rotation, steady or unsteady, can also be used, and its applications have no special limitations. In the following sections, we will discuss flow control through a few examples.

36.3.1 Steady Magnetic Field

A steady magnetic field is an effective way to suppress the flow if the melt is electrically conductive. Two examples for FZ and VB configurations are given here.

As discussed previously, in the FZ configuration, the free surface of the molten zone often induces significant Marangoni flow, and it often lead to unstable heat flows, striations, and distorted interfaces, even in microgravity environment [36.21]. The control of the unsteady Marangoni flow has been an important topic in crystal growth, and has attracted extensive research over the years [36.19]. For silicon growth, magnetic fields have been known useful for suppressing unsteady flows, so that steady growth can be obtained and the grown crystal is striation free [36.18, 21]. Both axial [36.18, 21] and transversal magnetic fields [36.28, 80] have been investigated. In the FZ Si growth experiments by *Dold* et al. [36.18], a striation-free core in the grown crystal was found, and the core size increased with increasing magnetic field. It was believed that the melt inside the core was significantly suppressed by the magnetic field. *Lan* [36.81] made the first attempt to simulate the growth using an axisymmetric model. The results on the core size and dopant (phosphorous) distribution are shown in Fig. 36.16a,b, respectively. As shown in Fig. 36.16a, the streamlines are stretched along the direction of the applied magnetic field. Also, the radius of the suppressed core increases with increasing magnetic field strength, and the calculated core sizes at different axial magnetic field strengths are consistent with the measured ones. In addition, the calculated phosphorous distributions in Fig. 36.16b are consistent with the measured ones as well. Although the results based on the axisymmetric model are consistent with the experimental observations, the oscillatory and 3-D behavior cannot be seen in the simulation.

Further 3-D numerical simulations have been carried out recently [36.60]. Figure 36.17 shows the calculated results for an axial magnetic field of 0.5 T. At this condition, a steady-state result is obtained. However, as shown in Fig. 36.17, the result is not exactly axisymmetric but has a fourfold symmetry, which can be seen from the thermal and velocity fields on the z–y plane (growth direction is in the x-axis). The results on the x–y and x–z planes also show that the flow in the core region is greatly suppressed, but the thermocapillary flow is still quite strong near the melt surface. This is consistent with previous 2-D simulations shown in Fig. 36.16 [36.81]. In Fig. 36.17, the flow and thermal fields on the x–z and x–y planes happen

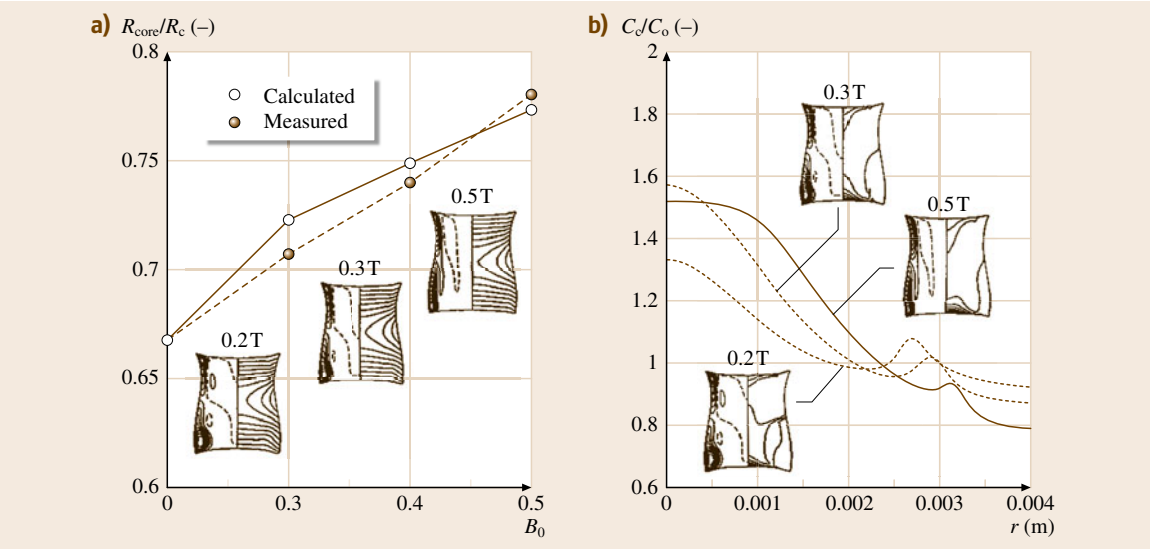


Fig. 36.16a,b Effect of axial magnetic fields on FZ silicon crystal growth: **(a)** the core size of the suppressed flow; **(b)** the radial phosphorous distribution; the simulated flow and solutal fields are also shown in **(a)** for comparison

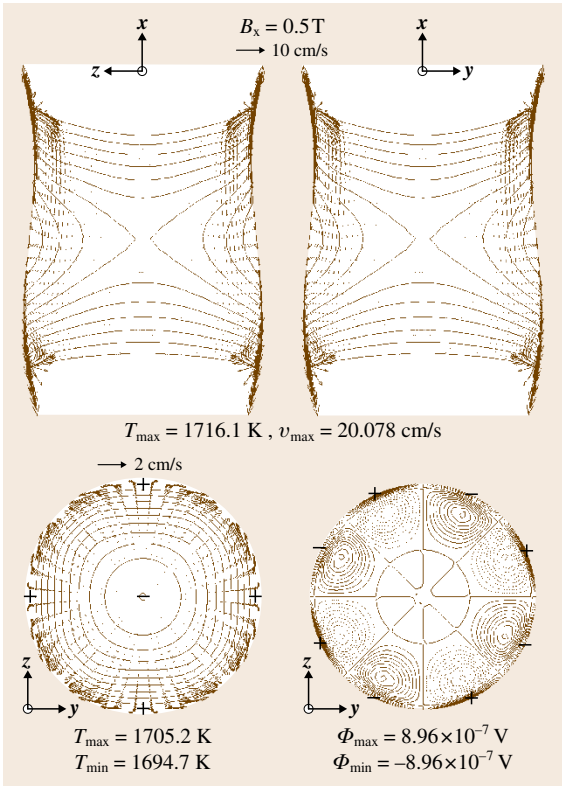


Fig. 36.17 Three-dimensional simulated flow, thermal, and potential fields in a FZ silicon growth under an axial magnetic field of 0.5 T

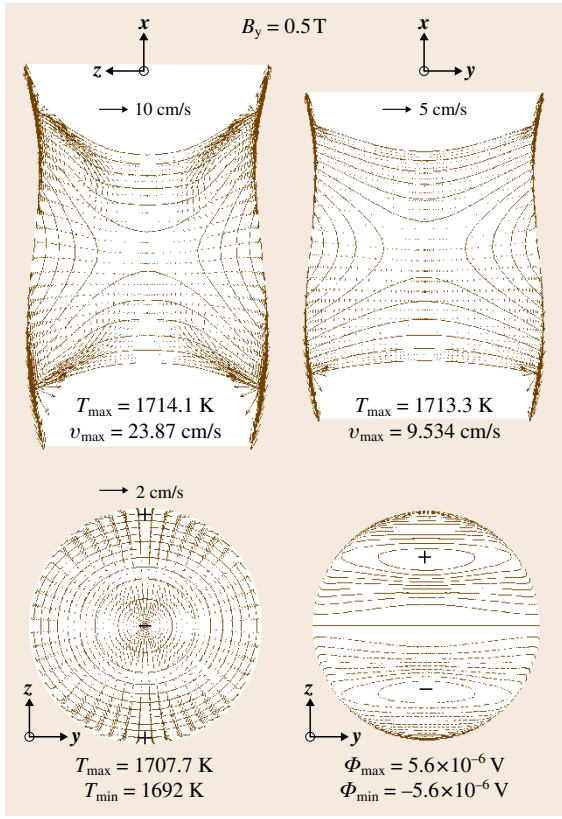


Fig. 36.18 Three-dimensional simulated flow, thermal, and potential fields in a FZ silicon growth under a transversal magnetic field of 0.5 T

to be identical due to symmetry. However, if we examine the result at $x = 4.1$ cm (the middle of the molten zone) on the y–z plane, the temperature is lower at the four corners. Due to such nonuniform thermal gradients, there are eight flow vortices caused by the thermocapillary force. As a result, the induced electrical potential distribution on the same plane also has an eight-cell structure.

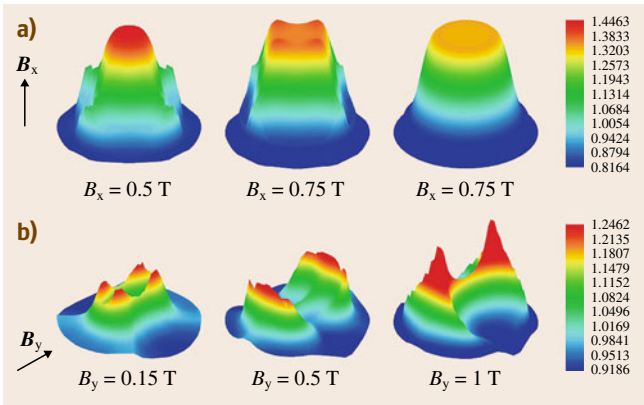


Fig. 36.19a,b Effect of axial (a) and transversal (b) magnetic fields on the dopant distribution in the growth crystal

When the magnetic field strength is less than 0.46 T, the symmetry disappears and the result becomes 3-D and time dependent. In the experiments by [36.18], a magnetic field strength greater than 0.24 T was found to be necessary to obtain nearly striation-free crystals. In fact, even at 0.5 T, irregular striation patterns were visible near the crystal surface. Therefore, our results are still in reasonable consistency with the observations.

With a transversal field, the steady-state result is obtained at a much lower magnetic field strength of 0.15 T. The calculated result for 0.5 T is shown in Fig. 36.18; the velocity scales at different planes are different for clarity. As shown, the results are asymmetric, but have a twofold symmetry when viewed from the y–z plane.

The surface zone length on the plane (x–z plane) perpendicular to the magnetic direction is much longer than that on the parallel plane. If we examine the flow fields, clearly, as mentioned previously, the melt flow on the plane perpendicular to the magnetic field is not as suppressed as that on the parallel plane, leading to a longer zone length there. If we view the results at the y–z plane (at $x = 4.1$ cm), the isotherms have an ellipsoid shape, while the maximum temperature appears at the surface of the x–z plane.

The dopant distribution is also significantly affected by the flow. Figure 36.19 summarizes the dopant distributions obtained from previous results with more results added. As shown, under axial fields, the segregation increases with increasing magnetic field strength due to flow damping. At 0.5 and 0.75 T of the axial field, the fourfold distribution is caused by flow such as that shown in Fig. 36.17. Under transversal fields, two concentration peaks align parallel to the field direction at 0.5 and 1 T. This is simply due to the poorer mixing

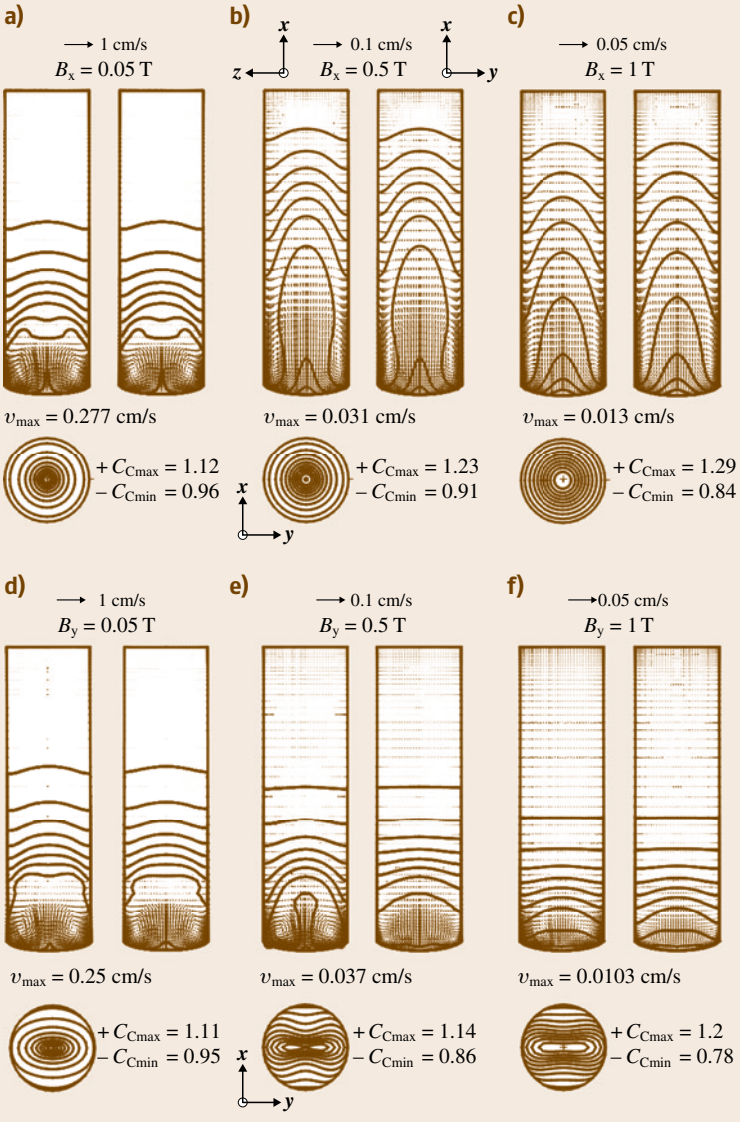


Fig. 36.20a–f Effect of axial and transversal magnetic fields on the flow and dopant fields

in the x–y plane, where the flow is suppressed more. On the other plane, the convection remains strong, so that the segregation is less. If we compare the segregation under both magnetic configurations, it is clear that the radial segregation is much less in the transversal field, even though it is highly asymmetric. Besides, as mentioned earlier, the minimum magnetic field strength to suppress the unsteady thermocapillary flow is also lower for the transversal field. Therefore, in real applications, it is believed that the transversal field may still be useful. If rotation is applied, the asymmetry may be reduced. The effect of rotation for the VB configuration has been investigated by *Lan* and *Yen* [36.60]. The effect of centrifugal pumping has also been discussed, where the conductivity of the crystal is important.

Flow suppression by static magnetic fields in the Bridgman configuration is even more effective, as could be clearly illustrated through numerical simulation. Again the gallium-doped germanium growth in a graphite ampoule discussed previously is used for illustration, with $\mathrm{Ra_T} = 2.489 \times 10^8$. Figure 36.20 shows some flow structures and dopant fields for several axial and horizontal fields. For the case of axial magnetic fields (Fig. 36.20a–c), two lower cells induced by radial thermal gradients are stretched in the axial direction by the axial magnetic field. Because the flow tending to across the magnetic lines will be suppressed, the flow cells are elongated by the axial magnetic field. Interestingly, as the cell is stretched axially, the solute penetrates more into the bulk melt. As will be shown shortly, for the case of $B_x = 0.5\,\mathrm{T}$, the bulk mixing is much enhanced. With the same magnetic strength, the horizontal field (Fig. 36.20d–f) is slightly more effective in suppressing the flow in terms of the maximum melt velocity. This is because the axial melt motion on the x–y plane is greatly suppressed; the axial melt motion is induced by radial thermal gradients. Although the flow on the plane (x–z plane) perpendicular to the magnetic direction is not suppressed effectively, the overall flow penetration into the upper bulk melt is significantly reduced. One can compare the dopant distributions in the x–z plane and the x–y plane to get a better idea of this. In addition, it is also clear that the solute mixing decreases monotonically with increasing magnetic field. Nevertheless, the flow and solute fields become asymmetric. One can further examine the solute concentration in the crystal ($C_\mathrm{c} = KC$) at the interface; the solute field is stretched in the applied magnetic direction.

To illustrate the bulk solute mixing and the radial segregation, we have also calculated the effective segregation coefficient K_eff from the pseudo-steady-state results for various magnetic fields, $K_\mathrm{eff} = \langle C \rangle / C_0$ [36.82, 83], where $\langle C \rangle$ is the average solute concentration in the melt. Figure 36.21a shows the effect of field strength (in terms of Ha number) on K_eff. As shown, the diffusion growth can be reached more quickly by applying a transversal field. However, for the axial field, there is a decrease in K_eff, i. e., better global dopant mixing, at

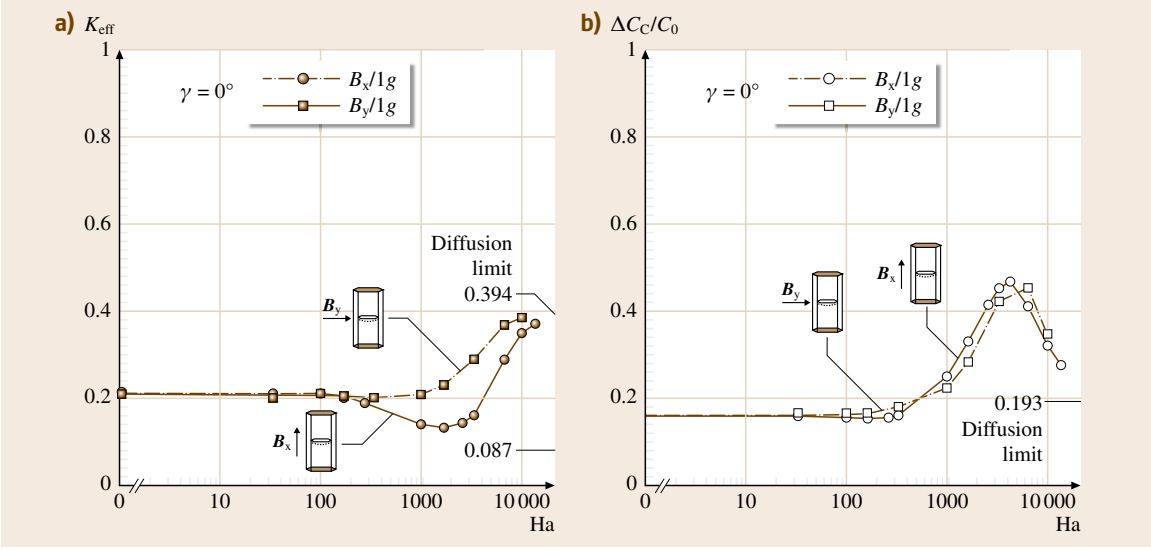


Fig. 36.21a,b Effect of axial and transversal magnetic fields: **(a)** effective segregation coefficient K_eff; **(b)** radial dopant segregation

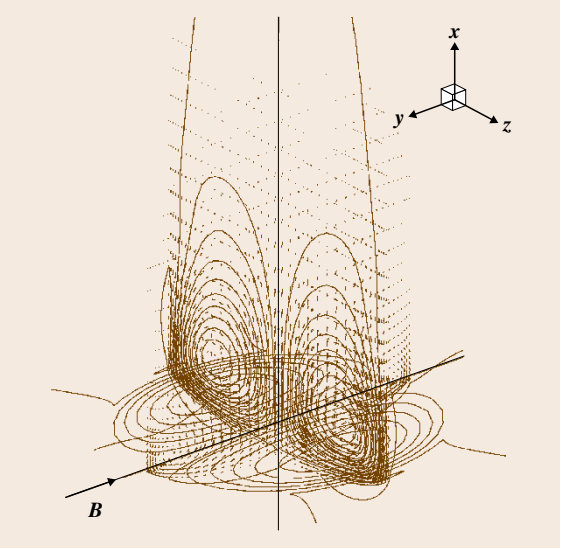


Fig. 36.22 Induced electrical potential and flow fields under a transversal magnetic field (1 T)

Ha ≈ 2000. Again, as discussed regarding Fig. 36.20b, the flow cells are stretched axially by the axial field and penetrate deeply into the bulk melt, leading to better bulk dopant mixing. The amount of radial segregation by both types of fields in Fig. 36.21b remains similar; $\Delta C = C_{c\,max} - C_{c\,min}$ is the maximum dopant concentration difference at the interface.

Apparently, even though the magnetic field is quite effective in terms of flow damping, it is clear that to suppress the flow completely is also difficult. It should be noticed that the damping effect is effective only for the flow in the plane parallel to the magnetic field. The flow damping in the plane perpendicular to the magnetic field is much less effective. This could be explained by a buildup of the electric potential that reduces the induced electric current for the Lorentz force. Fig. 36.22 illustrates the flow and the induced electric potential fields due to the transversal field at 0.5 T. As shown, the potential gradients build up due to the thermal toroidal cells and the transversal fields. The potential contour surfaces look like a pair of kidneys (with one positive and one negative potential value), with a symmetry plane at $z = 0$. In other words, due to the potential gradients, the net currents, $\boldsymbol{j} = \sigma(-\nabla\Phi + v \times \boldsymbol{e}_B)$, for the Lorentz force ($\boldsymbol{j} \times \boldsymbol{e}_B$) become smaller. Accordingly, the flow suppression becomes less effective. It can also be seen from the previous case for the FZ configuration that the flow perpendicular to the magnetic field cannot be suppressed effectively.

36.3.2 Rotation

Beside a static magnetic field, the use of rotation has been popular for crystal growth. There are a few ways of using rotation. The use of a centrifuge [36.51, 84], i. e., so-called centrifugal or high-gravity processing,

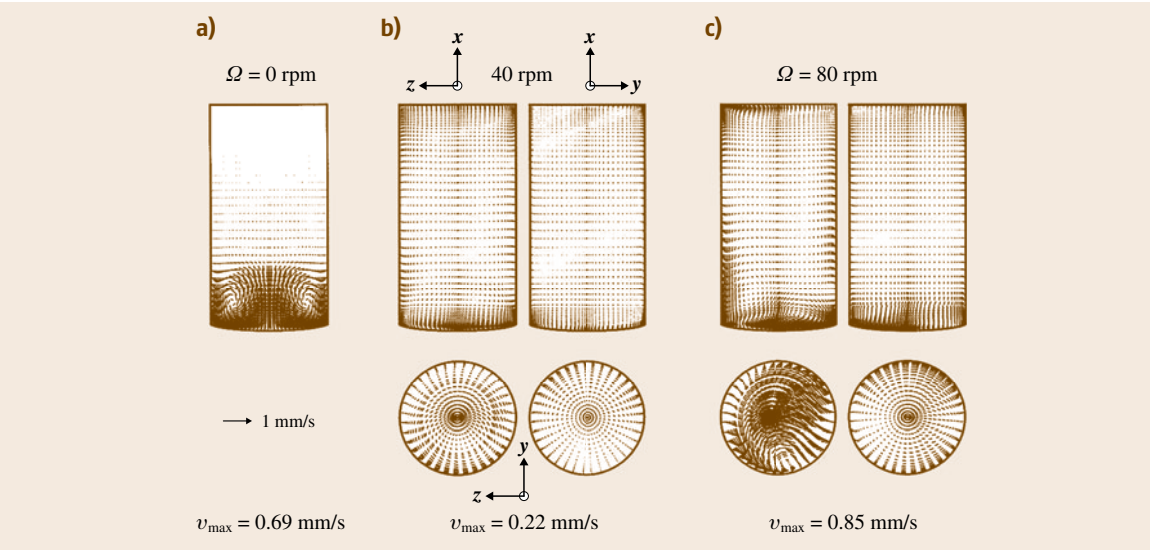


Fig. 36.23a–c Effect of free-swing rotation on the flow fields for a gradient freeze growth of GaAs: **(a)** 0 rpm; **(b)** 40 rpm; **(c)** 80 rpm

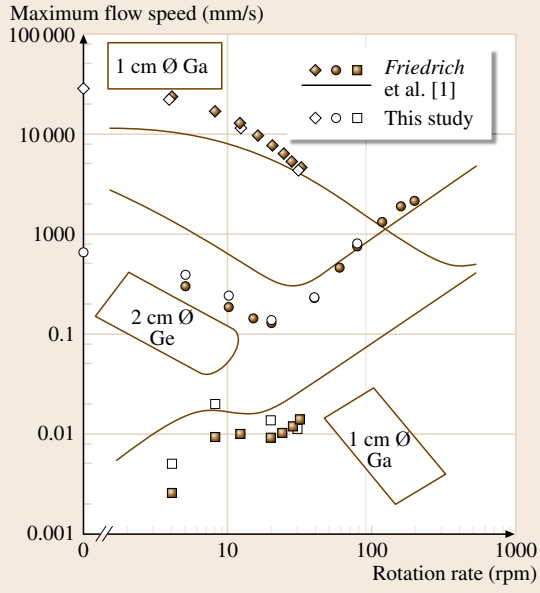


Fig. 36.24 Effect of rotation rate on the maximum melt velocity for various growth orientations

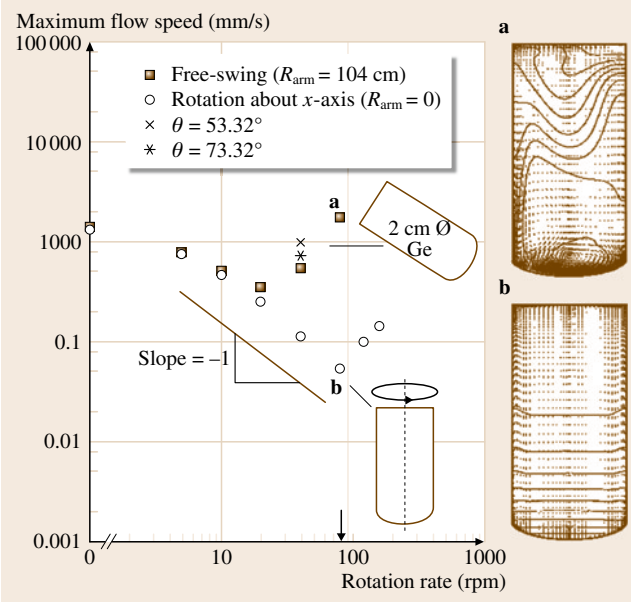


Fig. 36.25a,b Effect of rotation rate on the maximum melt velocity for rotation about the growth axis and the free-swing configuration. The flow and dopant fields for the free-swing **(a)** and rotation about growth axis **(b)** configurations are shown on the *right*

is particularly interesting. ACRT has also been widely used [36.34, 36–38]. The way of using the centrifuge in crystal growth has always been in the free-swinging configuration, in which the sample is placed at the end of a rotating arm. In such a case, the resultant acceleration is antiparallel to the axial thermal gradient, which is thus intuitively thermally stable. Accordingly, the convection can be suppressed at a certain rotation rate, or the so-called magic-g level [36.84], where the Coriolis force balances the gravitational acceleration. Beyond the magic-g level, the centrifugal acceleration becomes important and the centrifugal thermal convection increases. Figure 36.23 shows simulated flow patterns of Bridgman growth of germanium in a free-swing centrifuge at different rotation speed [36.51, 56]. For 0 rpm (Fig. 36.23a) the flow is axisymmetric and its structure is typical for the VB configuration, with a concave interface; the heating temperature profile is linear. At 40 rpm the flow near the growth interface is significantly suppressed by the Coriolis force and the flow structure is also changed dramatically. Although the averaged resultant gravity direction is still antiparallel to the growth axis, the centrifugal acceleration and the Coriolis force in the melt are asymmetric, leading to the 3-D flow. Due to the nonuniform forces, the global convection increases slightly away from the interface. As the rotation speed is further increased to 80 rpm, the centrifugal force becomes dominant and the convection increases, as can be seen from the larger velocity vectors. The flows in the x–y plane are also shown, but are in general featureless except for the flow near the growth interface.

We also present two flow patterns in the y–z plane, shown at the bottom of the figures for 40 and 80 rpm, respectively. One is at $x = 4.2$ cm and the other at 4.9 cm. The edge of the interface is at about $x = 4.0$ cm. As shown, near the interface the flow is mostly counterclockwise, but at some places the flow may be in the opposite direction. Interestingly, at 80 rpm, the flow pattern at $x = 4.9$ cm shows two cells with different flow directions. Therefore, the flow seen from the top does not have a well-defined structure. Closer to the growth interface, the counterclockwise flow seems to be clearer. Nevertheless, as the interface becomes flat, the flow pattern is also changed.

The maximum melt velocity as a function of rotation speed is further illustrated in Fig. 36.24 (open symbols), where the results of *Friedrich* et al. [36.51] (filled symbols and solid lines) are also included for comparison. We also performed calculations for gallium

melt (in both the free-swing and horizontal configurations). The comparison with the previous study is also shown in Fig. 36.24. As shown, they are all in good agreement. The solid lines in Fig. 36.24 are from the scaling analysis of *Friedrich* et al. [36.51]. As shown for the free-swing case of germanium, there is a minimum of convection at about 20 rpm. This is supposed to be the so-called magic-g level, having the least axial dopant segregation. At this critical rotation rate, the Coriolis force balances the two gravitational forces. Beyond this value, the centrifugal force becomes dominant and enhances the convection.

Interestingly, in a recent numerical study by *Lan* and *Tu* [36.85], rotation about the growth axis could give a much better result. As shown in Fig. 36.25, the melt flow can be suppressed more effectively by this configuration, and its flow and dopant fields are still axisymmetric. On the contrary, the free-swing, or near, configuration at the magic-g level generates 3-D flows and severe dopant nonuniformity; a side view of the flow and dopant fields at 80 rpm is shown in Fig. 36.25. *Lan* [36.55] also performed a numerical simulation for a similar system and found that the flow direction (thermal convection) near the solidification front could be reversed at high speeds. As a result, an inversion of radial dopant distribution was found, clearly due to the centrifugal acceleration. In fact, in some of the earlier numerical studies of Bridgman crystal growth in a rotating ampoule [36.53, 54, 86], the centrifugal acceleration was ignored. As a result, the convection decreases monotonically with increasing rotation speed. This is correct only at low rotation rate (small Fr number). When the rotation speed is higher than the magic-g level, the convection is enhanced by centrifugal acceleration. Clearly, the centrifugal force is perpendicular to the axial gradient, and the buoyancy flow due to the centrifugal force is generated. This flow direction happens to be in the opposite direction to that near the interface due to the concave interface.

To validate the idea proposed by *Lan* and *Tu* [36.85], visualization experiments on a rotating table using SCN doped with a small amount of ethanol or acetone were performed by *Lan* et al. [36.57, 58]. Figure 36.26 shows the effect of rotation on the interface morphology after 3 h of growth (the growth rate was 2.5 μm/s with 0.07 wt % of acetone). Before the critical rotation speed was reached (Fig. 36.26a–c), the breakdown location from a planar to a cellular interface was a good indication of the acetone accumulation at the center of the interface; the initial stage of the growth also showed pit formation, as illustrated in Fig. 36.15a (60 min). Again,

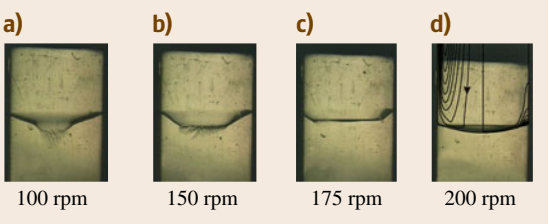


Fig. 36.26a–d Effect of rotation speed (about the growth axis) on interface morphology for the VB growth containing 0.007 wt % (after 3 h of crystal growth); the ampoule translation speed is 2.5 μm/s; **(a)** 100 rpm; **(b)** 150 rpm; **(c)** 175 rpm; **(d)** 200 rpm

the morphological breakdown of the interface was due to constitutional supercooling. With a high enough ro-

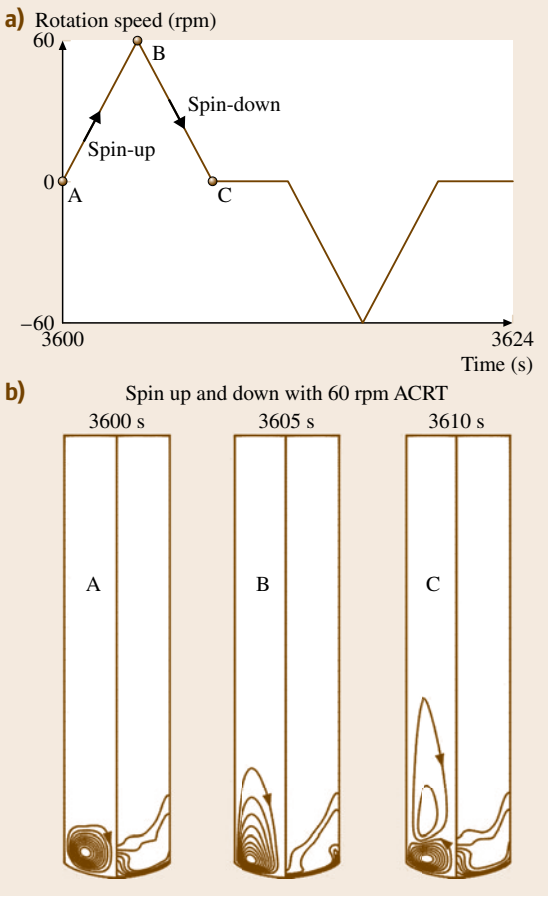


Fig. 36.27a,b Effect of ACRT on the flow and solute fields: **(a)** ACRT rotation cycle pattern; **(b)** instantaneous flow and solute fields at the moments indicated in **(a)**

tation speed, as shown in Fig. 36.26d at 200 rpm, the acetone was pushed outward to the edge of the interface, while the growth interface remained smooth. The calculated flow and solute fields near the growth interface are also illustrated in Fig. 36.26d. Therefore, by using an appropriate rotation rate, rotation about the growth axis is believed to be a useful approach for growth control, and this is not restricted to electrically conductive materials.

Beside steady rotation, ACRT is also a useful technique. Both the rotation cycle pattern and the period are critical to the flow control. Again, we take SCN as an example. The ACRT cycle pattern is shown in Fig. 36.27a, and some instantaneous flow patterns and acetone concentrations corresponding to the stages A, B, and C indicated in Fig. 36.27a are shown in Fig. 36.27b. The spin-up flows at 3600 and 3605 s near the interface are counterclockwise in direction. The spin-down flow at 3605 s is in the opposite flow direction. The acetone fields are significantly affected by the instantaneous flow as well. Similarly, the isotherms (not shown here) near the interface are found to be distorted toward the center of the interface. As a result,

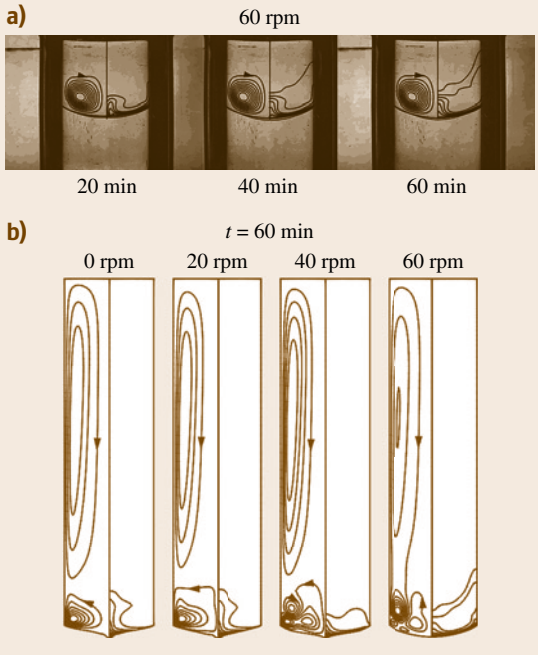


Fig. 36.28 **(a)** Flow and solute fields and the interface shapes at different growth periods for 60 rpm ACRT. **(b)** Average flow and solute fields and interface shapes for different maximum rotation magnitudes in ACRT

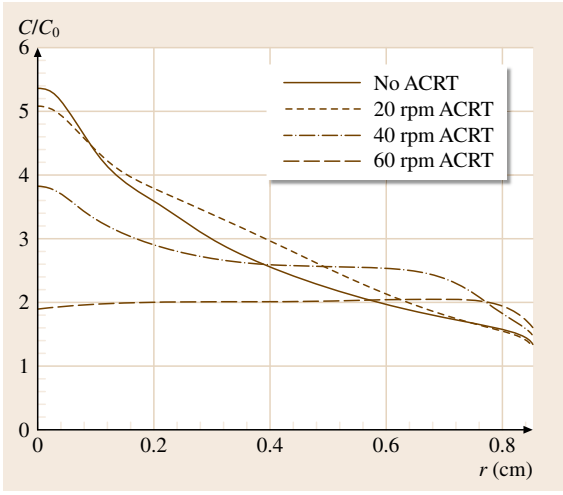


Fig. 36.29 Effect of the rotation amplitude in ACRT on the averaged radial acetone distribution

the interface becomes significantly concave at 60 rpm. The comparison of the calculated and observed interface shapes at different growth period for 60 rpm ACRT is shown in Fig. 36.28a. As shown, they are in good agreement. In addition, the interface when stationary is flat and it takes about 20 min to reach a steady shape.

The averaged flow and acetone fields for different ACRT amplitudes at 60 min after solidification are shown in Fig. 36.28b, from which the acetone concentration profiles at the interface are plotted in Fig. 36.29. As shown in Fig. 36.28b, it is clear that the original upper cell is not affected much by ACRT, except at 60 rpm. In other words, the solute mixing is confined to the region near the interface. However, at 60 rpm, the global mixing is slightly enhanced due to the connection of the lower and upper flow cells. Therefore, the maximum acetone concentration is significantly lower for 60 rpm. On the other hand, the lower cell is significantly affected by ACRT and this significantly affects the radial acetone segregation. As shown in Fig. 36.28, in general, as the rotation amplitude increases, except for 20 rpm, the acetone accumulation at the interface center decreases and the radial uniformity increases for the acetone concentration at the interface. From Fig. 36.29 it is thus clear that ACRT improves radial acetone uniformity and lowers its concentration due to improved mixing by ACRT. Particularly, at 60 rpm, the much slower acetone accumulation, as a result of better global mixing, reduces supercooling and enhances morphological stability.

36.3.3 Vibration

The ACRT mentioned previously usually requires a rotation cycle having a period long enough to develop Ekman flows. The Ekman time scale can be estimated by $R_c/\sqrt{\Omega\nu}$ [36.87], where R_c is the crystal radius, Ω is the rotation speed, and ν is the kinematic viscosity of the melt. For growth of SCN in small to medium-sized vertical Bridgman systems, the Ekman time scale is up to a few seconds. An alternative approach to applying ACRT is to use a cycle time that is much shorter than the Ekman time. This method is known as the angular vibration technique (AVT) [36.45]. In this technique the ampoule is vibrated at a frequency greater than 1 Hz in the rotational direction to generate a radial outward Schlichting flow near the growth front [36.45, 47].

Figure 36.30a,b shows some simulated results for AVT with different frequencies for SCN/acetone and SCN/salol, respectively. The observed interface shapes after 1 h growth are put together for comparison. As shown, the simulated interface concavity for both cases agrees quite well with the experiments. It should be noticed that the interface is at the upper boundary of the breakdown area. More importantly, from 0 to 5 Hz, the interface concavity decreases with frequency, while from 5 to 20 Hz the concavity increases with frequency. The reason is quite clear from the simulation. From 0 to 5 Hz, the flow above the interface is weakened by vibration because of the radial outward streaming flow induced by the angular vibration. As a result, the solute distribution becomes more uniform and this reduces the interface concavity cased by the local solute accumulation. On the other hand, from 5 to 20 Hz, the Schlichting flow becomes dominant. Since the flow is in the clockwise direction and the isotherms are distorted with the flow, the interface concavity increases with the vibration intensity (frequency).

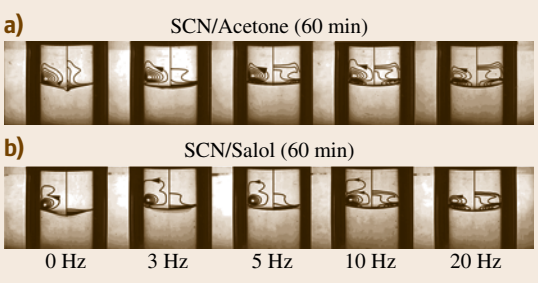


Fig. 36.30a,b Effect of angular vibration frequency on the flow and solute fields and the interface shape for VB growth of SCN: **(a)** containing 0.064 wt % acetone; **(b)** 0.15 wt % salol

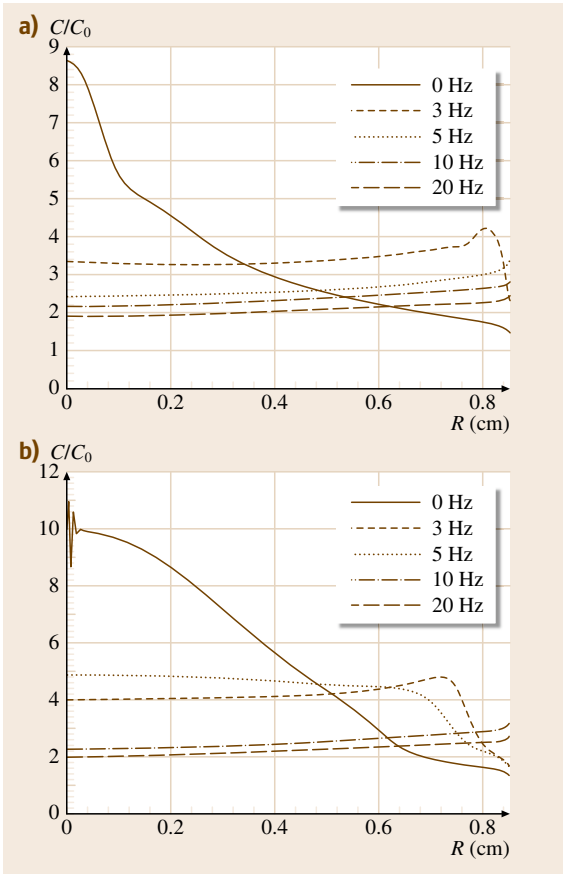


Fig. 36.31a,b Effect of angular vibration on the radial solute distributions for VB growth of SCN: **(a)** containing 0.064 wt % acetone; **(b)** 0.15 wt % salol

The radial solute concentrations extracted from the simulated results in Fig. 36.30 are plotted in Fig. 36.31 for SCN/acetone and SCN/salol, respectively. As shown, for both cases, the radial segregation reverses from 0 to 3 Hz. This indicates that the Schlichting streaming flow is strong enough to overcome the buoyancy force and is able to push the solute from the interface center to the rim. From both Figs. 36.30 and 36.31, it is clear that, when the frequency is greater than 10 Hz, the Schlichting flow dominates and the solutal effect becomes insignificant, which can be seen from the solutal fields as well as the radial solute segregation profiles.

Besides angular vibration, vibration can also be applied axially. However, as discussed by *Lan* [36.68], the

thermovibration requires radial gradients, and the effect is not as significant as that by angular vibration. Further discussion on flow control by vibration for crystal growth has been given by *Lyubimov* et al. [36.40].

36.4 Outlook

So far, we have discussed melt convection and its control for ZM and Bridgman bulk crystal growth through a few examples. As illustrated, the interface shape, which is a key factor for crystal quality, and the composition uniformity are significantly affected by the convective heat and mass transports due to the flow. We have introduced the basic flow structures through flow visualization experiments and numerical simulations. From these, one can better understand the interplay of the transport processes and the interface in crystal growth processes. The melt flow is affected by body forces, which can be the buoyancy (gravitational and centrifugal), Coriolis, and Lorentz forces. With a free surface, the thermocapillary force can also drive the melt flow. Thermal and solutal gradients are the sources for the buoyancy convection. Therefore, if these gradients are antiparallel to the gravitational or centrifugal acceleration, the convection can be minimized, and this is typically the case for vertical Bridgman growth in normal gravity or growth in a centrifuge under a free-swing configuration. In such a configuration, the flow is much more stable. Nevertheless, the residual flow can induce significant composition nonuniformity. The use of microgravity, magnetic fields, rotation or vibration is useful in manipulating the flow and thus improving crystal uniformity. Furthermore, suppressing the flow also helps improve axial composition uniformity, and a static magnetic field is particularly effective if the melt is electrically conductive.

Although we have not been able to discuss the flow and its control for Czochralski crystal growth, the concepts learned from previous examples are still useful. For example, similar to that in the FZ growth, in Czochralski growth of oxide crystals, the interface inversion can be easily controlled by crystal rotation [36.88]. However, the convection in the CZ configuration is much more complicated and hard to elucidate. For example, the buoyancy instability leading to spoke patterns at the melt surface [36.6, 89] and its coupling with rotation leading to baroclinic instability [36.90–92] are also typical in CZ oxide growth. The baroclinic instability has also been reported for CZ silicon growth [36.93]. Moreover, in large-scale CZ silicon growth, the flow is often turbulent. To suppress the turbulent flow, magnetic fields have been widely used in practice [36.27].

In contrast to flow damping, generating a favorable flow is also useful in bulk crystal growth. This includes the use of rotating and traveling magnetic fields [36.35, 94], as well as dynamic and alternating magnetic fields [36.4]. Recently, *Watanabe* et al. [36.30] proposed the electromagnetic Czochralski (EMCZ) method, in which the Lorentz force was used to generate a controlled melt flow. To generate this controlled melt flow, they created a Lorentz force by combining static magnetic fields with an electric current passing through the melt from a growing crystal. The controlled melt flow in the EMCZ method makes it easy to control the temperature distribution around the crystal, so that the growth interface can be better controlled [36.95]. Since the rotating flow washing over the crucible's inner surface can be adjusted, the oxygen content in the grown crystal can be controlled as well [36.32].

Over the years, extensive research has been applied to the investigation of melt flows, and many techniques have been proposed for their control. Still, with the increasing demands for crystal quality for advanced electronic and optoelectronic applications, this research will continue. Especially, their connection to as-grown defects and crystal quality still requires much more research to obtain deeper understanding. On the other hand, computer modeling has become an indispensable tool in the analysis and design of bulk crystal growth systems. Convection structures and their control in melt growth have been well stimulated and better understood [36.60, 96].

References

36.1 K. Kakimoto, M. Watanabe, M. Eguchi, T. Hibiya: Ordered structure in non-axisymmetric flow of silicon melt convection, J. Cryst. Growth **126**, 435–440 (1993)

36.2 K. Kakimoto, H. Ozoe: Oxygen distribution at a solid–liquid interface of silicon under transverse magnetic fields, J. Cryst. Growth **212**, 429–437 (2000)

36.3 I.Y. Evstratov, V.V. Kalaev, A.I. Zhmakin, Y.N. Makarov, A.G. Abramov, N.G. Ivanov, A.B. Korsakov, E.M. Smirnov, E. Dornberger, J. Virbulis, E. Tomzig, W. von Ammon: Numerical study of 3D unsteady melt convection during industrial-scale CZ Si-crystal growth, J. Cryst. Growth **237–239**, 1757–1761 (2002)

36.4 E. Tomzig, J. Virbulis, W. von Ammon, Y. Gelfgat, L. Gorbunov: Application of dynamic and combined magnetic fields in the 300mm silicon single-crystal growth, Mater. Sci. Semicond. Process. **5**, 347–351 (2003)

36.5 D.T.J. Hurle: *Crystal Pulling from the Melt* (Springer, Berlin Heidelberg 1993)

36.6 A.D.W. Jones: An experimental model of the flow in Czochralski growth, J. Cryst. Growth **61**, 235–244 (1983)

36.7 D.P. Lukanin, V.V. Kalaev, Y.N. Makarov, T. Wetzel, J. Virbulis, W. von Ammon: Advances in the simulation of heat transfer and prediction of the melt-crystal interface shape in silicon CZ growth, J. Cryst. Growth **266**, 20–27 (2004)

36.8 R.A. Brown: Theory of transport processes in single crystal growth from the melt, AIChE J. **34**, 881–911 (1989)

36.9 F. Dupret, N. van den Bogaert: Modeling Bridgman and Czochralski growth. In: *Handbook of Crystal Growth: Growth Mechanisms and Dynamics*, Vol. 2b, ed. by D.T.J. Hurle (North-Holland, Amsterdam 1994)

36.10 G. Müller, A.G. Ostrogorsky: Convection in melt growth. In: *Handbook of Crystal Growth Growth Mechanisms and Dynamics*, Vol. 2b, ed. by D.T.J. Hurle (North-Holland, Amsterdam 1994)

36.11 Y.F. Zou, G.-X. Wang, H. Zhang, V. Prasad: Mechanisms of thermo-solutal transport and segregation in high-pressure liquid-encapsulated Czochralski crystal growth, J. Heat Transf. **121**, 148–159 (1999)

36.12 C.W. Lan: Recent progress of crystal growth modeling and growth control, Chem. Eng. Sci. **59**, 1437–1457 (2004)

36.13 M.C. Liang, C.W. Lan: Three-dimensional thermocapillary and buoyancy convections and interface shape in horizontal Bridgman crystal growth, J. Cryst. Growth **180**, 587–596 (1997)

36.14 C.W. Lan, M.C. Su, M.C. Liang: A visualization and computational study of horizontal Bridgman crystal growth, J. Cryst. Growth **208**, 717–725 (1999)

36.15 C.W. Lan, J.H. Chian, T.Y. Wang: Interface control mechanisms in horizontal zone-melting with slow rotation, J. Cryst. Growth **218**, 115–124 (2000)

36.16 C.W. Lan: Heat Transfer, Fluid Flow, and Interface Shapes in Floating-Zone Crystal Growth. Ph.D. Thesis (University of Wisconsin, Madison 1991)

36.17 M. Levenstam, G. Amberg: Hydrodynamical instabilities of thermocapillary flow in a half-zone, J. Fluid Mech. **297**, 357–372 (1995)

36.18 P. Dold, A. Cröll, K.W. Benz: Floating-zone growth of silicon in magnetic fields. I. Weak static axial fields, J. Cryst. Growth **183**, 545–553 (1998)

36.19 S. Nakamura, T. Hibiya, K. Kakimoto, N. Imaishi, S. Nishizawa, A. Hirata, K. Mukai, S. Yoda, T.S. Morita: Temperature fluctuations of the Marangoni flow in a liquid bridge of molten silicon under microgravity on board the TR-IA-4 rocket, J. Cryst. Growth **186**, 85–94 (1998)

36.20 M. Prange, M. Wanschura, H.C. Kuhlmann, H.J. Rath: Linear stability of thermocapillary convection in cylindrical liquid bridges under axial magnetic fields, J. Fluid Mech. **394**, 281–302 (1999)

36.21 M. Schweizer, A. Croll, P. Dold, T. Kaiser, M. Lichtensteiger, K.W. Benz: Measurement of temperature fluctuations and microscopic growth rates in a silicon floating zone under microgravity, J. Cryst. Growth **203**, 500–510 (1999)

36.22 A. Mühlbauer, A. Muiznieks, J. Virbulis, A. Lüdge, H. Riemann: Interface shape, heat transfer and fluid flow in the floating zone growth of large silicon crystals with the needle-eye technique, J. Cryst. Growth **151**, 66–79 (1995)

36.23 G. Ratnieks, A. Muiznieks, A. Mühlbauer: Modelling of phase boundaries for large industrial FZ silicon crystal growth with the needle-eye technique, J. Cryst. Growth **255**, 227–240 (2003)

36.24 H.F. Utech, M.C. Flemming: Elimination of solute banding in indium antimonide crystals by growth in a magnetic Field, J. Appl. Phys. **37**, 2021–2023 (1966)

36.25 K.M. Kim, P. Smetana: Oxygen segregation in CZ silicon crystal-growth on applying a high axial magnetic-field, J. Electrochem. Soc. **133**, 1682–1686 (1986)

36.26 K.M. Kim: Suppression of thermal convection by transverse magnetic field, J. Electrochem. Soc. **129**, 427–429 (1982)

36.27 D.T.J. Hurle, R.W. Series: Use of a magnetic field in melt growth. In: *Handbook of Crystal Growth*, Vol. 2a, ed. by D.T.J. Hurle (North-Holland, Amsterdam 1994)

36.28 F.M. Herrmann, G. Müller: Growth of 20 mm diameter GaAs crystals by the floating-zone technique with controlled As-vapour pressure under microgravity, J. Cryst. Growth **156**, 350–360 (1995)

36.29 P. Dold, K.W. Benz: Rotating magnetic fields: fluid flow and crystal growth applications, Prog. Cryst. Growth Charact. Mater. **38**, 7–38 (1999)

36.30 M. Watanabe, M. Eguchi, T. Hibiya: Silicon crystal growth by the electromagnetic Czochralski (EMCZ) method, Jpn. J. Appl. Phys. **38**, L10–L13 (1999)

36.31 M. Watanabe, K.W. Yi, T. Hibiya, K. Kakimoto: Direct observation and numerical simulation of molten silicon flow during crystal growth under magnetic fields by x-ray radiography and large-scale computation, Progr. Crystal Growth Charact. Mater. **38**, 215–238 (1999)

36.32 M. Watanabe, M. Eguchi, W. Wang, T. Hibiya, S. Kuragaki: Controlling oxygen concentration and distribution in 200 mm diameter Si crystals using the electromagnetic Czochralski (EMCZ) method, J. Cryst. Growth **237–239**, 1657–1662 (2002)

36.33 A. Mitric, T. Duffar, C. Diaz-Guerra, V. Corregidor, L.C. Alves, C. Garnier, G. Vian: Growth of $Ga_{1-x}In_xSb$ alloys by vertical Bridgman technique under alternating magnetic field, J. Cryst. Growth **287**, 224–229 (2006)

36.34 H.J. Scheel: Accelerated crucible rotation: a novel stirring technique in high-temperature solution growth, J. Cryst. Growth **13/14**, 560–565 (1971)

36.35 P. Capper, J.J. Gosney: Method of growing crystalline cadmium mercury telluride grown by method, UK Patent 2098879 (1982)

36.36 P. Capper, J.J.G. Gosney, C.L. Jones: Application of the accelerated crucible rotation technique to the Bridgman growth of $Cd_xHg_{1-x}Te$: simulations and crystal growth, J. Cryst. Growth **70**, 356–364 (1984)

36.37 P. Capper, J.C. Brice, C.L. Jones, W.G. Coates, J.J.G. Gosney, C.K. Ard., I. Kenworthy: Interfaces and flow regimes in ACRT grown $Cd_xHg_{1-x}Te$ crystals, J. Cryst. Growth **89**, 171–176 (1988)

36.38 W.G. Coates, P. Capper, C.L. Jones, J.J.G. Gosney, C.K. Ard, I. Kenworthy, A. Clark: Effect of ACRT rotation parameters on Bridgman grown $Cd_xHg_{1-x}Te$ crystals, J. Cryst. Growth **94**, 959–966 (1989)

36.39 A.V. Anilkumar, R.N. Grugel, R.N. Shen, T.G. Wang: Control of thermocapillary convection in a liquid bridge by vibration, J. Appl. Phys. **73**, 4165–4170 (1993)

36.40 D.V. Lyubimov, T.P. Lyubimova, S. Meradji, B. Roux: Vibrational control of crystal growth from liquid phase, J. Cryst. Growth **180**, 648–659 (1997)

36.41 W.S. Liu, M.F. Wolf, D. Elwell, R.S. Feigelson: Low frequency vibrational stirring: a new method for radial mixing solutions and melts during growth, J. Cryst. Growth **82**, 589–597 (1987)

36.42 W. Yuan, M. Banan, L.L. Regel, W.R. Wilcox: The effect of vertical vibration of the ampoule on the direction solidification of InSb-GaSb alloy, J. Cryst. Growth **151**, 235–242 (1995)

36.43 V. Uspenski, J.J. Favier: High frequency vibration and natural convection in Bridgman-scheme crystal growth, Int. J. Heat Mass Transf. **37**, 691–698 (1994)

36.44 C.W. Lan: Effect of axisymmetric magnetic fields on heat flow and interface in floating-zone silicon crystal growth, Model. Simul. Mater. Sci. Eng. **6**, 423–445 (1998)

36.45 W.C. Yu, Z.B. Chen, W.T. Hsu, B. Roux, T.P. Lyubimova, C.W. Lan: Reversing radial segregation and suppressing morphological instability during Bridgman crystal growth by angular vibration, J. Cryst. Growth **271**, 474–480 (2004)

36.46 W.C. Yu, Z.B. Chen, W.T. Hsu: Effects of angular vibration on the flow, segregation, and interface morphology in vertical Bridgman crystal growth, Int. J. Heat Mass Transf. **50**, 58–66 (2007)

36.47 Y.C. Liu, W.C. Yu, B. Roux, T.P. Lyubimova, C.W. Lan: Thermal-solutal flows and segregation and their control by angular vibration in vertical Bridgman crystal growth, Chem. Eng. Sci. **61**, 7766–7773 (2006)

36.48 W.A. Arnold, W.R. Wilcox, F. Carlson, A. Chait, L.L. Regel: Transport modes during crystal growth in a centrifuge, J. Cryst. Growth **119**, 24–40 (1992)

36.49 G. Müller, G. Neumann, W. Weber: The growth of homogeneous semiconductor crystals in a centrifuge by the stabilizing influence of the Coriolis force, J. Cryst. Growth **119**, 8–23 (1992)

36.50 W.R. Wilcox, L.L. Regel: Influence of centrifugation on transport phenomena, 46th Int. Astronaut. Congr. (Oslo 1995)

36.51 J. Friedrich, J. Baumgartl, H.J. Leister, G. Müller: Experimental and theoretical analysis of convection and segregation in vertical Bridgman growth under high gravity on a centrifuge, J. Cryst. Growth **167**, 45–55 (1996)

36.52 W.R. Wilcox, L.L. Regel, W.A. Arnold: Convection and segregation during vertical Bridgman growth with centrifugation, J. Cryst. Growth **187**, 543–558 (1998)

36.53 C.W. Lan: Effects of ampoule rotation on flows and segregation in vertical Bridgman crystal growth, J. Cryst. Growth **197**, 983–991 (1999)

36.54 A. Yeckel, F.P. Doty, J.J. Derby: Effect of steady ampoule rotation on segregation in high-pressure vertical Bridgman growth of cadmium zinc telluride, J. Cryst. Growth **203**, 87–102 (1999)

36.55 C.W. Lan: Effects of centrifugal acceleration on flows and segregation in vertical Bridgman crystal growth, J. Cryst. Growth **229**, 595–600 (2001)

36.56 C.W. Lan, C.H. Chian: Three-dimensional simulation of Marangoni convection in floating-zone crystal growth, J. Cryst. Growth **230**, 172–180 (2001)

36.57 C.W. Lan, Y.W. Yang, C.Y. Tu: Reversing radial segregation and suppression morphological instability in directional solidification by rotation, J. Cryst. Growth **235**, 619–625 (2002)

36.58 C.W. Lan, Y.W. Yang, H.Z. Chen, I.F. Lee: Segregation and morphological instability due to double diffusive convection in rotational directional solidification, Metal. Mater. Trans. A **33**, 3011–3017 (2002)

36.59 A.F. Witt, H.C. Gatos, M. Lichtensteiger, M.C. Lavine, C.J. Herman: Crystal growth and steady state segregation under zero gravity, J. Electrochem. Soc. **122**, 276–283 (1975)

36.60 C.W. Lan, B.C. Yeh: Three-dimensional analysis of flow and segregation in vertical Bridgman crystal growth under a transversal magnetic field with ampoule rotation, J. Cryst. Growth **266**, 200–206 (2004)

36.61 M.C. Lan, M.C. Liang: A three-dimensional finite-volume/Newton method for thermal-capillary

problems, Int. J. Numer. Methods Eng. **40**, 621–636 (1997)
36.62 C.W. Lan, M.C. Liang: Multigrid methods for incompressible heat flow problems with an unknown interface, J. Comput. Phys. **152**, 55–77 (1999)
36.63 A. Lipchin, R.A. Brown: Hybrid finite-volume/finite-element simulation of heat transfer and melt turbulence in Czochralski crystal growth of silicon, J. Cryst. Growth **216**, 192–203 (2000)
36.64 C.W. Lan, D.T. Yang: A numerical study on heat flow and interface of the vertical zone-melting crystal growth, Numer. Heat Transf., Part A **129**, 131–145 (1996)
36.65 C.W. Lan: Newton's method for solving heat transfer, fluid flow and interface shapes in a floating molten zone, Int. J. Numer. Method Fluids **19**, 41–65 (1994)
36.66 C.W. Lan, S. Kou: Heat-transfer, fluid-flow and interface shapes in floating-zone crystal-growth, J. Cryst. Growth **108**, 351–366 (1991)
36.67 S. Kou, C.W. Lan: Contactless heater floating zone refining and crystal growth, US Patent 5217565 (1993)
36.68 C.W. Lan: Heat transfer, fluid flow, and interface shapes in zone melting processing with induction heating, J. Electrochem. Soc. **145**, 3926–3935 (1998)
36.69 C.W. Lan, M.C. Liang: Three-dimensional simulation of vertical zone-melting crystal growth: Symmetry breaking to multiple states, J. Cryst. Growth **208**, 327–340 (2000)
36.70 C.W. Lan, C.H. Wang: Three-dimensional bifurcations of a two-phase Rayleigh–Benard problem in a cylinder, Int. J. Heat Mass Transf. **44**, 1823–1838 (2001)
36.71 J. Baumgartl, W. Budweiser, G. Müller, G. Neumann: Studies of buoyancy driven convection in a vertical cylinder with parabolic temperature profile, J. Cryst. Growth **97**, 9–17 (1989)
36.72 W.G. Pfann: *Zone Melting* (Wiley, New York 1958)
36.73 C.W. Lan: Effects of axial vibration on vertical zone-melting processing, Int. J. Heat Mass Transf. **43**, 1987–1997 (2000)
36.74 W.W. Mullins, R.F. Sekerka: The stability of a planar interface during solidification of a dilute binary alloy, J. Appl. Phys. **35**, 444–451 (1964)
36.75 K.H. Lie, J.T. Hsu, Y.D. Guo, T.P. Chen: Influence of through-window radiation on the horizontal Bridgman process for rectangular shaped GaAs crystals, J. Cryst. Growth **109**, 205–211 (1991)
36.76 P.M. Adornato, R.A. Brown: Convection and segregation in directional solidification of dilute and non-dilute binary alloy: effects of ampoule and furnace design, J. Cryst. Growth **80**, 155–190 (1987)
36.77 D. Hofmann, T. Jung, G. Müller: Growth of 2 inch Ge:Ga crystals by the dynamic verical gradient freeze process and its numerical modeling including transient segregation, J. Cryst. Growth **128**, 213–218 (1992)
36.78 C.W. Lan, F.C. Chen: A finite-volume method for solute segregation in directional solidification and comparison with a finite-element method, Comput. Methods Appl. Mech. Eng. **31**, 191–207 (1996)
36.79 M.C. Liang, C.W. Lan: Three-dimensional convection and solute segregation in vertical Bridgman crystal growth, J. Cryst. Growth **167**, 320–332 (1996)
36.80 G.D. Robertson, D.J. O'Connor: Magnetic field effects on float-zone Si crystal growth: strong axial fields, J. Cryst. Growth **76**, 111–122 (1986)
36.81 C.W. Lan: Effect of axisymmetric magnetic fields on radial dopant segregation of floating-zone silicon growth in a mirror furnace, J. Cryst. Growth **169**, 269–278 (1996)
36.82 J.A. Burton, R.C. Prim, W.P. Slichter: The distribution of solute in crystals grown from the melt. Part I. Theoretical, J. Chem. Phys. **21**, 1987–1991 (1953)
36.83 D.H. Kim, P.M. Adornato, R.A. Brown: Effect of vertical magnetic field on convection and segregation in vertical Bridgman crystal growth, J. Cryst. Growth **89**, 339–356 (1988)
36.84 L.L. Rodot, M. Rodot, W.R. Wilcox: Material processing in high gravity-proceedings of the 1st international workshop on material processing in high gravity, J. Cryst. Growth **119**, R8 (1992)
36.85 C.W. Lan, C.Y. Tu: Three-dimensional analysis of heat flow, segregation, and interface shape of gradient-freeze growth in a centrifuge, J. Cryst. Growth **226**, 406–418 (2001)
36.86 M.R. Foster: The effect of rotation on vertical Bridgman growth at large Rayleigh number, J. Fluid Mech. **409**, 185–221 (2000)
36.87 A. Yeckel, J.J. Derby: Effect of accelerated crucible rotation on melt composition in high-pressure vertical Bridgman growth of cadmium zinc telluride, J. Cryst. Growth **209**, 734–750 (2000)
36.88 S.H. Lee, Y.J. Kim, S.H. Cho, E.P. Yoon: The influence of the Czochralski growth parameters on the growth of lithium niobate single crystals, J. Cryst. Growth **125**, 175–180 (1992)
36.89 Q. Xiao, J.J. Derby: Three-dimensional melt flows in Czochralski oxide growth: High-resolution, massively parallel, finite element computations, J. Cryst. Growth **152**, 169–181 (1995)
36.90 M.P. Gates, B. Cockayne: Purification of sodium tungstate, Nature **207**, 855 (1965)
36.91 C.D. Brandle: Flow transitions in Czochralski oxide melts, J. Cryst. Growth **57**, 65–70 (1982)
36.92 C.J. Jing, N. Imaishi, T. Sato, Y. Miyazawa: Three-dimensional numerical simulation of oxide melt flow in Czochralski configuration, J. Cryst. Growth **216**, 372–388 (2000)
36.93 M. Watanabe, M. Eguchi, K. Kakimoto: The baroclnic flow instability in rotating silicon melt, J. Cryst. Growth **128**, 288–292 (1993)
36.94 P. Dold, K.W. Benz: Modification of fluid flow and heat transport in vertical Bridgman configurations

by rotating magnetic fields, Cryst. Res. Technol. **32**, 51–60 (1997)

36.95 M. Watanabe, D. Vizman, J. Friedrich: Large modification of crystal–melt interface shape during Si crystal growth by using electromagnetic Czochralski method (EMCZ), J. Cryst. Growth **292**, 252–256 (2006)

36.96 G. Müller, J. Friedrich: Challenges in modeling of bulk crystal growth, J. Cryst. Growth **266**, 1–19 (2004)

37. Vapor Growth of III Nitrides

Dang Cai, Lili Zheng, Hui Zhang

Good understanding of transport phenomena in vapor deposition systems is critical to fast and effective crystal growth system design. Transport phenomena are complicated and are related to operating conditions, such as temperature, velocity, pressure, and species concentration, and geometrical conditions, such as reactor geometry and source–substrate distance. Due to the limited in situ experimental monitoring, design and optimization of growth is mainly performed through semi-empirical and trial-and-error methods. Such an approach is only able to achieve improvement in the deposition sequence and cannot fulfill the increasingly stringent specifications required in industry. Numerical simulation has become a powerful alternative, as it is fast and easy to obtain critical information for the design and optimization of the growth system. The key challenge in vapor deposition modeling lies in developing an accurate simulation model of gas-phase and surface reactions, since very limited kinetic information is available in the literature. In this chapter, GaN thin-film growth by iodine vapor-phase epitaxy (IVPE) is used as an example to present important steps for system design and optimization by the numerical modeling approach. The advanced deposition model will be presented for multicomponent fluid flow, homogeneous gas-phase reaction inside the reactor, heterogeneous surface reaction on the substrate surface, heat transfer, and species transport. Thermodynamic and kinetic analysis will be presented for gas-phase and surface reactions, together with a proposal for the reaction mechanism based on experiments. The prediction of deposition rates is presented. Finally, the surface evolution of film growth from vapor is analyzed for the case in which surface diffusion determines crystal grain size and morphology. Key control parameters for film instability are identified for quality improvement.

37.1 **Overview of Vapor Growth of III Nitrides** 1244
37.1.1 Various GaN/AlN Vapor-Growth Systems 1244
37.1.2 Modeling of AlN/GaN Vapor Deposition 1246

37.2 **Mathematical Models for AlN/GaN Vapor Deposition** 1248
37.2.1 Transport Equations 1248
37.2.2 Growth Kinetics 1249
37.2.3 Numerical Solution 1251

37.3 **Characteristics of AlN/GaN Vapor Deposition** 1251
37.3.1 Theoretical Analysis of Heat and Mass Transfer 1251
37.3.2 Thermodynamic and Kinetic Analysis of Chemical Reactions 1254

37.4 **Modeling of GaN IVPE Growth – A Case Study** 1258
37.4.1 Scaling Analysis 1258
37.4.2 Computational Issues 1258
37.4.3 Gas-Phase and Surface Reactions Analysis 1259
37.4.4 Geometrical and Operational Conditions Optimization 1264
37.4.5 Effect of Total Gas Flow Rate on Substrate Temperature 1264
37.4.6 Effect of Substrate Rotation on Deposition Rate and Deposition Uniformity 1269
37.4.7 Quasi-equilibrium Model for Deposition Rate Prediction 1270
37.4.8 Kinetic Deposition Model 1271

37.5 **Surface Evolution of GaN/AlN Film Growth from Vapor** 1274

37.6 **Concluding Remarks** 1275

References 1276

37.1 Overview of Vapor Growth of III Nitrides

Chemical vapor deposition (CVD) systems have been widely used to grow thin-film and bulk GaN/AlN crystals, which have a broad range of industrial applications, especially in the field of optoelectronics. For example, they have been used to manufacture optoelectronic devices such as light-emitting diodes (LEDs), laser diodes (LDs), and detectors [37.46, 47]. Due to their wide bandgap and high breakdown field, they have become important materials for high-temperature/high-power electronics [37.48–51].

In this section, GaN/AlN vapor-growth systems are briefly introduced first. An iodine vapor deposition system for GaN bulk growth is taken as an example to explain the modeling of comprehensive transport phenomena and chemical reactions in crystal growth from vapor. Different numerical models, their limitations, and future treads are then reviewed. GaN thin-film growth by the iodine vapor-phase epitaxy (IVPE) technique is presented in this chapter as an example for discussion.

37.1.1 Various GaN/AlN Vapor-Growth Systems

For conventional semiconductors such as silicon, the Bridgman or Czochralski methods are usually used to grow ingots from a melt. However, the high vapor pressure of nitrogen at the melting temperature

Table 37.1 Research groups of GaN/AlN growth

Researchers	Growth techniques	Substrate	Growth conditions	Size	Maximum growth rate
Speck et al. (1997–2006)	MBE [37.1–8]	MOCVD GaN/ sapphire template (0001)	GaN: 10^{-11} Torr 650–800 °C	–	750 nm/h
Karpov et al. (1998–2004)	HVPE [37.9, 10]	–	GaN: 760 Torr 900–1100 °C	–	GaN: 40 μm/h
	Sublimation growth [37.11–15]	GaN:6H-SiC AlN:SiC	GaN: 1100–1250 °C AlN: 760 Torr; > 1700 °C	GaN: $15\times15\times0.5$ mm^3 AlN: $10\times10\times0.08$ mm^3	GaN: 1 mm/h AlN: 1 mm/h
Edgar et al. (2000–2006)	MOCVD [37.16–18]	GaN:3C-SiC/ Si(100)	GaN: 76 Torr; 950 °C	–	GaN: 1.5 μm/h
	Sublimation growth [37.19–24]	AlN:6H-SiC	AlN: 100–800 Torr; 1700–1900 °C	–	AlN: 1 mm/h
Bliss et al. (1999–2005)	CVRP [37.25]	GaN:quartz	GaN: 1 atm; 900 °C	GaN: 9×2 mm$^2\times100$ μm	GaN: 25–100 μm/h
	IVPG [37.26, 27]	GaN:sapphire	GaN: 75–750 Torr; 910–1025 °C	GaN: 32 μm thickness 0.05–1 mm diameter	GaN: 10–11 μm/h
	HVTE [37.28, 29]	GaN:sapphire	AlN: 3.5–760 Torr, 1100–1300 °C	AlN: 75 μm thickness 2.5–5 mm diameter	AlN: 40 μm/h
Sitar (2001–2006)	IVPE [37.30]	GaN/AlN:sapphire	GaN: 400–800 Torr, 1000–1200 °C	GaN: 50 mm in diameter	GaN: 75 μm/h
	Sublimation growth [37.31–37]	AlN/SiC	AlN: 500–900 Torr; 2200–2300 °C	AlN: 18 mm in diameter	AlN: 5 mm/h
Slack et al. (2000–2005)	Sublimation growth [37.38–40]	–	AlN: ≈ 1800 °C (Rojo et al., 2001)	AlN: 150 mm^2	AlN: 0.9 mm/h
Spencer et al. (1993–2004)	MOCVD [37.41–45]	4H-SiC or 6H-SiC	AlN: 10 Torr, 1160–1190 °C		AlN: 1 μm/h

of GaN/AlN hampers melt growth of the materials. Other growth techniques were developed to fabricate GaN/AlN thick films at lower temperature. Metalorganic chemical vapor deposition (MOCVD) [37.52–55] and molecular-beam epitaxy (MBE) [37.56–59] technologies have been extensively used to grow high-quality GaN/AlN films for device applications. Both methods, however, have the drawback of high cost. Owing to its cost effectiveness, the halide vapor-phase epitaxy (HVPE) technique has gained more attention for depositing thick AlN/GaN layers [37.25, 26, 60–63]. In HVPE, the growth rate is mainly determined by the mass flow rate of the reactants, since surface reactions inside the reactor are close to equilibrium due to high temperature on the substrate. High growth rates ($> 20\,\mu m/h$) can easily be achieved in HVPE. Besides HVPE, sublimation methods [37.11, 31, 64–69] have also been used for AlN/GaN growth. In sublimation growth, high growth rate and good quality crystal are achieved [37.70] as the results of using high growth temperature ($\approx 2000\,°C$) and employing the repeated seeding method [37.65]. Different AlN/GaN vapor-growth systems are summarized in Table 37.1.

HVPE method has been used to produce GaN thick films or ingots with growth rates of up to $50\,\mu m/h$ and acceptable thickness uniformity [37.73, 74]. In the process, GaN is grown from the reactions of $NH_3/GaCl$ or $NH_3/GaCl_3$. The use of HCl gas to obtain $GaCl_x$ in the reactor creates contamination due to its corrosive effects. To prevent this problem, *Bliss* et al. [37.26, 27] developed an iodine vapor-phase epitaxy (IVPE) growth system for the first time, in which solid iodine is vaporized and reacts with gallium source to form GaI_x. Using a horizontal IVPE system, GaN film at growth rate on the order of $10\,\mu m/h$ has been obtained, and a low level of yellow luminescence of thus-produced GaN film has been observed, indicating the beneficial effect of reduced contamination in the process gases. To investigate the feasibility of higher GaN growth rate and better deposition quality, a vertical up-flow IVPE system for GaN has been designed and built at North Carolina State University to grow GaN thick film on a squared SiC substrate [37.30, 71, 72].

A schematic of the vertical reactor is presented in Fig. 37.1a. Iodine vapor is used to transport Ga from source to substrate to grow GaN. Ammonia gas is introduced to provide the nitrogen source for GaN growth. The reactor consists of four concentric tubes: furnace tube, reactor tube, outer silica tube, and inner silica tube. Iodine is carried by N_2 and H_2 gases flowing upwards through the inner silica. At the top of the inner silica tube is the gallium source within a BN holder. The iodine reacts with Ga melt to form GaI_x. A silica nozzle is placed above the gallium source to enhance Ga replacement. The gas flowing out of the silica nozzle is, therefore, a mixture of nitrogen, iodine, hydrogen, GaI_x, and gallium. N_2 gas flows between the inner and outer silica tubes to shield the Ga from the ammonia in the gas phase and enhance reactions on the substrate. Finally, the ammonia and nitrogen mixture flows through the space between the reactor tube and the outer silica tube. The system is heated by a re-

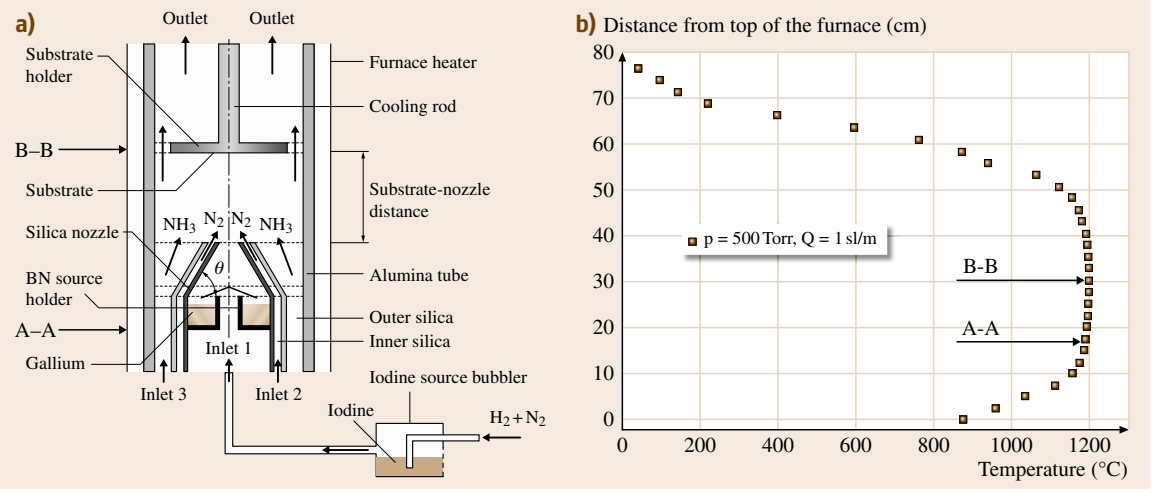


Fig. 37.1 **(a)** Schematic of the reaction chamber and **(b)** measured temperature profile from top to bottom along the furnace wall (after [37.30, 71, 72])

Table 37.2 List of simulation models for CVD process

Model	Length scale	Function
Ab initio methods, density function theory (DFT) [37.75, 76]	Electronic scale: 10^{-13}–10^{-9} m	Calculate transition state structures, surface reaction pathways, etc.
Molecular dynamics [37.75, 77, 78]	Atomic scale: 10^{-10}–10^{-6} m	Simulate the real atomic motion
Kinetic Monte Carlo (KMC) [37.79–84]	Microstructure scale: 10^{-9}–10^{-5} m	Bridge atomic scale and microscopic scale in dynamic simulations
Level-sets or hybrid models [37.77, 85, 86]	Film scale 10^{-7}–10^{-3} m	Simulate grain growth, recrystallization, etc.
Continuum models [37.87, 88]	Reactor scale 10^{-5}–10 m	Simulate heat and mass transport in the CVD reactor

sistant heater. Calibration of the furnace temperature has been conducted under reactor pressure of 500 Torr and N_2 flow rate of 1 slm. Figure 37.1b shows the temperature profile measured by a thermocouple along the furnace wall. By setting the growth temperature at 1200 °C, the uniform-temperature zone achieved is about 37.1 cm. The gallium source (position A–A) and the substrate (position B–B) are both positioned in the uniform-temperature zone. The gallium source is positioned close to the bottom of the zone to ensure that the entire growth area is located in the uniform-temperature zone.

37.1.2 Modeling of AlN/GaN Vapor Deposition

Physics-based theoretical modeling and simulation are widely used for better understanding of growth mechanisms and identifying the important issues related to growth processes, material characteristics, and dopant incorporation. Advanced models integrating the key aspects in the growth process can help identify the material stability limits, define the ideal parametric window(s), and improve crystal perfection. In addition, the physical mechanisms that limit the growth process can be determined.

Vapor deposition modeling presents challenges in two aspects. First, except for some well-studied important reactions such as homogeneous and heterogeneous pyrolysis of silane, the kinetic data of gas-phase and surface reactions are limited. In this chapter, combined with experimental observation, theoretical methods are used to obtain insight into the chemical mechanisms and reaction kinetics of IVPE. Second, vapor deposition involves complex transport phenomena occurring on different length and time scales. The characteristic length and time for atomic surface diffusion and relaxation and chemical reactions are approximately 10^{-10} m and 10^{-12} s, respectively. The typical reactor size is about 10^{-1} m and the deposition time is 10^{3}–10^{5} s. A simulation model accurately describing both the microscale molecular motion and the macroscale species transport remains elusive due to the tremendous memory and extremely fast central processing unit (CPU) required for such a model. Currently, most numerical models have focused only on particular scales, either macroscale or microscale. Different models and their corresponding scales are summarized in Table 37.2.

Among the various models, only continuum models have been extensively used for vapor-growth reactor design. More advanced models tend to predict

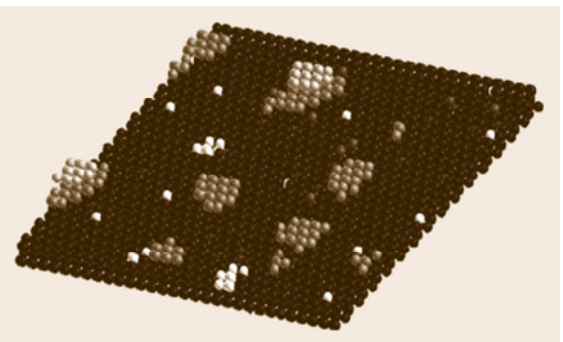


Fig. 37.2 Simulated Si(111) surface after deposition of 12% of C monolayer. *Black* and *white spheres* denote substrate atoms and vacancies created by thermal fluctuations, respectively. *Dark grey spheres* denote carbon atoms incorporated into the substrate. Atoms that are combined into the clusters are shown in *light grey* (*lighter tint* is used for upper cluster layers) (after CMS, with permission from Elsevier)

Table 37.3 Researchers and their models of GaN/AlN vapor growth

Authors	Models
Liu and *Edgar* [37.90, 91]	AlN sublimation growth: 2-D thermal convection; Stefan flow model; surface kinetics model
Karpov et al. [37.9–11, 14, 92, 93]	GaN sublimation growth: 2-D gas dynamics and gaseous species model AlN sublimation growth: Diffusive and transport kinetic model GaN MBE growth: Surface kinetic model GaN HVPE growth: 3-D gas flow and growth rate model GaN MOCVD growth: Surface kinetic model
Safvi et al. [37.94, 95]	GaN HVPE growth: Effects of flow rate, V/III ratio, and geometry on growth rate GaN MOCVD growth: Growth rate model
Theodoropoulos et al. [37.96]	GaN MOCVD growth: 2-D transport model; kinetic model
Aujol et al. [37.97]	GaN HVPE growth: Thermodynamic and kinetic model of growth rate
Dollet et al. [37.98]	AlN HVPE growth: Gas-phase and surface reaction analysis of $AlCl_3$ with NH_3; 2-D species transport modeling
Zhang et al. [37.29, 37, 72, 99–101]	AlN sublimation growth: 2-D induction heating and vapor transport; growth morphology and growth rate prediction AlN HVPE growth: Gas-phase and surface reactions modeling; growth optimization of geometric and operating conditions GaN IVPE growth: Thermodynamic and kinetic analysis and modeling of gas-phase and surface reactions; growth rate prediction

the microstructure and properties of materials. Figure 37.2 shows three-dimensional SiC clusters on a Si substrate modeled by the kinetic Monte Carlo (KMC) method [37.89]. Parameters used in the KMC model are estimated from molecular-dynamics simulation and by fitting data to experimental transmission electron microscopy (TEM) results. Formation of three-dimensional clusters and pits in the surface is successfully demonstrated.

These models, however, are still under development and their use is still limited to research purposes because of numerous approximations used in the models. A model combining two or more of the aforementioned models is called a multiscale model. A few reviews of multiscale modeling of vapor deposition thin-film growth are available in the open literature [37.102–106]. It is difficult for this method to consider the effect of feedback from the lower-scale model on the results of the upscale model [37.102].

Due to the lack of detailed reaction mechanisms, simulations of AlN/GaN vapor growth have been focused on the continuum modeling and solved gas flow, heat transfer, species diffusion, and chemical reactions in a vapor deposition reactor. A summary of the relevant research groups and their models is given in Table 37.3.

The continuum model will be used here to study the transport phenomena in the vapor-growth reactors for AlN/GaN growth. Gas-phase and surface reactions in the reactor will be studied and modeled to predict the species concentrations above the surface of the substrate accurately. The information obtained from the model will then be passed onto a surface deposition model to predict the AlN/GaN deposition rate distribution on the substrate.

37.2 Mathematical Models for AlN/GaN Vapor Deposition

Characteristics of chemical vapor deposition process are determined by gas hydrodynamics and chemical kinetics, which are affected by reactor geometry and process conditions. To reduce the complexity, the following assumptions are usually made for simulations:

1. The gas mixture is treated as a continuum
2. An ideal gas law is used
3. Gas flow is laminar
4. Gas mixture in the CVD reactor is radiative-transparent
5. Viscous dissipation of gas mixture is neglected.

It should be noted that these assumptions do not essentially limit the accuracy of the model for a wide range of vapor deposition process conditions.

37.2.1 Transport Equations

Based on these assumptions, the following equations can be used to describe the vapor deposition process:

Continuity
$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{V}) = 0 \,. \tag{37.1}$$

Momentum
$$\frac{\partial}{\partial t}(\rho \boldsymbol{V}) + \nabla \cdot (\rho \boldsymbol{V}\boldsymbol{V}) = -\nabla p + \rho \boldsymbol{g} \,. \tag{37.2}$$

Energy
$$\frac{\partial}{\partial t}(\rho h) + \nabla \cdot (\rho \boldsymbol{V} h) = \nabla\left(\frac{k}{C_\mathrm{p}} \nabla h\right) + \tau_{ij} \frac{\partial u_i}{\partial x_j} + q'''_\mathrm{latent} + q'''_\mathrm{radi} + q'''_\mathrm{eddy} + q'''_\mathrm{react} \,. \tag{37.3}$$

Species transport (in mass fractions Y_i)
$$\frac{\partial}{\partial t}(\rho Y_i) + \nabla(\rho \boldsymbol{V} Y_i) = \nabla(\rho D \nabla Y_i) + \dot{w}_i \,. \tag{37.4}$$

Ideal gas law
$$p = \rho \mathrm{RT} \sum_i \frac{Y_i}{W_i} \,. \tag{37.5}$$

Here ρ is the density, $\boldsymbol{g}$ is the gravitational acceleration vector, $\boldsymbol{V}$ is the gas velocity, p is the pressure, τ is the viscous stress tensor, T is the temperature, k is the thermal conductivity, D is the binary diffusion coefficient of reactant in carrier gas, Y_i is the mass fraction of species i, $\dot{w}_i$ is the gas-phase reaction rate of species i, W_i is the molecular weight of species i, and R is the gas constant. $q'''_\mathrm{latent} = \Delta H_\mathrm{vs} \dot{M}_\mathrm{vs}$ is the latent heat of phase change with ΔH_vs as the specific latent heat and $\dot{M}_\mathrm{vs}$ as the deposition rate; q'''_radi is the heat exchange rate due to radiation; $q'''_\mathrm{eddy} = \frac{1}{2}\sigma_\mathrm{c}\omega A_0 A_0^*$ is the power dissipation rate due to an induced eddy current in the susceptor, in which σ_c is the electrical conductivity of the susceptor, ω is the alternate current frequency in the induction coils, and A_0 is the induced magnetic potential; q'''_react is the heat release or absorption due to reactions. Calculation of q'''_radi will be detailed in the radiative heat transfer model.

Critical Condition for Radiation Dominance. Radiative heat transfer becomes important when temperature is high ($\approx 1000\,^\circ\mathrm{C}$). Heat fluxes transferred between the hot and cold walls due to heat conduction, convection, and radiation are calculated analytically to show the importance of radiative heat transfer. Considering a typical reactor for GaN/AlN growth, the distance between two walls d is chosen as 0.1016 m (5 inches), and the length of the walls is 0.508 m (20 inches). The radiative emissivity of the walls is 0.25 (if the reactor is made of metal). Temperature for the cold wall is 300 K, and the hot-wall temperature varies from 300 to 2000 K. Nitrogen is running between the walls with a Reynolds number of 100. Such a system can be approximated as a one-dimensional (1-D) problem for a first analysis. Heat flux due to heat conduction, convection, and radiation can then be calculated. The conductive heat flux is

$$q'' = k\frac{\Delta T}{d} = k\frac{T_\mathrm{hot} - T_\mathrm{cold}}{d} \,. \tag{37.6}$$

The convective heat flux is [37.107]

$$q'' = h\Delta T = \frac{k\mathrm{Nu}}{d}\Delta T = 0.46 k \mathrm{Re}_\mathrm{f}^{0.5} \mathrm{Pr}_\mathrm{f}^{0.43} \left(\frac{\mathrm{Pr}_\mathrm{f}}{\mathrm{Pr}_\mathrm{w}}\right)^{0.25} \times \left(\frac{d}{L}\right)^{0.4} (T_\mathrm{hot} - T_\mathrm{cold})/d \,. \tag{37.7}$$

The radiative heat flux is

$$q'' = \frac{\sigma\left(T_\mathrm{hot}^4 - T_\mathrm{cold}^4\right)}{\frac{2(1-\varepsilon)}{\varepsilon} + \frac{1}{F_\mathrm{h\text{-}c}}} \,, \tag{37.8}$$

where k is the thermal conductivity of nitrogen, h is the convective heat transfer of cold nitrogen on the hot wall surface, Nu is the Nusselt number, Re_f is the Reynolds number of the nitrogen flow, Pr_f is the Prandtl number with the reference temperature of the gas, Pr_w

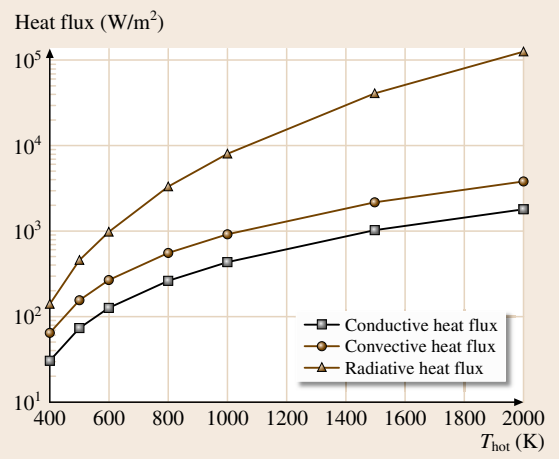


Fig. 37.3 Comparison of heat flux due to conduction, convection, and radiation

is the Prandtl number with the reference temperature of the hot wall, σ is the Stefan–Boltzmann constant (5.672×10^{-8} W/m^2 K^4), and F_{h-c} is the view factor between the hot and cold walls, which is calculated to be 0.82. The heat fluxes calculated from (37.6–37.8) are summarized in Fig. 37.3.

It is found that the radiative heat flux is an order of magnitude higher than conduction or convection when the hot-wall temperature is 1000 K. Radiation will therefore be dominant when the temperature is higher than 1000 K.

The integro-differential radiative heat transfer equation for calculating the radiative heat flux can be written as follows

$$\frac{\mathrm{d}I(\boldsymbol{r},\boldsymbol{s})}{\mathrm{d}s} = -(\alpha_a+\sigma_s)I(\boldsymbol{r},\boldsymbol{s}) + \alpha_a I_b(\boldsymbol{r}) + \frac{\sigma_s}{4\pi}\int_0^{4\pi} I(\boldsymbol{r},\boldsymbol{s}')\Phi(\boldsymbol{s}\cdot\boldsymbol{s}')\,\mathrm{d}\Omega', \qquad (37.9)$$

where I is the radiation intensity, $\boldsymbol{r}$ is the position vector, $\boldsymbol{s}$ is the direction vector, α_a is the absorption coefficient, and σ_s is the scattering coefficient. I_b is the intensity of the black body, Φ is the phase function of energy transfer, and Ω' is the solid angle of the radiation beam.

Thermal radiation can be transported by the following four mechanisms:

1. From surface to surface
2. From surface to volume
3. From volume to surface
4. From volume to volume.

Equation (37.9) is applicable for an absorbing, emitting, and scattering medium. For some media, a simplified model can be used instead of (37.9). For inert gases, which can be considered as totally transparent to radiation, only surface-to-surface radiation should be considered instead of solving (37.9). It is therefore important to select an appropriate radiation model for a certain radiation problem to save calculation time. Optical thickness is usually used as a critical parameter to determine which radiation model should be used. The optical thickness L is defined as follows:

$$L = \alpha R\,, \qquad (37.10)$$

where α is the absorption coefficient and R is the characteristic length of the medium. Some of the most commonly used radiation models and their applicable regimes in reference to optical thickness are summarized in Table 37.4, together with the radiation transportation considered in the model and the property of the radiative surface.

The first three models in Table 37.4 do not require (37.9) to be solved. The discrete ordinate model that needs to solve (37.9) will be used in most calculations presented in this chapter.

37.2.2 Growth Kinetics

In a typical chemical vapor deposition system, a thin film is formed by deposition from surface chemical reactions. Reactive molecules that contain atoms of the material to be deposited are introduced as a mixture of gases (usually diluted with inert carrier gases such as argon and nitrogen) into the temperature- and pressure-controlled environment of the reacting chamber in which the targets (wafers) on which deposition takes place are located. Heterogeneous reactions take place on the substrate. For example, in the case of epi-silicon production from silane pyrolysis, some initial portions of the overall reaction may occur in the gas phase, but the final stage of the reaction forming elemental silicon occurs on the wafer surface. Homogeneous gas-phase reactions may also occur, either preceding, parallel with, or in competition with the heterogeneous surface reactions. As the precursor gas approaches the wafer surface, it may react in the gas phase due to the high temperature (> 400 °C at 1 atm). The reactions may form intermediate species, such as SiH_2, or a series of reactions may proceed to form el-

Table 37.4 Summary of radiation models

Radiation model	Optical thickness (m)	Radiation transport mechanisms	Radiation surface
P1 model [37.108, 109]	$L > 1$	(a) (c)	Gray
Rosseland model [37.108]	$L > 3$	(a) (c)	Gray
Surface-to-surface model [37.110]	$L \ll 1$	(c)	Gray
Discrete exchange factor model [37.111]	$0 < L < +\infty$	(a) (b) (c) (d)	Gray or non-gray
Discrete ordinate mode [37.112]	$0 < L < +\infty$	(a) (b) (c) (d)	Gray or non-gray
Discrete transfer model [37.113]	$0 < L < +\infty$	(a) (b) (c) (d)	Gray

emental silicon. In the latter case, silicon atoms may form small silicon grains in the gas phase. These grains then flow out of the system (sometimes causing an additional problem) or migrate to the wafer surface, where they form an irregular porous deposit, as well as loosely adherent particles. This gas-phase or homogeneous reaction does not form the dense uniform films needed for integrated-circuit applications and is usually suppressed in favor of heterogeneous reactions.

The sequence of events that may take place during vapor deposition process is shown in Fig. 37.4. It includes:

1. Reactant gases enter the reaction area inside the reactor by convection and diffusion.
2. Homogeneous gas-phase reactions take place, and intermediate species may be formed.
3. Species diffuse trough the boundary layer onto the surface.
4. Adsorption and diffusion of these species occur on the surface.
5. Heterogeneous reaction, nucleation, and lattice incorporation take place on the surface, leading to the formation of a solid film.
6. Desorption of adsorbed species away from the surface through the boundary layer.
7. Byproduct is transferred away from the reaction area in the reactor chamber.

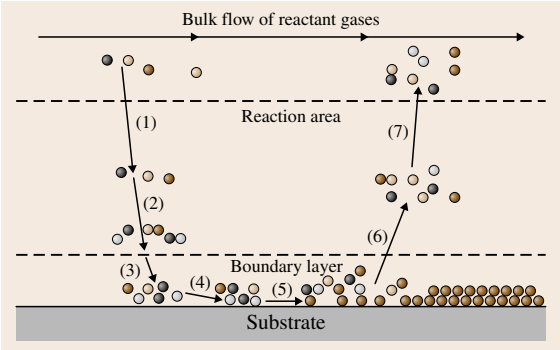


Fig. 37.4 Sequence of events during vapor deposition

Among these steps, the slowest will determine the deposition rate. A species concentration layer only exists when the reactor pressure is relatively high. If the gases pressure inside the reactor is low (i. e., in the range of mTorr), the species concentration boundary layer is no longer applicable, since the species transportation speed due to diffusion is much faster than the speed due to convection.

It is clear that many of the steps involved are transport related. The extent and role of the transport phenomena are, however, determined by process parameters such as substrate temperature, flow rate, and reactant partial pressure, as well as the chemistries involved. Depending on which of these is the dominant factor, the deposition process may be thermodynamical, diffusion or kinetics controlled.

In a thermodynamically controlled process, the mass transfer of species to and from the deposition zone is much slower than either mass transfer between the main flow and the substrate or mass transfer from the surface processes. Steps 1 and 7 are therefore rate-controlling steps. The process is assumed to proceed under thermodynamic equilibrium, and the deposition rate is generally determined by the equilibrium values of the partial pressures of species in the system.

In a diffusion- or mass-transport-controlled process, the rate-determining step is the diffusion or transport of the reactant gases to the substrate surface. The gas flow, heat transfer, and species diffusion play the dominant roles in determining the deposition characteristics.

Finally, in a kinetics-controlled system, the surface processes are not as fast as steps 1 and 7, nor 3 and 6. The rate-determining step is therefore the slowest of steps 4 and 5. Substrate temperature then

plays an important role whereas the flow field has only a marginal influence. For some combinations of temperature and concentration, the homogeneous gas-phase reaction plays an important role and determines the deposition rate.

37.2.3 Numerical Solution

The simulation procedures are summarized as follows:

1. Grids are generated for calculating the electromagnetic field and heat and mass transfer, respectively.
2. The electromagnetic potentials are calculated. If the electrical/magnetic properties are dependent on temperature, this procedure will be repeated at a certain time interval until a steady-state temperature distribution is achieved. The calculated power dissipation distribution will be used as a source term for the energy transport calculation.
3. Pressure and velocity fields are calculated.
4. Energy transport and radiative heat transfer equations are solved iteratively at each time step until convergence is reached.
5. The species transport equations are solved for every species.
6. The thermoelastic stress distribution in the crystal may be solved based on the temperature field obtained in the crystal.

It should be noted that, in the program, steps 2–5 have to be repeated in an iterative way until the final variable fields stabilize.

37.3 Characteristics of AlN/GaN Vapor Deposition

The purpose of studying the characteristics of transport phenomena in a CVD system is to achieve fast and effective design of CVD reactors. In this section GaN growth using an IVPE system will be used as an example to identify the critical transport parameters. Heat and mass transfer will be analyzed theoretically for the following two reasons. First, this can provide useful information for the control of the transport process. Second, analytical results can be used to validate the accuracy of numerical simulation. Gas-phase and surface reactions will be analyzed thermodynamically and kinetically. Based on this analysis, a model for gas-phase and surface reactions can be provided.

37.3.1 Theoretical Analysis of Heat and Mass Transfer

Important Dimensionless Groups

An order-of-magnitude study provides a basic understanding of the complex heat and mass transport phenomena in a vapor-growth system. The important dimensionless groups and their definitions are listed in Table 37.5.

For a typical vapor deposition system, the Reynolds number (Re) is usually in the range of 1–100, which corresponds to laminar flow. The Grashof number is usually kept below 10^5. A large value of Gr indicates strong natural convection in the system, leading to a complicated flow pattern in the growth chamber. The ratio of the Grashof number Gr to the Reynolds number squared, Re^2, is used to quantify whether heat transfer is mainly controlled by natural or forced (inertia) convection. The regimes for natural convection, forced convection, and mixed convection are shown in Table 37.6 [37.114].

The value of the Prandtl number (Pr) is usually in the range of 0.1–0.7 for gas. The Schmidt number (Sc) is in the range of 0.01–1.0, depending strongly on pressure. At low pressure the species exhibits strong diffusion and the Schmidt number will lie towards the low end of this range. The thermal and mass Péclet numbers (Pe^T and Pe^M) are usually in the range of 1–100. The Damköhler numbers (Da) for the gas phase and surface are reaction dependent. A small value ($\ll 1$) means that the gas-phase or surface reactions are negligible. These dimensionless groups, along with the governing equations and appropriate boundary conditions, can provide the entire picture of a vapor deposition process.

Analytical Analysis of Gas Flow

For a conventional vapor-growth reactor, one of the key issues is maintaining a suitable growth environment near the substrate area, which is usually achieved by controlling the flow, temperature, and species concentration. To achieve an appropriate gas flow environment in the reactor, the relationship between the reactor diameter, total gas flux, and Reynolds number is studied. The Reynolds number can have the form

$$\mathrm{Re} = \frac{\rho Q d}{A\mu}\,, \tag{37.11}$$

Table 37.5 Important dimensionless groups in a CVD reactor

Dimensionless group	Definition	Physical interpretation
Reynolds (Re)	$\frac{\rho_0 V_0 L_c}{\mu_0}$	Inertial forces/ viscous forces
Grashof (Gr)	$g\rho_0^2 L_c^2 (T_w) - \frac{T_{in}}{(\mu_0^2 T_0)^2}$	Buoyancy forces/ viscous forces
Prandtl (Pr)	$\frac{\mu_0 C_{p0}}{k_0}$	Momentum diffusivity/ thermal diffusivity
Schmidt (Sc)	$\frac{\mu_c}{\rho_0 D}$	Momentum diffusivity/ species diffusivity
Thermal Péclet (Pe^T)	$Re \cdot Pr$	Convective heat transfer/ conductive heat transfer
Mass Péclet (Pe^M)	$Re \cdot Sc$	Convective mass transfer/ conductive mass transfer
Gas-phase Damköhler (Da^g)	$\frac{r_0^g L_c}{V_0}$	Typical time for flow/ typical gas-phase reaction time
Surface Damköhler (Da^s)	$\frac{r_0^s L_c}{V_0}$	Typical time for flow/ typical surface reaction time

ρ_0, μ_0, k_0, C_{p0}, D_0, and β_0 are the values of ρ, μ, k, C_p, D, and β ($\beta = 1/T_0$ for ideal gas) at the average process gas temperature $T_0 = (T_w + T_{in})/2$, where T_w and T_{in} are the temperature on the reactor wall and the temperature at inlet of the reactor, respectively. L_c is the characteristic length, defined as the diameter of the reactor, V_0 is the characteristic velocity, defined as the inlet process gas velocity, r_0^g is the gas-phase reaction constant, and r_0^s is the surface reaction constant.

where ρ is the density of the mixing gases, Q is the total volume flow rate, d is the diameter of the reactor, A is the area, and μ is the dynamic viscosity. In the GaN growth system presented here, nitrogen and ammonia are the mass-dominant species in the reactor, and their molar ratio is assumed to be 1 : 1. The dynamic viscosities of nitrogen and ammonia are very similar. Sutherland's law is used to determine the dynamic viscosity $\mu = (AT^{3/2})/(B+T)$, where the coefficients are $A = 1.461 \times 10^6$ and $B = 79.96$. The relationships of the volume flow rate and density with temperature and pressure can be expressed as $Q = Q_0(T/T_0)(P_0/P)$ and $\rho = \rho_0(T_0/T)(P/P_0)$, respectively.

Equation (37.11) can be rewritten as

$$\mathrm{Re_D} = \frac{0.907\rho_0 Q_0 (111.5 + T)}{\mathrm{d}T^{3/2}}\,, \qquad (37.12)$$

where ρ_0 (1.005 kg/m^3) and Q_0 are the density and flow rate of the mixture under standard conditions (0 °C and 1 atm). For a typical vapor-growth system, the Reynolds number is usually kept in the range 1–100 to maintain a stable, laminar flow environment, which allows a sufficient residence time of reactants on the substrate for surface deposition. Figure 37.5 shows the relationship between the volume flow rate and reactor diameter at different Reynolds numbers. Experimental operating temperature of 1050 °C and pressure of 200 Torr are used in the calculation. It is shown in Fig. 37.5 that the total flow rate should be controlled at about 1 slm in order to achieve a Reynolds number of 10 for a reactor diameter of 5 cm.

Analytical Analysis of Heat and Mass Transfer

Besides forming a laminar-flow condition, a uniform species concentration distribution on the substrate is also preferred, as this is essential for deposition quality. Since gas flow is laminar in the reactor, there is no direct mixing. The transport of NH_3 from the near wall to the substrate surface will be mainly achieved through molecular diffusion in the stream. It is therefore impor-

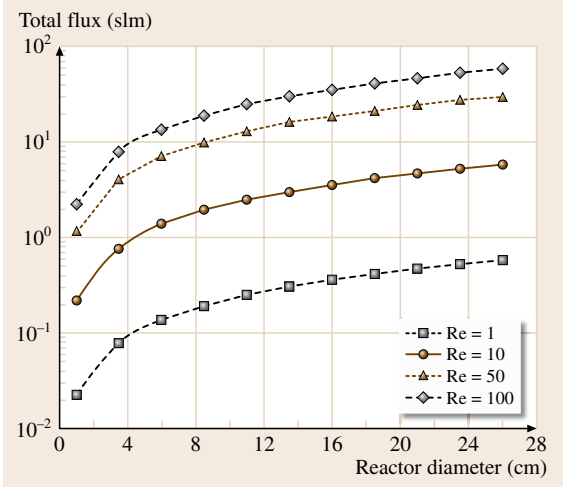


Fig. 37.5 Relationship between chamber diameter and gas flow rate with different Reynolds numbers

Table 37.6 Criteria for natural and forced convection

Convection	Gr/Re^2 criteria
Natural	> 16
Mixed	$\in [0.3, 16]$
Forced	$\in [0, 0.3]$

tant to allow ammonia sufficient time to diffuse to the substrate surface. This can be controlled by the total flow rate Q, the diameter of the reactor d, the diffusion coefficient of NH_3 in N_2 D, the distance between the substrate and the silica nozzle L, and the temperature and pressure in the reactor. To estimate the mixing time needed, or to select an appropriate substrate–nozzle distance (Fig. 37.1a) at different flow rates, it is assumed that the reactor tube is filled with N_2 with a flow rate of Q, and at $t = 0$ an NH_3 concentration of C_s is added on the wall of the reactor. According to [37.115], by solving the diffusion problem for a laminar flow in a rounded tube, the dimensionless average NH_3 concentration can be expressed as

$$\Delta' = \frac{C_s - C_{av}}{C_s - C_0} = 0.819\,e^{-14.6272\beta''} + 0.0976\,e^{-89.22\beta''} + 0.0135\,e^{-212.2\beta''} + \ldots \quad (37.13)$$

where C_{av} is the averaged NH_3 concentration at a certain axial position X, $\beta'' = DX/4r^2\bar{U}_{av}$, D is the binary diffusion coefficient, r is the radius of the cylinder, $\bar{U}_{av}$ is the averaged gas velocity, and $C_0 = 0$ is the initial concentration of NH_3 in the tube. According to (37.13), when $\beta'' = 0.2$, $C_{av} = 0.95C_s$ with an exposure length of X. It is therefore concluded that, when $\beta'' = 0.2$, different species have been mixed very well due to molecular diffusion. In the growth reactor, $\bar{U}_{av} = Q/A$, and the total exposure length, or the substrate–nozzle distance, is L. The following equation has to be satisfied for a uniform species concentration distribution on the substrate surface

$$\beta'' = 0.2 = DL\frac{A}{4r^2Q}\,. \quad (37.14)$$

The diffusion coefficient of NH_3 in N_2 can be estimated from [37.90]

$$D = D_0\left(\frac{T}{T_0}\right)^n\left(\frac{P_0}{P}\right), \quad (37.15)$$

with $D_0 = 2.3\times10^{-5}\,m^2/s$, $T_0 = 298\,K$, $P_0 = 760\,Torr$, and $n = 1.8$. The influence of species concentration on the diffusion coefficient can be neglected. Equation (37.14) can therefore be rewritten as

$$Q_0 = \frac{\pi D_0\left(\frac{T}{T_0}\right)^{0.8}}{4\beta''}L\,, \quad (37.16)$$

with $\beta'' = 0.2$ and $T = 1373\,K$. The relationship between the standard flow rate and the substrate–nozzle distance is given by

$$Q_0 = 1.62L\,. \quad (37.17)$$

According to (37.17), when the total flow rate increases, a larger substrate–nozzle distance is required to achieve a uniform species concentration distribution on the substrate. For a typical flow rate of 3 slm, the corresponding substrate–nozzle distance for complete mixing is about 11.9 cm.

Temperature control in a vapor-growth system is critical. To obtain sufficiently high activation energy, reactants have to be preheated before they reach the substrate. The analytical method can be used to calculate the temperature variation of gases in the tubular reactor subject to different operational conditions such as flow rate, pressure and temperature of the furnace, and geometric conditions such as the diameter of the reactor. For the resistance heating system discussed here, we assume a constant temperature of $T_w = 1373\,K$ on the reactor wall and that the inlet gas mixture is composed of N_2 and NH_3 with a molar ratio of 1 : 1 and a temperature of $T_{in} = 300\,K$.

The analytical solution of the problem can be shown to be [37.116]

$$T_{m(x)} = T_w - (T_w - T_{in})\exp\left(-\frac{\alpha Nu}{r^2\bar{U}_{av}}z\right), \quad (37.18)$$

where $T_{m(x)}$ is the bulk longitudinal average temperature at position z, $\alpha_T = (k/(\rho Cp))$ is the thermal diffusion coefficient of the mixture, and $Nu = hd/k$ is the Nusselt number. Solution of the above problem gives $Nu = 3.66$. Since the average velocity in the round tube is $\bar{U}_{av} = Q/A = Q_0(T_{av}/T_0)(P_0/P)/(\pi r^2)$,

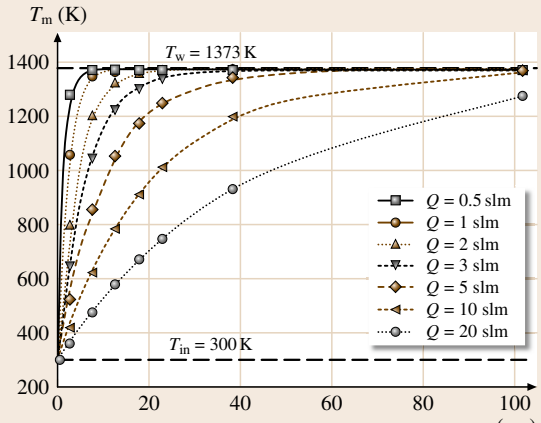


Fig. 37.6 Bulk mean temperature along the z-axis

(37.18) can be rewritten as

$$T_{m(z)} = T_w - (T_w - T_{in}) \times \exp\left(-\frac{3.66\alpha}{Q_0\left(\frac{T_{av}}{T_0}\right)\left(\frac{P_0}{P}\right)z}\right). \quad (37.19)$$

Equation (37.19) indicates that temperature is independent of the reactor diameter. With a reference pressure of 200 Torr and $T_{av} = 1000$ K, the thermal diffusion coefficients for NH_3 and N_2 are 8.49×10^{-4} and $6.22\times10^{-4}\,m^2/s$, respectively. We used the averaged thermal diffusion of $7.36\times10^{-4}\,m^2/s$ in (37.19) to calculate the mean temperature distribution.

Figure 37.6 shows the bulk mean temperature variation along the z-axis, i.e., the distance from the gas inlet in the axial direction, under different total flow rates. The inlet gas temperature in this case is 300 K and the wall temperature is 1373 K (dashed line). For a reactor tube length of 102 cm, the gas can be heated to the reactor wall temperature at a flow rate of less than 10 slm. For a typical flow rate of 3.0 slm, the temperature of mixed gas can reach 1300 K within a heating length of 17.8 cm.

In experiments, a uniform temperature distribution on the substrate surface is essential to achieve uniform deposition and reduce the thermal stress in thick films. Numerical simulations can be conducted to optimize operating and geometric conditions.

37.3.2 Thermodynamic and Kinetic Analysis of Chemical Reactions

Thermodynamic and kinetic analysis of reactions allows us to predict the growth rate on the substrate in order to improve geometric and operating conditions for fast and uniform deposition. Assuming that reactions are in quasithermodynamic equilibrium, reaction constants can be calculated, and the quasi-equilibrium partial pressures of the reactants and products can be determined. The rate of the gas-phase reaction and surface deposition can be calculated. Thermodynamic calculations, however, only provide an upper limit. For the situation with high flow rate and limited reaction space, reactions in a vapor deposition system could be far from the equilibrium status. Kinetic analysis is therefore needed to understand how fast the reactions can proceed in the system. Kinetic data for gas-phase and surface reactions required for numerical simulations can be obtained from experiments. By solving the gas-phase reaction as a volumetric term and the surface reaction as a boundary condition on the substrate surface, the distributions of species concentration in the CVD reactor and the deposition rate on the substrate surface can be predicted under various operating conditions.

Thermodynamic Analysis of Gas-Phase and Surface Reactions

Prediction of Gas-Phase Reactions. Gas-phase reactions can be analyzed thermodynamically under different species concentrations, temperatures, and pressures. There are two methods to predict gas-phase reactions. In the first method, the species present in the system have to be specified, however no reaction step will be assigned. The concentration of difference species will be determined by minimizing the Gibbs energy of the system. In the second method, reaction steps will be assigned, and the reaction constants will be used to determine the reaction rates.

In the first method, chemical equilibrium is reached at constant temperature and pressure when the Gibbs energy is minimized. The Gibbs energy per unit mass of a system with N species can be written as follows:

$$G = \sum_1^N n_i\mu_i = \sum_{i=1}^N n_i\left(\mu_i^0 + RT\ln\frac{p_i}{p_0}\right), \quad (37.20)$$

where p_0 is the reference pressure of 1 atm and n_i is the molar amount of species i. Since the chemical potential μ_i is a function of temperature and pressure, the Gibbs energy is minimized at constant T and p for the correct combination of n_i. Elements must be conserved by the change in composition, which adds additional constraints to the system of the form

$$\sum_{i=1}^N a_{ji}n_i = b_j \quad \text{for} \quad j = 1,\ldots,M, \quad (37.21)$$

where a_{ji} is the number of atoms of element j in species i, M is the total number of elements in the system, and b_j is the total number of moles of element j per unit mass.

The composition that minimizes the Gibbs energy while satisfying the element balances is obtained by introducing the function

$$\Theta = G + \sum_{j=1}^M \lambda_j\left(\sum_{i=1}^N a_{ji}n_i - b_j\right). \quad (37.22)$$

The quantities λ_j are termed Lagrangian multipliers. Since (37.21) must be satisfied to conserve elements, the second term on the right-hand side vanishes and the composition that minimizes Θ also minimizes G.

Differentiating (37.22) with respect to n_i gives

$$\frac{\partial \Theta}{\partial n_i} = \mu_i - \sum_{j=1}^{M} a_{ji}\lambda_j \quad \text{for } i = 1, \ldots, N . \tag{37.23}$$

Differentiating (37.22) with respect to λ_i gives

$$\frac{\partial \Theta}{\partial \lambda_i} = b_i - \sum_{j=1}^{N} a_{ji} n_i \quad \text{for } i = 1, \ldots, N . \tag{37.24}$$

Setting (37.23) and (37.24) to zero generates $N + M$ equations which can be solved to give the composition at the chemical equilibrium.

Thermodynamic gas-phase reaction analysis in the second method makes calculations simpler than in the first method since the reaction steps are already known. The gas-phase reaction from I_2 and H_2 will be used as an example to demonstrate the analysis process. The gas-phase reaction from I_2 and H_2 is given by

$$\tfrac{1}{2}H_2(g) + \tfrac{1}{2}I_2(g) \rightleftharpoons HI(g) . \tag{37.25}$$

The following equations can be obtained according to the above reaction

$$K = \exp\left(-\frac{\Delta G_r^0}{RT}\right) = \frac{P_{HI}^e}{\sqrt{P_{I_2}^e P_{H_2}^e}} , \tag{37.26}$$

$$P_{H_2}^0 - P_{H_2}^e = P_{I_2}^0 - P_{I_2}^e , \tag{37.27}$$

$$P_{HI}^e = 2\left(P_{I_2}^0 - P_{I_2}^e\right) , \tag{37.28}$$

where $P_{H_2}^0$ and $P_{I_2}^0$ are the initial partial pressures of H_2 and I_2; $P_{H_2}^e$, $P_{I_2}^e$, and P_{HI}^e are the equilibrium partial pressures of H_2, I_2, and HI; K is the reaction equilibrium constant; and ΔG_r^0 is the Gibbs free energy change. The dependence of ΔG_r^0 on temperature is shown in Fig. 37.7 [37.117, 118].

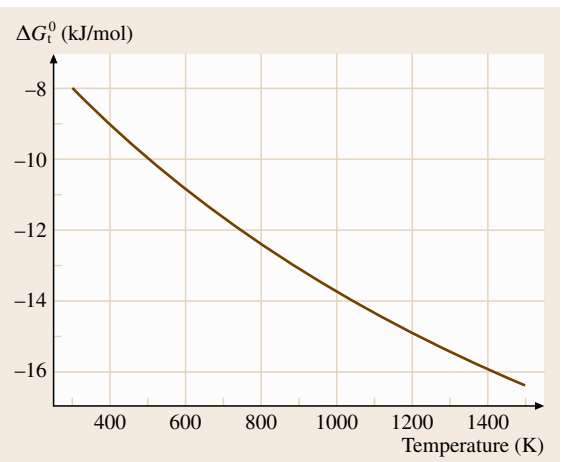


Fig. 37.7 Free energy change of reaction (37.25)

With the given reaction temperature, the Gibbs free energy change can be obtained. The equilibrium constant K can, therefore, be calculated. The reaction equilibrium pressure of each species in reaction (37.25) can be solved from (37.26–37.28) when the initial partial pressures of H_2 and I_2 are given.

Thermodynamic Prediction of Surface Reactions

Based on the assumption that surface kinetic limitations occur at the stage of species adsorption/absorption on the substrate surface, a quasithermodynamic approach can be used to predict the surface deposition rate with different sticking coefficients, growth temperatures, and gas supersaturations [37.92]. Surface deposition of GaN from GaI and NH_3 is used as an example to predict the deposition rate

$$GaI(g) + NH_3(g) \rightleftharpoons GaN(s) + HI(g) + H_2(g) . \tag{37.29}$$

Figure 37.8a and b show a schematic of GaN deposition on the substrate surface and a typical reactant species concentration boundary layer formed on the substrate due to surface reactions, respectively. The concentration of GaI before the mixing gas reaches the substrate area is defined as $C = C_0$. Due to the surface reaction of GaI with NH_3 on the substrate, the GaI concentration decreases to $C = C_w$ on the substrate surface. The concentration boundary layer χ is defined as the distance where the GaI concentration changes from C_0 in the bulk gas to C_w on the substrate surface.

Species are transported to the concentration boundary layer by the concentration gradient due to the surface deposition. Assuming that the decrease of species concentration from the bulk gas to the substrate surface is linear with the concentration boundary-layer thickness, a species mass flux in the gas phase can be expressed as

$$J_i = D_i M_i \frac{\partial C}{\partial n} = \frac{D_i M_i}{\chi}\left(C_i^0 - C_i^w\right) \ [\mathrm{kg/m^2 s}] , \tag{37.30}$$

where n is the coordinate normal to the substrate surface, C_i^0 and C_i^w are the species molar concentration in the bulk gas and on the deposition surface at the interface temperature, respectively, D_i is the

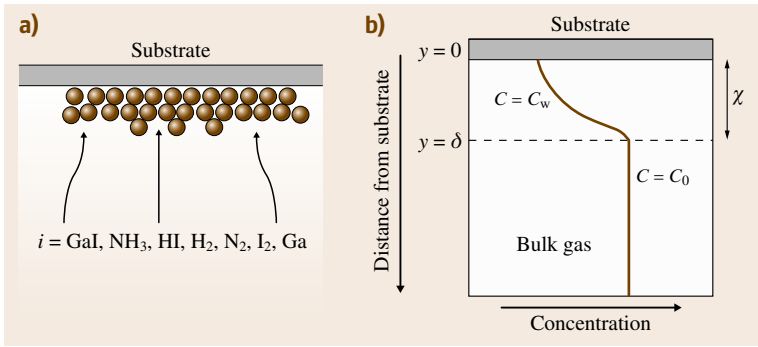


Fig. 37.8 (a) Schematic of GaN deposition on a substrate and (b) concentration boundary formed on the substrate due to surface reactions

diffusion coefficient of species i, M_i is the molecular weight of species i, and χ is the concentration boundary-layer thickness. According to the ideal gas law, $P_i = C_i \mathrm{RT}$, (37.30) can be rewritten as

$$J_i = \frac{D_i M_i}{\mathrm{RT}\chi}\left(P_i^0 - P_i^{\mathrm{w}}\right) \qquad [\mathrm{kg/m^2 s}]\,, \tag{37.31}$$

where P_i^0 and P_i^{w} are the species partial pressure in the bulk gas and on the deposition surface. At the substrate surface, according to the Hertz–Knudsen law, the interface mass flux of species i is given by

$$J_i = \alpha_i \beta_i \left(P_i^{\mathrm{w}} - P_i^{\mathrm{e}}\right) \qquad [\mathrm{kg/m^2 s}]\,, \tag{37.32}$$

where α_i $(0 \le \alpha_i \le 1)$ is the sticking coefficient of species i which accounts for all interfacial kinetics controlled by step and kink density, provided by two-dimensional nucleation and dislocations in the modern version of the Burton–Cabrera–Frank theory [37.119]; $\beta_i = \left(\frac{M_i}{2\pi \mathrm{RT}}\right)^{1/2}$ is the thermodynamic factor of species i and P_i^{e} is the equilibrium partial pressure of species i on the substrate surface.

The following equations can be obtained according to reaction (37.29)

$$\frac{J_{\mathrm{GaI}}}{M_{\mathrm{GaI}}} = \frac{J_{\mathrm{NH_3}}}{M_{\mathrm{NH_3}}} = \frac{\rho_{\mathrm{GaN}} V_{\mathrm{g}}}{M_{\mathrm{GaN}}}\,, \tag{37.33}$$

$$\frac{J_{\mathrm{GaI}}}{M_{\mathrm{GaI}}} + \frac{J_{\mathrm{HI}}}{M_{\mathrm{HI}}} = 0\,, \tag{37.34}$$

$$\frac{J_{\mathrm{GaI}}}{M_{\mathrm{GaI}}} + \frac{J_{\mathrm{H_2}}}{M_{\mathrm{H_2}}} = 0\,, \tag{37.35}$$

$$P_{\mathrm{HI}}^{\mathrm{e}} P_{\mathrm{H_2}}^{\mathrm{e}} = k_1 P_{\mathrm{GaI}}^{\mathrm{e}} P_{\mathrm{NH_3}}^{\mathrm{e}}\,. \tag{37.36}$$

According to (37.33–37.36), there are four unknowns, i. e., $P_{\mathrm{HI}}^{\mathrm{e}}$, $P_{\mathrm{GaI}}^{\mathrm{e}}$, $P_{\mathrm{NH_3}}^{\mathrm{e}}$, and $P_{\mathrm{H_2}}^{\mathrm{e}}$, and four equations. Given the wall partial pressures $P_{\mathrm{HI}}^{\mathrm{w}}$, $P_{\mathrm{GaI}}^{\mathrm{w}}$, $P_{\mathrm{NH_3}}^{\mathrm{w}}$, and $P_{\mathrm{H_2}}^{\mathrm{w}}$ and the sticking probabilities α_i of the species HI, GaI, NH_3, and H_2 on the substrate surface, the reaction rate can be obtained by solving (37.33–37.36).

Modeling of Gas-Phase and Surface Reactions Kinetically

Kinetic Modeling of Gas-Phase Reactions. A system that has a number N_{r} of gas-phase reactions involving a number N_{sp} of species can be expressed in general notation by

$$\sum_{i=1}^{N_{\mathrm{sp}}} \nu'_{ij} \Lambda_i = \sum_{i=1}^{N_{\mathrm{sp}}} \nu''_{ij} \Lambda_i, \quad j = 1, \ldots, N_{\mathrm{r}}\,, \tag{37.37}$$

where ν'_{ij} and $\nu_{ij}{}''$ are the forward and backward stoichiometric coefficients for the ith species in the jth reaction. The above equation can be written more compactly as

$$\sum_{i=1}^{N_{\mathrm{sp}}} \nu_{ij} \Lambda_i = 0, \quad j = 1, \ldots, N_{\mathrm{r}}\,, \tag{37.38}$$

where $\nu_{ij} = \nu_{ij}{}'' - \nu'_{ij}$. The stoichiometric coefficients are integers for elementary reactions and are normally 0, 1 or 2.

The molar production rate of species i due to gas-phase reaction is

$$\omega_i^{\mathrm{g}} = \sum_{j=1}^{N_{\mathrm{r}}} \nu_{ij} q_j\,, \tag{37.39}$$

where q_j is the rate-of-progress variable for the jth reaction, which can be expressed as

$$q_j = (k_{\mathrm{f}})_j \prod_{i=1}^{N_{\mathrm{sp}}} c_i^{a'_{ij}} - (k_{\mathrm{r}})_j \prod_{i=1}^{N_{\mathrm{sp}}} c_i^{a''_{ij}}\,, \tag{37.40}$$

where $(k_{\mathrm{f}})_j$ and $(k_{\mathrm{r}})_j$ are temperature-dependent forward and backward rate coefficients, c_i is the concentration of species i, and a'_{ij} and a''_{ij} are constants. For elementary reactions which obey the mass action law,

Table 37.7 An example of gas species, adsorbed species, and bulk species

Surface reactions	Gas species	Adsorbed species	Deposited species
$NH_3(g) \rightleftharpoons NH_{3\,ads}$	NH_3	$NH_{3\,ads}$	None
$GaI(g) + NH_{3\,ads} \rightleftharpoons GaN(s) + HI(g) + H_2$	GaI, HI, H_2	$NH_{3\,ads}$	GaN
$AlCl_3(g) + NH_3(g) \rightleftharpoons AlN(s) + 3HCl(g)$	$AlCl_3$, NH_3, HCl	None	AlN

$a'_{ij} = \nu'_{ij}$ and $a''_{ij} = \nu''_{ij}$, the rate coefficients are assumed to have the Arrhenius form of

$$k_f = AT^n \left(\frac{p}{p_{atm}}\right)^m e^{\frac{-E_{af}}{RT}} , \tag{37.41}$$

$$k_r = AT^n \left(\frac{p}{p_{atm}}\right)^m e^{\frac{-E_{ar}}{RT}} , \tag{37.42}$$

where A is pre-exponential constant, n indicates the temperature dependence, m is the exponent of the pressure dependency, E_{af} is the activation energy for the forward reaction, and E_{ar} is the activation energy for the backward reaction. All these constants are provided by experiment results of kinetic reaction analysis. For a typical reaction of the form

$$\alpha A + \beta B \rightarrow C + D , \tag{37.43}$$

the rate of the gas-phase reaction is expressed as

$$\dot{\omega}_g = A_p T^n \exp\left(\frac{-E_a}{RT}\right)\left(\frac{p}{p_{atm}}\right)^m \times C_A^{\alpha} C_B^{\beta} \quad [\mathrm{kmol/m^3 s}] , \tag{37.44}$$

where C_A and C_B are the molar concentrations of reactants A and B, and α and β are the concentration exponents of A and B, respectively.

Kinetic Modeling of Surface Reactions. The general form of the surface reaction is

$$\sum_{i=1}^{N_g} a'_{ij} A_i + \sum_{i=1}^{N_s} b'_{ij} B_i(s) + \sum_{i=1}^{N_b} c'_{ij} C_i(b) = \sum_{i=1}^{N_g} a''_{ij} A_i + \sum_{i=1}^{N_s} b''_{ij} B_i(s) + \sum_{i=1}^{N_b} c''_{ij} C_i(b) , \tag{37.45}$$

where a_{ij} is the gas species stoichiometric, b_{ij} is the adsorbed species stoichiometric, c_{ij} is the bulk species stoichiometric, N_g is the total number of gas-phase species, N_s is the total number of adsorbed species, and N_b is the total number of deposited species. For example, in Table 37.7 the gas species, adsorbed species, and bulk species are listed for each reaction.

The surface reaction rate corresponding to reaction (37.43) may be expressed as

$$\dot{\omega}_{sj} = k_{fj} \prod_{i=1}^{N_g} C_{aiw}^{a'_{ij}} \prod_{i=1}^{N_s} C_{bi(s)}^{b'_{ij}} - k_{rj} \prod_{i=1}^{N_g} C_{aiw}^{a''_{ij}} \prod_{i=1}^{N_s} C_{bi(s)}^{b''_{ij}} \quad [\mathrm{kmol/m^2 s}] , \tag{37.46}$$

where k_{fj} and k_{rj} are the forward and backward reaction rates, C_{aiw} is the gas-phase concentrations of species i at the surface and can be expressed as $C_{aiw} = \rho_w Y_i^w / M_i$, where ρ_w is the gas-phase mass density, Y_i^w is the gas-phase mass fraction of species i near the wall, and M_i is the molecular weight of gas species i. $C_{bi(s)}$ is the surface concentration of adsorbed species i and can be expressed as $C_{bi(s)} = \rho_s X_i$, where ρ_s is the surface site density and X_i is the surface site fraction of adsorbed species i. It is seen from (37.46) that the concentration dependence of bulk species is neglected.

Two different approaches, the sticking probability (or sticking coefficient) method and finite rate method, are usually used to estimate the surface reaction rate in numerical simulations. The sticking probability method calculates the reaction rate based on sticking probability and precursor thermal flux, while the finite rate chemistry uses the kinetic expression to evaluate the reaction rate. When using the sticking probability method, the surface rate becomes

$$\dot{\omega}_{sj} = \gamma_j J_A \prod_{i=1}^{N_s} (X_i)^{b'_{ij}} , \tag{37.47}$$

where the sticking probability γ_j is described in Arrhenius form as

$$\gamma_j = f(\theta^s)_j \exp\left(\frac{-E_{aj}(\theta^s)}{RT}\right) , \tag{37.48}$$

where $f(\theta^s)_j$ is the function of the existing surface coverage of adsorbed species j, θ^s is the Langmuir definition of surface coverage of certain species or the fraction of sites that are occupied on the substrate surface, and E_{aj} is the activation energy for adsorption of species j.

The thermal flux of precursor A, or the incident rate of precursor onto the substrate (Hertz–Knudsen law) can be expressed as

$$J_A = \left(\frac{RT_w}{2\pi M_A}\right)^{1/2} C_{Aw} , \tag{37.49}$$

where C_{Aw} is the concentration of precursor A near the wall.

For the finite rate method, the surface reaction rate is the same as (37.47). For example, the reaction rate using a finite rate method can be expressed as

$$\dot{\omega}_s = A_s T^n \exp\left(\frac{-E_{as}}{RT}\right)\left(\frac{p}{p_{atm}}\right)^m C_{Aw}^{\alpha} C_{B(s)}^{\beta} , \tag{37.50}$$

for the reaction $\alpha A(g) + \beta B_{ads} \rightarrow C(s) + D(g)$.

37.4 Modeling of GaN IVPE Growth – A Case Study

In this section, a case study is presented to demonstrate optimization of GaN IVPE growth by computer modeling. In this example, the growth surface evolution mechanism will be presented and thermodynamic and kinetic methods will be used to predict gas-phase and surface reactions.

37.4.1 Scaling Analysis

The important dimensionless groups, their physical definitions, and calculated values in the vapor-growth reactor are listed in Table 37.8.

The values of ρ_0, μ_0, k_0, C_{p0}, and D_0 obtained from simulation results are 6.0×10^{-2} kg/m^3, 4.8×10^{-5} kg s/m, 1.3×10^{-1} J s/m K, 1740.0 J K/kg, and 1.0×10^{-3} m^2/s, respectively. L_c is selected as the diameter of the furnace reactor, which is 0.045 m, and V_0 is estimated as 0.67 m/s based on a total flow rate of 3.0 slm and a reactor pressure of 200 Torr.

According to Table 37.8, the flow is mainly laminar in the reactor; heat and mass transfer due to convection is much more important than conduction/diffusion. The ratio of the Grashof number Gr to the Reynolds number squared Re^2 is used to examine whether heat transfer is mainly controlled by natural convection, forced convection, or both. For the case presented here Gr/Re^2 = 0.66, indicating that both forced convection and natural convection are important.

Table 37.8 Important dimensionless groups in the IVPE reactor

Dimensionless group	Value
Reynolds (Re)	37.7
Grashof (Gr)	940.0
Prandtl (Pr)	0.64
Schmidt (Sc)	0.73
Thermal Péclet (Pe^T)	24.2
Mass Péclet (Pe^M)	27.4
Gas-phase Damköhler (Da^g)	–
Surface Damköhler (Da^s)	–

37.4.2 Computational Issues

To better understand the heat and mass transfer inside the system, numerical simulations are conducted by solving governing equations together with gas-phase and surface reactions. A schematic of the system and numerical grids is shown in Fig. 37.9.

Since the growth cell is positioned in a uniform-temperature region, a constant temperature was as-

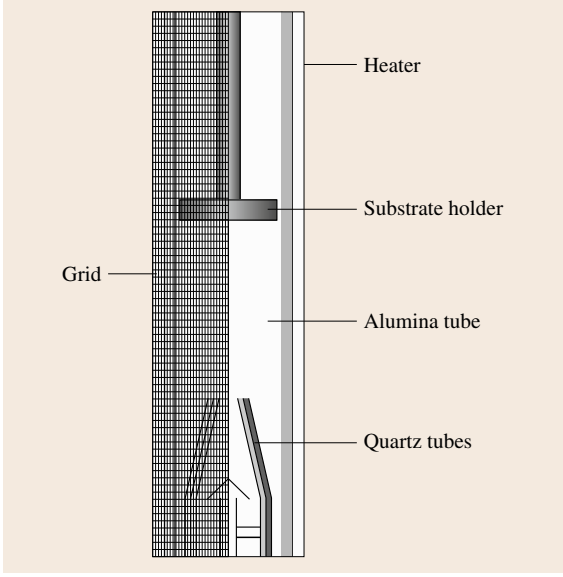


Fig. 37.9 Schematic of the calculation domain and grids for the system designed by *Mecouch* et al. [37.30, 72, 120]

Table 37.9

Wall on the graphite heater:	$\boldsymbol{V}=0$, $T=1323\,\mathrm{K}$, $\varepsilon=1.0$
Wall on the substrate holder:	$\boldsymbol{V}=0$, $\varepsilon=0.65$
Top surface of the gallium source:	$\dot{w}_{\mathrm{gallium}}=\dot{w}_1$, $T=700\,\mathrm{K}$, $\varepsilon=0.2$;
Wall on the surface of the substrate:	$\boldsymbol{V}=0, J_i=M_i\dot{w}_i^{\mathrm{s}}$, $\varepsilon=0.65$
Wall on the alumina tube:	$\boldsymbol{V}=0$, $\varepsilon=0.7$
Wall on the quartz surface:	$\boldsymbol{V}=0$, $\varepsilon=0.14$
Inlet carrier gas (NH_3+N_2):	$\boldsymbol{V}=\boldsymbol{V}_1$, $T=700\,\mathrm{K}$, $P_{\mathrm{in}}=P_{\mathrm{ref}}$, $Y_{NH_3}=Y_1$, $Y_{N_2}=Y_2$, $\varepsilon=0.15$
Inlet I_2, N_2, and H_2:	$\boldsymbol{V}=\boldsymbol{V}_2$, $T=700\,\mathrm{K}$, $P_{\mathrm{in}}=P_{\mathrm{ref}}$, $Y_{I_2}=Y_3$, $Y_{N_2}=Y_4$, $Y_{H_2}=Y_5$, $\varepsilon=0.15$
Shield gas at the inlet (N_2):	$\boldsymbol{V}=\boldsymbol{V}_3$, $T=700\,\mathrm{K}$, $P_{\mathrm{in}}=P_{\mathrm{ref}}$, $Y_{N_2}=Y_6$, $\varepsilon=0.15$
Outlet:	$T=300\,\mathrm{K}$, $P_{\mathrm{out}}=P_{\mathrm{ref}}$, $\varepsilon=1.0$

signed to the surface of the furnace wall with emissivity of unity. The experimentally measured weight losses of iodine and Ga sources are converted to mass flow rates that are used as inlet gas flow rates. The surface reaction rate $\dot{w}_i^{\mathrm{s}}$ after being converted to the mass flux J_i can be incorporated into the simulation as boundary conditions on the substrate surface.

General boundary conditions are prescribed as in Table 37.9.

37.4.3 Gas-Phase and Surface Reactions Analysis

Gas-Phase Reaction Analysis

In the GaN epitaxy growth system, the carrier gas mixture of N_2 and H_2 enters the I_2 source bubbler, where liquid I_2, heated by a jacket heater, vaporizes and will be transported to the reactor chamber by the carrier gas. H_2 and I_2 react to form HI. The gas mixture flowing over the gallium source surface includes N_2, H_2, I_2, and HI. Both I_2 and HI may react with Ga to form GaI_x, with $x=1$ or 3. From [37.26], it is deduced that GaI is the main product of the reactions between gallium and iodine gas in the reactor chamber being studied. It is found that GaI_3 is more stable thermodynamically at the operational conditions of $T=1323\,\mathrm{K}$ and $P=200\,\mathrm{Torr}$. On the substrate surface, NH_3 will be adsorbed and cracked to form activated nitrogen atoms that will react with GaI_x to grow GaN [37.121]. It is important to control the concentrations of NH_3 and GaI_x achieved on the substrate surface in experiments. Gas-phase species reaching the substrate area include NH_3, HI, H_2, N_2, GaI, GaI_3, and maybe I_2 and Ga. The gas pathways can be found in Fig. 37.1a.

Due to the high-temperature environment in the growth reactor ammonia may decompose to nitrogen and hydrogen gases before it reaches the substrate area. This will reduce the amount of ammonia reaching the substrate. The global equation for the ammonia gas phase decomposition can be expressed as

$$\mathrm{NH_3(g)} \rightleftharpoons \tfrac{1}{2}\mathrm{N_2(g)} + \tfrac{3}{2}\mathrm{H_2(g)}\,. \tag{37.51}$$

This reaction is a mildly endothermic process ($\Delta H=46\,\mathrm{kJ/mol}$). Thermodynamically, almost all NH_3 is decomposed into N_2 and H_2 at temperature higher than 300 °C. In the actual experiments, the NH_3 is decomposed slowly and the decomposition rate strongly depends on the growth conditions and the equipment. This reaction is, in fact, far from the kinetic equilibrium state.

The Damköhler number of the gas phase is calculated for the ammonia gas-phase decomposition reaction to determine its importance under the aforementioned operational conditions. The physical interpretation of the gas-phase Damköhler number is the ratio of the gas-phase reaction time to the gas flow residence time. The gas flow residence time can be estimated from $t_{\mathrm{F}}=L_{\mathrm{F}}/\bar{U}_{\mathrm{av}}$, where L_{F} is the distance between the gas inlet and the substrate, which is about 0.16 m. The averaged velocity is about 0.67 m/s with a typical flow rate of 3.0 slm under a pressure of 200 Torr and a furnace wall temperature of 1323 K. The fluid flow residence time is about 0.24 s. The comprehensive mechanism for NH_3 pyrolysis can be found in [37.122]. The controlling step of the reactions is found to be

$$\mathrm{NH_3 + M} \rightleftharpoons \mathrm{H_2 + H + M}\,, \tag{37.52}$$

where M is the third-party molecule that serves as a catalyst. Reaction (37.52) is a second-order reaction with a reaction coefficient of $k_{\mathrm{r}}=A\exp(-E_{\mathrm{a}}/\mathrm{RT})$ and a reaction rate of $\dot{w}_{\mathrm{g}}=k_{\mathrm{r}}[\mathrm{NH_3}][\mathrm{M}]$, where the pre-exponential coefficient (molecular collision frequency) A is 2.2×10^{10}, and the reaction activation energy E_{a} is $3.93\times10^5\,\mathrm{J/mol}$ [37.122]. $k_{\mathrm{r}}=2.55\times10^{-5}\,\mathrm{m^3/(mol\,s)}$ is calculated with the reference temperature of 1323 K. The third-party molecular concentration at temperature of 1323 K and pressure of 200 Torr is $1.17\,\mathrm{mol/m^3}$, if the molar ratio of the mixing gas of NH_3 and N_2 is assumed to be 1 : 1. The reaction rate is calculated as $r_0^{\mathrm{g}}=k_{\mathrm{r}}[\mathrm{M}]=2.98\times10^{-5}\,\mathrm{s^{-1}}$. The gas-phase Damköhler number (Da) for ammonia pyrolysis is 7.12×10^{-6},

which is much smaller than 1. Based on this analysis, it can be concluded that ammonia gas-phase decomposition can be neglected under the discussed operational conditions. It should be noted that ammonia also tents to decompose on reactive solid surfaces. The heterogeneous reaction rate under the current operating conditions is, however, expected to be very low according to [37.9].

As well as the substrate surface, ammonia may also react with GaI_x in the gas phase. To reduce gas-phase reactions between NH_3 and GaI_x before their mixture reaches the substrate, shield gas N_2 runs between the inner and outer silica nozzle to prevent mixing in the area above the silica nozzle (Fig. 37.1a) in order to increase the NH_3 and GaI_x concentrations on the substrate surface. The gas-phase reactions of NH_3 and GaI_x are expected to be weak and can therefore be neglected due to the effect of the shield gas and short residence time of the reacting gases.

If the equilibrium vapor pressure of iodine has been reached in the iodine source bubbler, the molar concentration fraction of iodine $[I_2]$ can be written as

$$[I_2] = \frac{P^0_{I2}}{P_{total}} \,. \tag{37.53}$$

The total mass reduction rate of I_2 in the source can be expressed as

$$\dot{m}_{I_2} = \rho_{I_2} Q_{total} \frac{P^0_{I_2}}{P_{total}} \,, \tag{37.54}$$

where Q_{total} is the total flow rate of the carrier gas.

The gas-phase reaction in the iodine source bubbler has been given in reaction (37.25). No reaction between N_2 and other gases are expected. Before reaction (37.25) starts, the initial partial pressures of H_2, I_2, HI, and N_2 are $P^0_{H_2}$, $P^0_{I_2}$, P^0_{HI}, and $P^0_{N_2}$, respectively, where the initial partial pressure of HI obviously equals zero. When the reaction equilibrium is reached in the source bubbler, the equilibrium partial pressure of the species can be expressed as $P^e_{H_2}$, $P^e_{I_2}$, P^e_{HI}, and $P^e_{N_2}$. The total pressure inside the bubbler is maintained at 1 atm and remains unchanged since the partial pressure reduction of H_2 and I_2 due to reaction equals the partial pressure production of HI in reaction (37.25). The partial pressure of N_2 will not change during reaction. Controlled by the mass flow controller, the molar ratio of H_2 and N_2 in the carrier gas is 9 : 91, and it is assumed to be the same in the bubbler. Therefore, at time zero, the follow equations can be obtained:

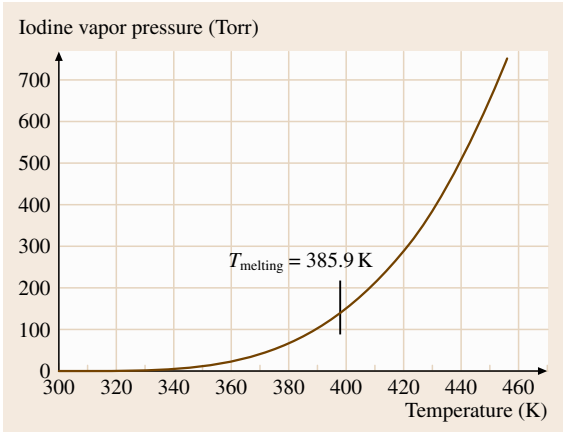


Fig. 37.10 Temperature dependence of iodine vapor pressure

$$\sum P^0_i = 1 = P^0_{H_2} + P^0_{I_2} + P^0_{N_2} \,, \tag{37.55}$$

$$\frac{P^0_{H_2}}{P^0_{N_2}} = \frac{9}{91} \,, \tag{37.56}$$

$$\frac{P^0_{I_2}}{\sum P^0_i} = [I_2] \,. \tag{37.57}$$

The initial vapor pressure of iodine over the source surface $P^0_{I_2}$ can be obtained using Fig. 37.10 [37.123].

The equilibrium expression of the reaction (37.25) and two additional equations obtained from reaction (37.25) are provided in (37.26–37.28).

Given the initial partial pressure of iodine vapor (or the source bubbler temperature), the reaction equilibrium pressure of each species in reaction (37.25) can be solved from (37.26–37.28) and (37.56–37.64).

Figure 37.11a shows the equilibrium partial pressures of different species at the inlet of the reactor under different iodine concentrations in the source. The highest iodine vapor pressure that can be achieved in the source bubbler is limited by the temperature of the valve connected to the bubbler. Gas temperature cannot exceed 150 °C for a regular needle valve, so the highest I_2 vapor pressure is about 300 Torr in the bubbler. When the equilibrium for reaction (37.25) is established before the mixture flows over the gallium source, the HI gas concentration will be around 0.1. When the I_2 concentration in the source is larger than 0.3, I_2 will be the main species to react with gallium, through GaI_x formation.

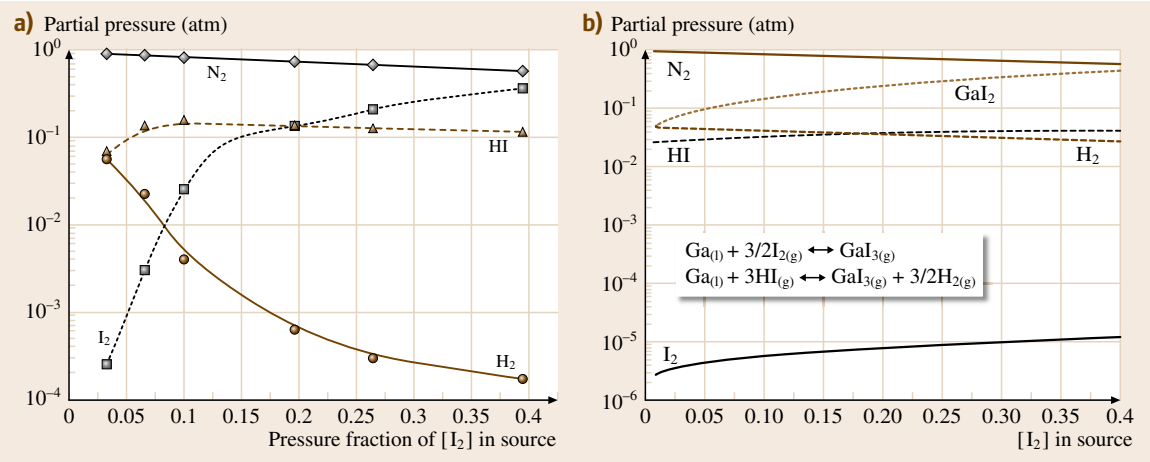


Fig. 37.11 **(a)** Equilibrium partial pressures with [I_2] in the source, and **(b)** equilibrium partial pressures above the liquid gallium at 1050 °C

The gas mixture coming out of the iodine source bubbler enters the reactor through the inner silica tube and reacts with gallium at the operating pressure of 200 Torr and growth temperature of 1050 °C. Instead of reacting with Ga vapor, both I_2 and HI will mainly react with gallium on its liquid surface, due to the very low gallium vapor pressure of 0.01 Torr under the given operating conditions. The reactions between I_2 and gallium can be written as

$$Ga_{(l)} + \tfrac{1}{2}I_{2(g)} \rightleftharpoons GaI_{(g)} \,, \tag{37.58}$$

$$Ga_{(l)} + \tfrac{3}{2}I_{2(g)} \rightleftharpoons GaI_{3(g)} \,, \tag{37.59}$$

$$GaI_{(l)} + I_{2(g)} \rightleftharpoons GaI_{3(g)} \,. \tag{37.60}$$

The reactions between gallium and HI can be written as

$$Ga_{(l)} + HI_{(g)} \rightleftharpoons GaI_{(g)} + \tfrac{1}{2}H_{2(g)} \,, \tag{37.61}$$

$$Ga_{(l)} + 3HI_{(g)} \rightleftharpoons GaI_{3(g)} + \tfrac{3}{2}H_{2(g)} \,, \tag{37.62}$$

$$GaI_{(g)} + 2HI_{(g)} \rightleftharpoons GaI_{3(g)} + H_{2(g)} \,. \tag{37.63}$$

At a typical growth temperature of 1050 °C, the calculated reaction equilibrium constants for reactions (37.58–37.60) are 8900, 840 000, and 50, respectively. The equilibrium reaction constants are 3400, 13 000, and 4 for reactions (37.61–37.63), respectively. The formation of GaI_3 in both cases is strongly favored thermodynamically. It is, therefore, assumed that only GaI_3 will be formed at equilibrium. A group of equations similar to the formation of HI can be derived and the equilibrium partial pressure of species above the gallium source calculated.

Figure 37.11b shows the calculated partial pressures on the gallium source. Comparing Fig. 37.11a with Fig. 37.11b, it is revealed that the partial pressure of GaI_3 is seen to track the initial partial pressure of iodine.

The favorability of reactions (37.59) and (37.62) indicates that any GaI formed will further react either with I_2 or HI to form GaI_3 under equilibrium conditions. However, the thermodynamic calculation only provides an upper limit for what to expect. Given the small surface area of the gallium source (1.1×10^{-4} m^2) and the relatively high velocity of the carrier gas over the gallium source (on the order of 0.1 m/s), it is unlikely that there is sufficient time for equilibrium condition to be established above the gallium source. The thermodynamic model describing transport of iodine and gallium species has to be compared with the experimental data to evaluate its accuracy.

Reported by *Mecouch* et al. [37.72,120], Fig. 37.12a shows the measured gallium loss rate compared with the measured iodine loss rate. The data points indicated by the diamond and square correspond to a carrier gas mixture of 9% H_2/91% N_2, whereas the triangle and circle correspond to carrier gas of pure hydrogen. The data were measured based on 20 h growth runs. Also two lines represent iodine transport as GaI (Ga : I_2 = 2 : 1) and GaI_3 (Ga : I_2 = 2 : 3), respectively. It is revealed that the measured iodine loss rate falls between the values calculated based on the two transport species. This indicates that there is enough iodine loss to account for GaI transport, but not enough iodine is lost to account for GaI_3 transport. It is concluded that the dominant trans-

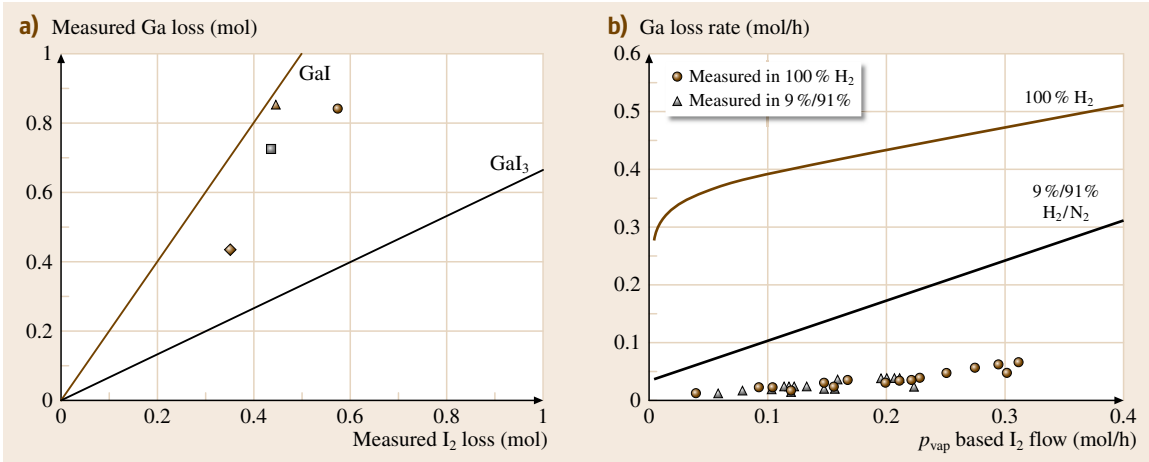


Fig. 37.12 **(a)** Measured gallium loss rate versus measured iodine loss rate and **(b)** gallium loss rate predicted from GaI_3-based thermodynamic equilibrium (*lines*) and measured during experiment growth runs (*bullets* and *triangles*) (after [37.72, 120])

port species is GaI, with some GaI_3 formed either on the Ga surface, or through subsequent reaction with the excess iodine species.

Figure 37.12b shows the predicted and measured rates of gallium loss versus the flow rates of I_2 [37.72, 120]. The solid lines represent the results obtained based on the thermodynamic equilibrium of the GaI_3 reaction, and diamonds and triangles represent the measured Ga loss with a pure H_2 carrier gas and a carrier gas mixture of 9% H_2/91% N_2, respectively. The experimental values fall well below the equilibrium prediction, and show no difference in measured loss rate of Ga between a carrier gas of pure H_2 and 9% H_2/91% N_2 mixture. This indicates that the source does not operate near the equilibrium as predicted by thermodynamics, and the reaction of HI formation does not significantly increase the transport of iodine. Thermodynamically calculated iodine loss rate is about 10–30 times greater than the measured iodine loss. Even if iodine is assumed to be transported from the source at the vapor equilibrium without any reaction to form HI, the calculated iodine flow rate is still 6–9 times higher than the observed loss rate from the iodine source. It is likely that the high flow rate of carrier gas and the crystal shape change in the iodine source as temperature fluctuates will both contribute to the difficulty in achieving equilibrium iodine vapor pressure in the source bubbler.

Based on the direct measurement of weight loss from both iodine and gallium sources after each experiment run, *Tassev* [37.26] reported that transport species in their reactor was GaI, with a slight excess of I_2. Furthermore, *Rolsten* [37.124] reported that GaI can be formed by heating either GaI_3 or the reaction product of gallium and iodine. With respect to these data and the relatively high flow rate of carrier gas and small volume for the reaction, it is assumed that the measured gallium loss was transported as GaI.

In the numerical model, only reaction (37.58) is therefore used to describe the gas-phase reaction for GaI_x formation. Measured iodine and gallium mass reduction rates will be converted to the flow rate of I_2 and Ga, respectively. The gas-phase reaction rate of (37.58) can be described as

$$\dot{\omega}_g = A_p T^n \exp\left(\frac{-E_a}{RT}\right)\left(\frac{p}{p_{atm}}\right)^m \times [A]^\alpha [B]^\beta \quad [\mathrm{kmol/m^3\,s}]\,, \tag{37.64}$$

where $[A]$ and $[B]$ are the molar concentrations of reactants Ga and I_2, and α and β are the concentration exponents of $[A]$ and $[B]$, respectively. $E_a/R = 0\,\mathrm{K}$ and $n = m = 0$ are assumed in the simulation. The value of A_p is used to determine the reaction rate and it is estimated as 4×10^8, corresponding to more than 95% conversion of experimentally weighed Ga loss to GaI. $\alpha = 1$ and $\beta = 0.5$ are assigned, corresponding to the stoichiometric coefficients of the reactants.

Surface Reaction Analysis

The overall surface reaction rate depends on the partial pressures of gas species such as GaI and NH_3, available free sites on the surface, surface concentrations of adsorbed species, surface diffusion coefficients, rate

constants of individual reaction step, and surface characteristics. Unlike the gas-phase reactions, predicting reaction paths and rate constants is more difficult for the heterogeneous surface reactions since the interactions between gas-phase and surface entities are more complicated than those between gas-phase molecules. A semi-empirical approach is usually used to simulate the surface mechanisms and kinetics.

Tassev et al. [37.26] discussed surface reactions from GaI/NH_3. Based on the byproducts detected in the GaN reactor, the rate-limiting step for the growth of GaN is given as

$$\mathrm{GaI(g)} + 2\mathrm{NH_3(g)} \rightleftharpoons \mathrm{GaN(s)} + \mathrm{NH_4I(g)} + \mathrm{H_2(g)}\,. \tag{37.65}$$

However, NH_4I detected in the cold reaction zone might be formed from NH_3 and HI at a temperature lower than 800 K since NH_4I decomposes above 800 K [37.118]. In the growth zone from silica nozzle to the substrate, the gas temperature will be close to 1323 K. It is therefore unlikely that NH_4I will be present. Considering different product gases which are stable at a temperature around 1323 K, the following analogs of element reactions to GaN growth from GaCl/NH_3 mixture in a halide vapor-phase epitaxy system [37.121] can be expressed:

$$V + \mathrm{NH_3} \rightleftharpoons \mathrm{NH_{3\,ads}}\,, \tag{37.66}$$
$$\mathrm{NH_{3\,ads}} \rightleftharpoons \mathrm{N} + \tfrac{3}{2}\mathrm{H_{2(g)}}\,, \tag{37.67}$$
$$\mathrm{N} + \mathrm{GaI} \rightleftharpoons \mathrm{NGaI}\,, \tag{37.68}$$
$$2\mathrm{NGaI} + \mathrm{H_{2(g)}} \rightleftharpoons 2\mathrm{NGa-IH}\,, \tag{37.69}$$
$$\mathrm{NGa-IH} \rightleftharpoons \mathrm{NGa} + \mathrm{HI_{(g)}}\,, \tag{37.70}$$
$$2\mathrm{NGaI} + \mathrm{GaI_{(g)}} \rightleftharpoons 2\mathrm{NGa-GaI_3}\,, \tag{37.71}$$
$$2\mathrm{NGa-GaI_3} \rightleftharpoons 2\mathrm{NGa} + \mathrm{GaI_3}\,, \tag{37.72}$$

where V is the vacant surface site of ammonia adsorption.

Since a very large NH_3 partial pressure is maintained over the substrate in the experiments, reaction (37.66) is close to thermodynamic equilibrium, which means that the $NH_{3\,ads}$ concentration on the substrate will be close to constant. The GaN deposition rate is mostly limited by reaction (37.68). This is supported by the experimental results that show the independence of GaN growth rate from the NH_3 partial pressure, but strong dependence on the GaI concentration. Since it is difficult to obtain the reaction constant for each elemental step, simplified overall surface reactions will be used in the simulations. Two overall reactions corresponding to the above element reactions are obtained as follows, corresponding to the most energetically favorable reactions for the iodine vapor-growth system:

$$\mathrm{GaI_{(g)}} + \mathrm{NH_{3(g)}} \rightleftharpoons \mathrm{GaN_{(s)}} + \mathrm{HI_{(g)}} + \mathrm{H_{2(g)}}\,, \tag{37.73}$$
$$3\mathrm{GaI_{(g)}} + 2\mathrm{NH_{3(g)}} \rightleftharpoons 2\mathrm{GaN_{(s)}} + \mathrm{GaI_{3(g)}} + 3\mathrm{H_{2(g)}}\,. \tag{37.74}$$

The following reaction might also be energetically favorable since GaI_3 is thermodynamically preferred

$$\mathrm{GaI_{3(g)}} + \mathrm{NH_{3(g)}} \rightleftharpoons \mathrm{GaN_{(s)}} + 3\mathrm{HI_{(g)}}\,. \tag{37.75}$$

Reaction (37.75) will also be considered in our numerical simulation model, though the thermodynamics analysis shows that the contribution of reaction (37.75) to the GaN deposition rate will be less than 2%.

Figure 37.13 shows the free energy of reaction for the above three surface reactions on the substrate surface. Since GaI is assumed to be the dominant species for gallium transport, reactions (37.73) and (37.74) are expected to be more important than reaction (37.75). It is revealed that the reaction free energies of the above three reactions are all positive at a typical growth temperature of 1323 K, which means that the equilibrium reaction constants are all smaller than 1, and none of the reactions are spontaneous. An appropriate effective supersaturation, however, can still be achieved by controlling the reactants' partial pressure on the substrate, which serves as the driving force for GaN deposition. Supersaturation of a typical vapor-growth system is

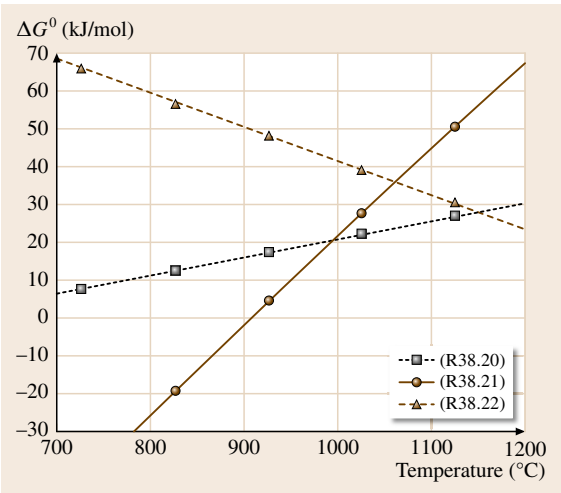


Fig. 37.13 Free energy of reaction for GaN surface deposition

about 5. For reaction (37.73), it can be expressed as

$$\sigma = \ln\left(\frac{P_{\mathrm{GaI}} P_{\mathrm{NH_3}}}{P_{\mathrm{HI}} P_{\mathrm{H_2}} K_{20}}\right) , \tag{37.76}$$

where K_{20} is the equilibrium constant. At a growth temperature of 1323 K, this can be calculated as

$$K_{20} = \frac{P^{\mathrm{e}}_{\mathrm{HI}} P^{\mathrm{e}}_{\mathrm{H_2}}}{P^{\mathrm{e}}_{\mathrm{GaI}} P^{\mathrm{e}}_{\mathrm{NH_3}}} = \exp\left(\frac{-\Delta G^0_{\mathrm{r}}}{\mathrm{RT}}\right) = 0.12 . \tag{37.77}$$

Since reactions (37.74) and (37.75) can be combined to form reaction (37.73), reaction (37.73) will therefore be used as the overall surface deposition step at first in the numerical model. The surface reaction rate for reaction (37.73) is determined by

$$\dot{\omega}_{\mathrm{s}} = \delta A_{\mathrm{p}} T^{\mathrm{n}} \exp\left(\frac{-E_{\mathrm{a}}}{\mathrm{RT}}\right)\left(\frac{p}{p_{\mathrm{atm}}}\right)^{\mathrm{m}} \times [A]^{\alpha} [B]^{\beta} \quad [\mathrm{kmol/m^2\,s}] , \tag{37.78}$$

where δ is the deposition coefficient of GaN on the substrate surface. The rate constant r^{s} for reaction (37.73) has not been reported in the literature. An activation energy of 103 000 J/mol is reported in reference to the data from *Shintani* et al. [37.125] for the following surface reaction of GaN HVPE growth:

$$\mathrm{GaCl_{(g)}} + \mathrm{NH_{3(g)}} \rightleftharpoons \mathrm{GaN_{(s)}} + \mathrm{HCl_{(g)}} + \mathrm{H_{2(g)}} . \tag{37.79}$$

The temperature range of their data is 860–1020 °C, which is similar to the IVPE growth temperature (1050–1100 °C). The activation energy of 1.03×10^5 J/mol is estimated for the reaction (37.73). $n = m = 0$ is assumed here. $\alpha = x$ and $\beta = 0$ are assigned for the concentration coefficients of GaI and NH_3. The value of α is set as undetermined since the concentration dependency of GaI should be 13 according to reactions (37.73–37.75). The ammonia concentration dependence is set to zero due to the fact that ammonia is always in excess on the substrate surface in the experiment. By matching the experimental data for the deposition rate with simulation results under different NH_3/GaI molar ratios, the value of δA_{p} and α can be determined using an optimization procedure, the detail of which will be covered in Sect. 37.4.8.

To determine the importance of reactions (37.73–37.75) in the GaN deposition rate, all three equations are included in the simulation. The contribution from individual reaction to the GaN growth rate will be determined under the assumption that all of the reactions are stoichiometric. No kinetic data for reactions (37.74–37.75) are available in the literature. *Shaw* et al. [37.126] reported that the activation energy of GaAs epitaxial growth is between 6.18×10^4 J/mol and 1.648×10^5 J/mol in the Ga–As–Cl_3 system due to surface adsorption, surface reaction, and surface diffusion. The activation energy for reaction (37.74) is therefore chosen as $E_{\mathrm{a}} = 1.133 \times 10^5$ J/mol, which is the medium value of the above energy range. For reaction (37.74), $n = m = 0$, $\alpha = 3$, and $\beta = 0$, and the value of δA_{p} is again determined by matching the simulation results with the experiment ones. $E_{\mathrm{a}} = 116\,396$ K is reported by *Dollet* et al. [37.98] for the following surface reaction of AlN epitaxial growth:

$$\mathrm{AlCl_{3(g)}} + \mathrm{NH_{3(g)}} \rightleftharpoons \mathrm{AlN_{(s)}} + 3\mathrm{HCl_{(g)}} , \tag{37.80}$$

which is used in the simulation as the activation energy for reaction (37.75). Also for reaction (37.75), we use $n = m = 0$, $\alpha = 1$, and $\beta = 0$, and $\delta A_{\mathrm{p}} = 0.36$, as reported by *Cai* [37.29] for the reaction (37.80) of AlN epitaxial growth with a temperature of 1100 °C and pressure of 760 Torr.

Finally the GaN deposition rate can be calculated as

$$G = \frac{\dot{\omega}_{\mathrm{s}} M_{\mathrm{A}}}{\rho_{\mathrm{A}}} \quad [\mathrm{m/s\ or}\ 3.6 \times 10^9\ \mu\mathrm{m/h}] , \tag{37.81}$$

where M_{A} and ρ_{A} are the molecular weight (84 kg/kmol) and density (6.15×10^3 kg/m^3) of GaN, respectively.

37.4.4 Geometrical and Operational Conditions Optimization

Many parameters pertinent to reactor geometry and mixed gas injection are important to optimal design of the vapor-growth reactor. In this section, the effects of geometrical configurations such as the diameter of the substrate holder and operating conditions such as the process and shield gas flow rate on mixing process, deposition rate distribution, and deposition uniformity on the substrate are studied.

37.4.5 Effect of Total Gas Flow Rate on Substrate Temperature

A uniform and sufficient high temperature on the substrate is required to achieve high growth rate, uniform film thickness, and good film quality. In the experiments conducted by *Mecouch* et al. [37.30, 72, 120], the temperature on the furnace wall is monitored using thermocouples. In situ observation of temperature achieved

Table 37.10 Inlet gas conditions for an experiment run by *Mecouch* et al. [37.30, 72, 120]

Inlet	Species	Initial volume flow rate (slm)	Adjusted volume flow rate (slm)
1	I_2	0.008	0.008
	H_2	0.026	0.026
	N_2	0.5	0.25
2	N_2	1.0	0.62
3	NH_3	2.0	1.0
	N_2	–	1.25

on the substrate surface, however, is difficult to realize. Temperature difference between the substrate surface and furnace wall, and the effect of total gas flow rate on the temperature achieved on the substrate surface will be investigated numerically. Gas species at each inlet and their flow rates used are listed in Table 37.10.

The initial inlet gas conditions are tested both experimentally and numerically. The measured gallium weight reduction rate of 0.0133 slm is used in the simulation. This flow rate corresponds to a molar ratio of $Ga : I_2 = 1.7 : 1$, which favors the formation of GaI in the gas phase. The temperature (1050 °C) of the heating unit and the reactor pressure (200 Torr) remain unchanged for all the simulations presented here. Gas-phase and surface reactions are not activated in the simulation unless stated, since they are expected to have insignificant influence on the gas flow and heat transfer in the reactor. Thermodynamic properties of different substances are listed in Table 37.11.

Figure 37.14 shows the streamline and temperature distributions in the simulated system. Figure 37.14a shows that a reverse flow is formed near the alumina tube wall, which means that the radial mixing of different species is enhanced. The gas-phase reactions of GaI/NH_3 in the center part and the GaN deposition on the alumina tube wall will therefore increase, and the GaN deposition rate will decrease on the substrate. In addition, the reverse flow might be unstable in the experiment, which could be larger or smaller

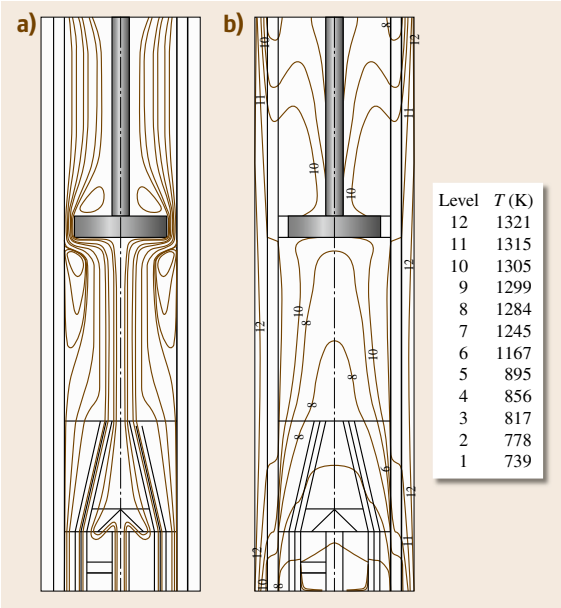


Fig. 37.14a,b Distributions of (a) streamlines and (b) temperature inside the VPE reactor designed by *Mecouch* et al. [37.30, 72, 120]

as time progresses. The GaN deposition quality on the substrate could therefore be worsened. For the above reasons, it is necessary to make sure that no big reverse flow is formed in the reactor. Figure 37.14b shows the temperature distribution for the simulated system. The temperature of the gas mixture will be heated to more than 1200 K above the silica nozzle. Attributed to such high temperature, chemical deposition may occur on the surfaces of the reactor wall, substrate, and substrate holder.

The velocity ratio of NH_3 from inlet 3, shield gas N_2 from inlet 2, and carrier gas from inlet 1 is 1 : 11 : 13 at the height of the silica nozzle outlet. It is concluded that the reverse flow is formed mainly due to a large gas velocity difference at the silica nozzle outlet. To prevent the reverse flow, adjusted inlet gas flow rates are

Table 37.11 Thermodynamic properties used in the simulation

Properties	All gases	Al_2O_3	GaN	Silica	BN
Specific heat (J/(kg K))	Mix JANAF method	900	490	710	1610
Density (kg/m^3)	Ideal gas law	3900	6150	2198	1900
Thermal conductivity (W/(m K))	Mix kinetic theory	30	130	1.38	28
Dynamic viscosity (kg/(m s))	Mix kinetic theory	–	–	–	–
Diffusivity	Sc = 0.72	–	–	–	–
Adsorption coefficient	0	1	1	0.145	1

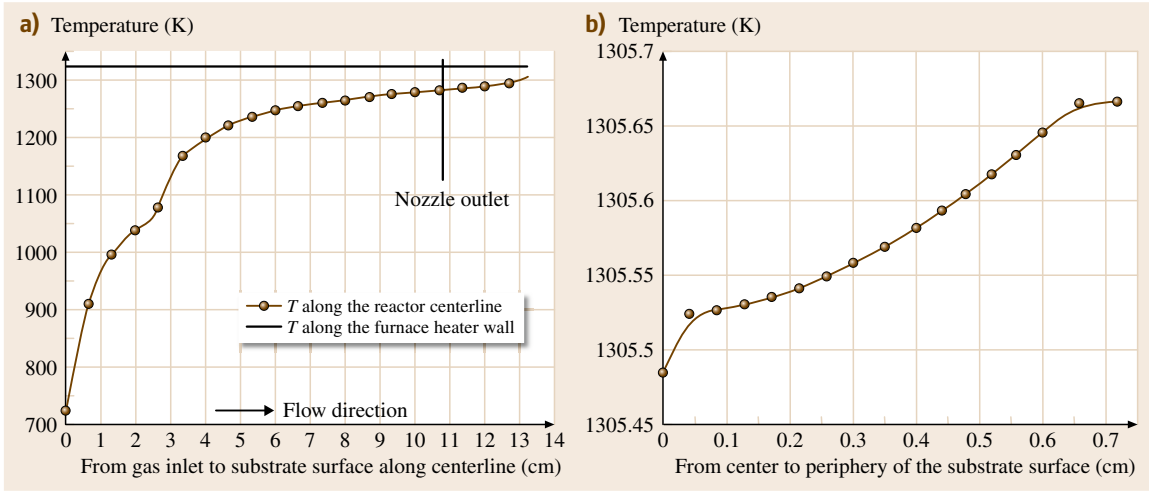


Fig. 37.15a,b Temperature distributions (a) along the reactor centerline and furnace heater wall and (b) from the periphery to the center on the substrate surface in the VPE reactor designed by *Mecouch* et al. [37.30, 72, 120]

used in the experiment (Table 37.10). The velocity ratio under the new inlet conditions is 1 : 6 : 6. Simulation results show that the reverse flow is eliminated under this condition.

Figure 37.15a shows the temperature distributions from inlet to the substrate surface along the reactor centerline and along the substrate surface, based on the adjusted inlet gas conditions. It is revealed that the mixing gases can be heated to about 1250 K when reaching the nozzle outlet, which further increases to 1300 K at the substrate area. Figure 37.15b shows the temperature distribution from the periphery to the center of the substrate surface. It is found that the temperature on the substrate surface will be high at the periphery but low in the middle. The temperature difference along the substrate surface is only about 0.18 °C, and the substrate temperature is about 17 °C lower than the temperature on the furnace heater.

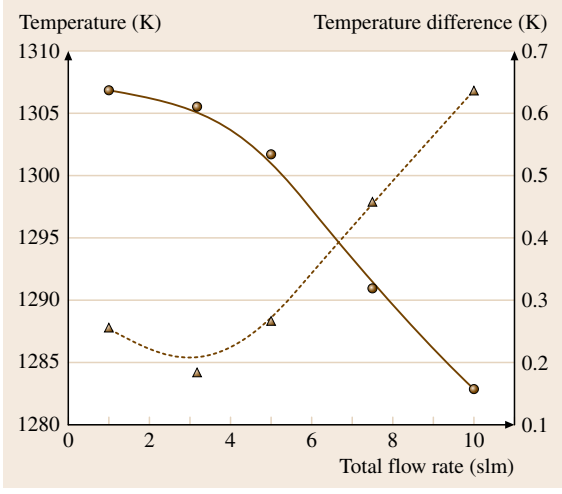


Fig. 37.16 Averaged temperature (*solid line*) and largest temperature difference (*dashed line*) on the substrate as functions of total flow rate

By reducing or increasing the flow rate at each gas inlet with the same ratio, the effect of the total flow rate on the temperature achieved on the substrate was investigated. Figure 37.16 shows the averaged temperature and the largest temperature difference achieved on the substrate with different total flow rates. It is revealed that the substrate temperature drops from 1307 °C to 1283 °C as the flow rate changes from 1 to 10 slm. Temperature nonuniformity on the substrate also changes with total flow rate. The largest temperature difference is defined as the highest minus the lowest temperature on the substrate. It is revealed that a total flow rate of 3 slm gives the lowest temperature difference, or the best uniformity, on the substrate surface. In the experiments, the total flow rate is controlled to around 3 slm [37.72, 120].

Effect of the Shield Gas Flow Rate

The shield gas, N_2, is used to prevent Ga from mixing with NH_3 directly in the area above the silica nozzle in order to reduce the gas-phase reactions of GaI and NH_3. Four shield gas flow rates of 0.5, 0.8, 1.2, and 1.5 slm were studied. The only reaction activated is (37.58).

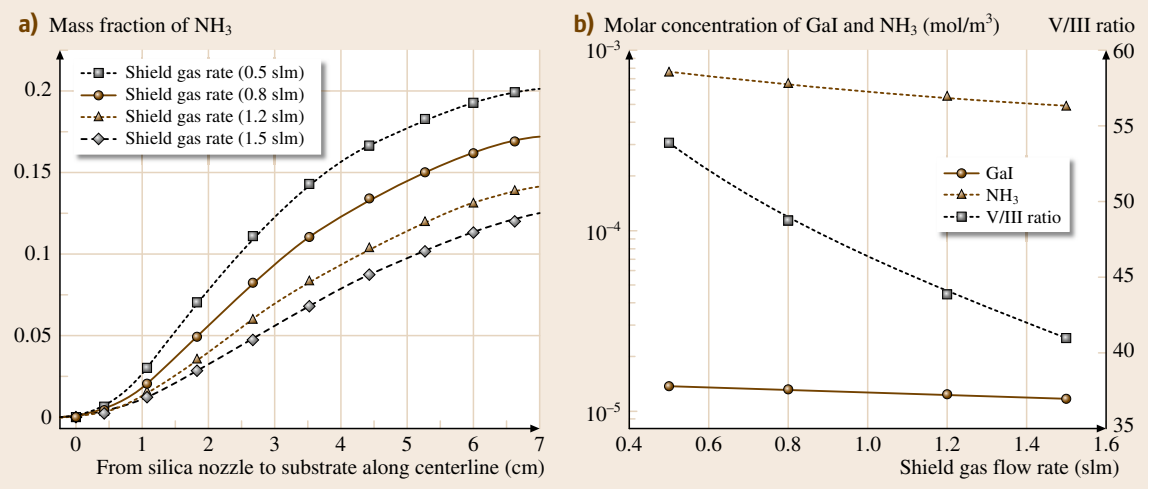


Fig. 37.17 **(a)** Mass fraction distribution of NH_3 along the centerline from the silica nozzle to the substrate and **(b)** averaged concentrations of GaI and H_3 on the substrate and V/III ratio achieved on the substrate as functions of the shield gas flow rate

Figure 37.17a shows the mass fraction change of NH_3 with the shield gas flow rate along the reactor centerline from the outlet of the silica nozzle to the substrate surface. It is seen that, at the silica nozzle outlet, the NH_3 concentration is zero, and then increases gradually with distance. The concentration of NH_3 with 1.5 slm shield gas flow rate is about half of that with 0.5 slm shield gas flow rate along the centerline. Figure 37.17b shows the averaged concentrations of GaI and NH_3 and the V/III ratio on the substrate surface. It is shown that the concentrations of *GaI* and NH_3 decrease as the shield gas flow rate increases. It is also observed that the effect of the shield gas flow rate on the NH_3 concentration is more significant than its effect on GaI concentration. When the shield gas flow rate increases from 0.5 to 1.5 slm, concentrations of GaI and NH_3 drop by 17% and 34%, respectively. This is attributed to the increase of the N_2 concentration on the substrate surface as the result of the increase of the shield gas flow rate. Furthermore, increasing the shield gas flow rate can reduce the residence time for NH_3 to diffuse into the center area, causing reduction of the NH_3 concentration and the V/III ratio.

Experiments were conducted to study the effect of the shield gas flow rate on the GaN deposition rate by *Mecouch* et al. [37.30]. Shield gas flow rates of 0.62 slm and 1.0 slm were used in their experiments. The flow rate of 0.62 slm corresponds to a velocity ratio of 1 : 1 between the shield gas and other gases at the outlet of the silica nozzle. It is revealed that the shield gas flow rate of 0.62 slm produces a higher GaN growth rate under different iodine concentrations in the source bubbler compared with the shield gas flow rate of 1.0 slm. A shield gas flow rate of 0.62 slm was therefore used in the subsequent experiments. Figure 37.18a,b shows the mass fraction distributions of GaI and NH_3 in the growth reactor. The gas-phase reaction between Ga and

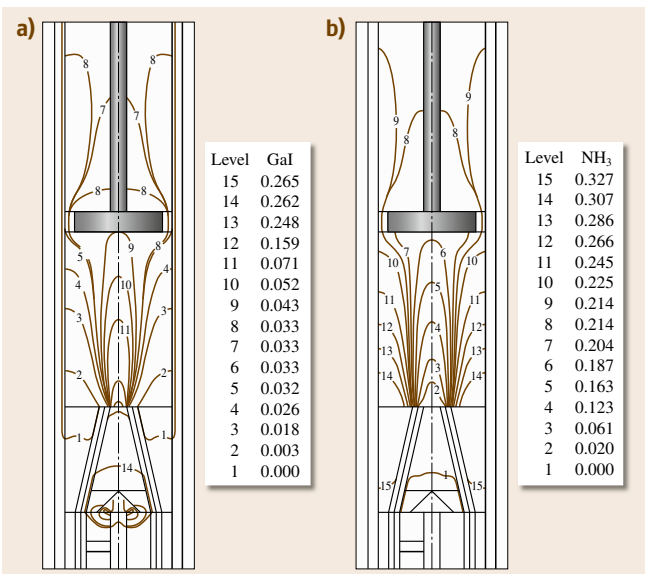


Fig. 37.18a,b Mass fraction distributions of **(a)** GaI and **(b)** NH_3 inside the VPE reactor designed by *Mecouch* et al. [37.30, 72, 120]

Table 37.12 Species concentrations on the substrate

	Ga	GaI	H_2	I_2	N_2	NH_3
Y_i	2.4×10^{-6}	0.042575	0.000855	0.005494	0.760977	0.190096
C_i (kmol/m^3)	2.16×10^{-9}	1.36×10^{-5}	2.69×10^{-5}	1.36×10^{-6}	1.71×10^{-3}	7.02×10^{-4}
P_i (Torr)	1.76×10^{-4}	$1.11\times10^{+0}$	$2.19\times10^{+0}$	1.11×10^{-1}	$1.39\times10^{+2}$	$5.73\times10^{+1}$

I_2 takes place mainly inside the inner silica nozzle. Since the gas flow in the reactor is laminar (Fig. 37.18b), the NH_3 has been transferred into the center area of the reactor by mass diffusion.

The mass fraction Y_i, the molar concentration C_i, and the partial pressure P_i of species i on the substrate are listed in Table 37.12 for the shield gas flow rate of 0.62 slm. It is seen that the partial pressure of Ga is close to zero, and the achieved partial pressure of GaI is around 1 Torr on the substrate at a reactor pressure of 200 Torr.

Effect of Silica Nozzle Angle

The GaN growth rate is directly proportional to the partial pressure of GaI achieved on the substrate. A silica nozzle is used in experiments to force more GaI to the substrate area [37.30]. Adding a silica nozzle will also help prevent ammonia from diffusing towards the gallium source area because ammonia reacts with Ga directly on the liquid gallium surface to form polycrystalline GaN through the following reaction:

$$2\mathrm{Ga(g)} + 2\mathrm{NH_3(g)} \rightleftharpoons 2\mathrm{GaN(s)} + 3\mathrm{H_2(g)}\,. \qquad (37.82)$$

This reaction (37.82) could terminate the transport of gallium and stop the GaN growth on the substrate.

Design of silica nozzle angle θ (Fig. 37.1a) is an important task. To test the effect of nozzle angle on concentration distributions of NH_3 and GaI between the outlet of silica nozzle and substrate surface, growth systems with four nozzle angles of 77°, 81°, 86°, and 90° were simulated. Figure 37.19a shows the mass fraction distribution of NH_3 along the reactor centerline from the silica nozzle outlet to the substrate surface. It is found that, at the silica nozzle outlet, the NH_3 concentration is close to 0 when using a nozzle with angle $< 81°$. The ammonia mass fraction is up to 0.03 at the outlet of the inner silica nozzle when its angle is $> 81°$, indicating that ammonia may diffuse into the inner silica tube. Using a nozzle with small angle, such as 77°, can significantly reduce the NH_3 concentration in the area above the silica nozzle. This influence will be weakened when the mixed gas approaches the substrate. Thus, the concentration of NH_3 on the substrate can be maintained at an appropriate level even with a small nozzle angle. Figure 37.19b shows the average molar concentrations of GaI and NH_3 on the substrate surface. The concentra-

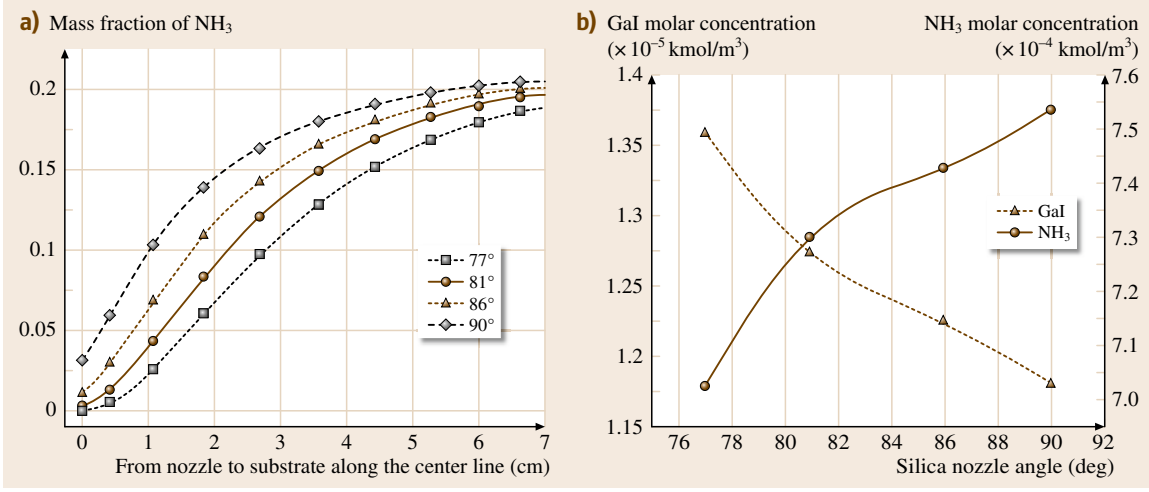


Fig. 37.19 (a) Mass fraction distribution of NH_3 along the centerline from the silica nozzle to the substrate and (b) averaged concentrations of GaI and NH_3 on the substrate under different silica nozzle angles

tion of NH_3 increases with the nozzle angle, while that of GaI decreases with the nozzle angle.

The silica nozzle used in the experiment has an angle of 77° [37.72]. This design can reduce the NH_3 concentration in the area above the silica nozzle outlet significantly, subsequently preventing NH_3 from diffusing into the inner silica nozzle and increasing the GaI concentration on the substrate. The simulation shows that using a nozzle angle of 77° can increase the GaI concentration on the substrate by about 10% compared with a silica nozzle angle of 90°.

Effect of Substrate–Nozzle Distance

The substrate–nozzle distance is one of the key parameters that affect the GaN growth rate and deposition quality. It controls the length and time allowed for different gases to mix with each other. An appropriate substrate–nozzle distance corresponds to a suitable V/III ratio above the substrate for fast and uniform GaN growth. The simulated substrate–nozzle distance ranges from 6.35 to 17.8 cm. Figure 37.20a shows the molar concentrations of GaI and NH_3 as a function of substrate–nozzle distance. The concentration of GaI decreases with substrate–nozzle distance until the mixing is completed at a substrate–nozzle distance of 13.0 cm. The error bars in Fig. 37.20a indicate the GaI concentration variation across the substrate surface. It is seen that, for a substrate–nozzle distance of 6.35 cm, the variation of GaI concentration on the substrate surface is small (around 0.6%). This indicates that species uniformity is not a problem. The concentration of NH_3 increases with the distance since more time is available for ammonia to diffuse into the center area of the reactor. The variation of the V/III ratio with substrate–nozzle distance is shown in Fig. 37.20b. When the substrate–nozzle distance is less than 13.0 cm, an increase of substrate–nozzle distance of about 2.5 cm corresponds to a V/III ratio increase of 10.

It is concluded that complete mixing of different species needs a substrate–nozzle distance larger than 13.0 cm. The species concentrations across the substrate surface is fairly uniform even with a small substrate–nozzle distance such as 6.35 cm. To achieve uniform GaN deposition with high growth rate, the substrate–nozzle distance of 6.35 cm is used. This distance can achieve uniform species concentrations across the substrate surface and a high GaI concentration of 1.4×10^{-5} kmol/m^3. The V/III ratio in this case will be about 50. For the growth of GaN from GaI_{1-3}/NH_3 or $GaCl_{1-3}/NH_3$, the concentration of NH_3 is maintained at a high value on the substrate, resulting in an appropriate supersaturation as the driving force for the surface deposition. For the current GaN vapor-growth system, the V/III ratio is kept above 50 for most experiments [37.30, 72, 120].

37.4.6 Effect of Substrate Rotation on Deposition Rate and Deposition Uniformity

In the experiment, the substrate is rotated by revolving the substrate holder to improve deposition uniformity,

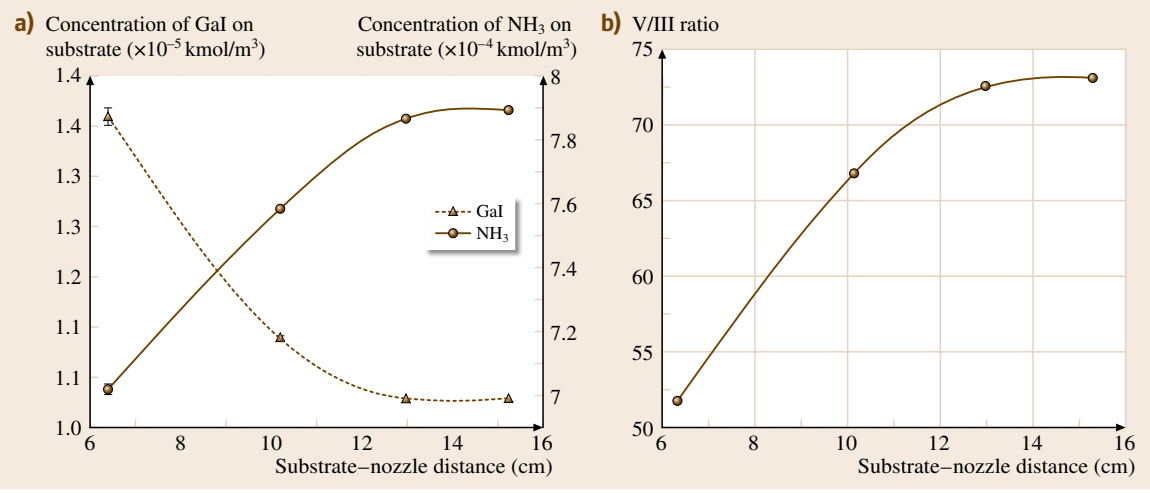


Fig. 37.20 (a) Averaged molar concentrations of GaI and NH_3 and (b) the V/III ratio achieved on the substrate under different substrate–nozzle distances

and consequently to improve deposition quality. It is shown by the simulation that temperature and species concentration distributions on the substrate are fairly uniform under the current geometrical and operational conditions. It is expected that the deposited GaN layer will be uniform. Experiments have been conducted using the gas inlet conditions listed in Table 37.10 for a growth time period of 1 h [37.30]. This experiment shows that the thickness of deposited GaN crystal layer varies from 35 to 75 μm/h on the substrate surface. To improve the deposition uniformity, the substrate was rotated in the experiments. The instantaneous GaN deposition rate distribution on the substrate will be averaged due to rotation and the deposition uniformity will be improved. Cross-sectional scanning electron microscopy (SEM) images of GaN layer with a substrate rotation speed of 1 rpm show that the deposition uniformity is significantly improved since the thickness difference in the GaN layer is reduced from 40 to 2 μm/h [37.30].

It should be noted that, in the simulation, since the entire system is assumed to be perfectly axisymmetric, the calculated temperature and species concentration distributions will be axisymmetric. In the experiment, the components in the growth reactor are mostly not axisymmetric. It is found experimentally that the deposition nonuniformity is mainly attributed to misalignment between the inner and outer silica nozzles and the seed holder. Deposition uniformity could therefore be improved by improving these alignments.

Another factor that could cause deposition nonuniformity across the substrate is reactant depletion in the species concentration boundary layer formed on the substrate due to surface reactions. Figure 37.8b shows that a typical boundary layer of the reactant species concentration is formed on the substrate due to surface reactions. The concentration of GaI before the mixing gas reaching the substrate area is defined as $C = C_0$. Due to the surface reactions (37.73–37.75) of GaI with NH_3 on the substrate, the GaI concentration drops to $C = C_w$ on the substrate surface. The concentration boundary layer thickness χ is defined as the distance where the GaI concentration changes from C_0 in the bulk gas to C_w on the substrate surface. The low-concentration layer will gradually accumulate as the gases flow from the center to the perimeter of the substrate. The growth rate is therefore expected to be low in the periphery area. To improve the deposition uniformity, one may reduce the concentration boundary-layer thickness. Substrate rotation can be used to reduce the thickness of the low-concentration layer. To estimate the rotation speed at which the boundary-layer thickness will be affected, the following equation is used to calculate the ratio of the Reynolds number due to forced convection and rotation:

$$\frac{\text{Forced convection}}{\text{Rotation}} = \frac{\mathrm{Re_d}}{\mathrm{Re}_\theta} = \frac{UD/v}{\omega d^2/v} = \frac{1500}{\varpi} , \tag{37.83}$$

where ω and ϖ are the rotation frequency in units of rps and rpm, respectively.

It is revealed that the rotation speed has to be on the order of 10^3 rpm for rotation to affect the boundary-layer thickness. Such a high rotation speed will cause the growth environment to become unstable near the substrate. In experiments, only a small rotation speed (1 rpm) is recommended in order to even out the instantaneous deposition rate on the periphery of the substrate.

37.4.7 Quasi-equilibrium Model for Deposition Rate Prediction

Predicting the GaN deposition rate is the most important issue for modeling. A thermodynamic surface reaction model (local model) combined with a numerical simulation model (global model) for mass transfer in the growth reactor are used to calculate the GaN deposition rates under different substrate temperatures. Since reactions (37.74–37.75) can be combined to form reaction (37.73), reaction (37.73) is assumed to prevail on the substrate when reaction equilibrium is reached.

Figure 37.8a shows a schematic of GaN deposition on the substrate surface. Species on the substrate include GaI, NH_3, HI, H_2, N_2, I_2, and Ga. From the global heat and mass transfer model, we can calculate the concentration of these seven species, which will be used as the initial species conditions. As shown in Fig. 37.8b, the transport of species into the concentration boundary layer is driven by the concentration gradient due to surface deposition.

According to (37.33–37.36), there are four unknowns, i. e., $P^{\mathrm{e}}_{\mathrm{HI}}$, $P^{\mathrm{e}}_{\mathrm{GaI}}$, $P^{\mathrm{e}}_{\mathrm{NH_3}}$, and $P^{\mathrm{e}}_{\mathrm{H_2}}$, and four equations. The partial pressures of each species, $P^{\mathrm{w}}_{\mathrm{HI}}$, $P^{\mathrm{w}}_{\mathrm{GaI}}$, $P^{\mathrm{w}}_{\mathrm{NH_3}}$, and $P^{\mathrm{w}}_{\mathrm{H_2}}$, can be obtained using the global simulation model, whereas the sticking probabilities of each species α_i are required. NH_3 will crack on the substrate surface to provide nitrogen atoms for GaN growth. The sticking coefficient of NH_3 on the substrate is defined as the ratio of the NH_3 flux incorporated as GaN to the total NH_3 flow rate incident on the substrate. This definition assumes that all NH_3 flux incident on the substrate is involved in the reaction process, i. e.,

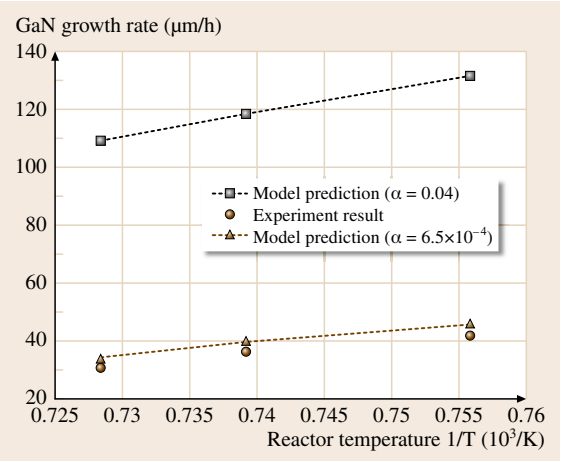


Fig. 37.21 Predicted and experimental results of GaN growth rate with ammonia sticking coefficient of 0.04 or 6.5×10^{-4} (after [37.72])

no saturation of surface sites occurs on the substrate. *Mesrine* et al. studied the efficiency of NH_3 as the nitrogen source for GaN molecular-beam epitaxy [37.127]. It was found that, at temperature of 830 °C and pressure of 10^{-5} Torr, the sticking coefficient of NH_3, α_{NH_3} is about 0.04. The sticking coefficient at temperature of 1050 °C is expected to be slightly higher, therefore 0.04 is used as the sticking coefficient of NH_3 in the simulation. The sticking coefficient of GaI is assumed to be unity. The sticking probabilities of H_2 and HI are also assumed to be unity, considering the high reactivity of these species under high growth temperature. In Fig. 37.21 the square line and the black dots indicate the experimental and predicted GaN deposition rates, respectively. The predicted results are 3–4 times the experiment ones. The difference between the experiments and simulations is attributed to overprediction of the sticking coefficient of ammonia in the modeling, which is due to several reasons. First, the assumption that all the incident ammonia flux on the substrate is involved in the reaction process is not correct for the current GaN growth system. To study the ammonia sticking coefficient in the MBE system, *Mesrine* et al. [37.127] controlled the ammonia flow rate low enough to ensure that the above assumption is valid. However, in the GaN growth system, the ammonia flow rate is high and it is found experimentally that the GaN deposition rate is independent of the ammonia flow rate, as discussed in the next section. The efficiency of the ammonia deposition coefficient will be lowered greatly in this case. Second, since a much higher pressure, 200 Torr, is used in the current system, the available surface sites for nitrogen atoms to deposit will be further reduced since it has to compete with other atoms on the substrate surface. The sticking coefficient will therefore be reduced further. Finally, since a much higher growth temperature is used, GaN decomposition on the substrate surface will be enhanced, causing the sticking coefficient of ammonia to drop further. By adjusting the ammonia sticking coefficient, it is found that, when the value is lowered to the order of 10^{-4}, the GaN growth rate from prediction and experiment matches very well, as shown in Fig. 37.21 for an ammonia sticking coefficient of 6.5×10^{-4}.

37.4.8 Kinetic Deposition Model

Combined with the experimental data, the GaN deposition rate at a constant growth temperature of 1050 °C was tested numerically while the flow rates at the inlet were changed to obtain different reactant concentrations and V/III ratios on the substrate. More importantly, simulations were performed to test the surface reaction pathway of GaN growth and the contribution of different reactions to the final GaN growth rate.

The energetically preferred surface reactions on the substrate for GaN deposition have been obtained previously. Reaction (37.73) was first used as the overall surface reaction step on the substrate in the simulation. Figure 37.22 shows the experimental data for the GaN deposition rate at different ammonia flow rates and iodine vapor fractions [37.30,72]. The highest growth rate

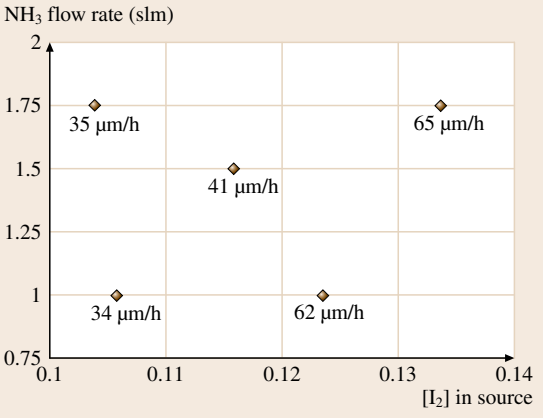


Fig. 37.22 Effect of ammonia flow rate and iodine vapor fraction in the source on the GaN deposition rate (after [37.30,72])

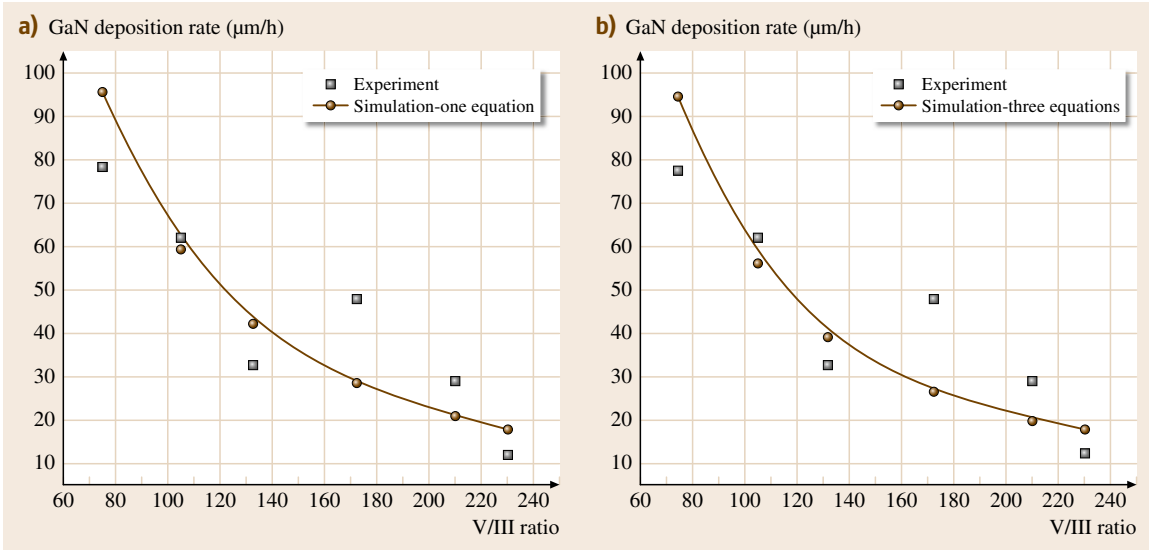


Fig. 37.23a,b GaN deposition rate as a function of the V/III ratio for **(a)** one surface reaction and **(b)** three surface reactions (after [37.72])

achieved is 65 μm/h. It is found that the GaN growth rate is independent of the NH_3 flow rate but dependent on the GaI concentration.

Experiments were conducted to examine the effect of V/III ratio on the GaN deposition rate [37.72, 120]. The V/III ratio was varied by maintaining the NH_3 flow rate at 1 slm while changing the I_2 concentration in the source. Using the same operating conditions as in the experiment, simulations were performed with reaction (37.73) as the boundary condition. The surface reaction rate coefficient δA_p and the concentration dependency of GaI α were determined by fitting the simulation results to the experimental data. Figure 37.23a shows the variation of the GaN deposition rate with the V/III ratio.

The surface reaction rate of reaction (37.73) used to obtain the numerical results in Fig. 37.23a is determined by the following equation

$$\dot{w} = 6.5\times10^{9}\exp\left(\frac{-12\,390}{T}\right)[\mathrm{GaI}]^{2.2}\,. \qquad (37.84)$$

It is concluded that, when one overall surface reaction step is used to predict the GaN deposition rate, the reaction rate coefficient is 6.5×10^{9} and the GaI concentration dependency is 2.2.

When reactions (37.73–37.75) are used, the reaction rate coefficient of each reaction step is determined by matching the simulation results to the experimental data. Figure 37.23b shows the GaN deposition rate calculated using all three surface reactions and the deposition rates from the experiment. The surface reactions and their reaction rates used here are summarized in Table 37.13.

The calculated GaN deposition rate with reactions (37.73) and (37.74) is presented in Table 37.14. Since GaI is considered as the only medium for Ga source transport, the contribution of reaction (37.75) to the deposition rate is found to be less than 0.1% for all cases and is therefore not listed in Table 37.14. In the experiment, the contribution of reaction (37.75) to the GaN deposition rate might be larger. It is, however, not expected to be as significant as reactions (37.73)

Table 37.13 Surface reaction rates obtained by matching the simulation data with the experimental data

Reaction		Rate expression (kmol/m^2 s)
(37.73)	$\mathrm{GaI(g)+NH_3(g)\rightleftharpoons GaN(s)+HI(g)+H_2(g)}$	$0.07\exp(-12\,390/T)[\mathrm{GaI}]$
(37.74)	$\mathrm{3GaI(g)+2NH_3(g)\rightleftharpoons 2GaN(s)+GaI_3(g)+3H_2(g)}$	$5.88\times10^{9}\exp(-13\,630/T)[\mathrm{GaI}]^3$
(37.75)	$\mathrm{GaI_3(g)+NH_3(g)\rightleftharpoons GaN(s)+3HI(g)}$	$0.36\exp(-14\,000/T)[\mathrm{GaI_3}]$

Table 37.14 GaN deposition rate and rate contribution from reactions (37.73) and (37.74)

V/III ratio	GaN deposition rate (μm/h)		Growth rate contribution (%)	
	Reaction (37.73)	Reaction (37.74)	Reaction (37.73)	Reaction (37.74)
75	21.80	73.10	22.97	38.03
105	17.62	38.61	31.33	68.67
132	15.04	24.01	38.52	61.48
172	12.54	13.89	47.41	52.59
210	10.86	9.03	54.60	45.40
230	10.16	7.40	57.86	42.14

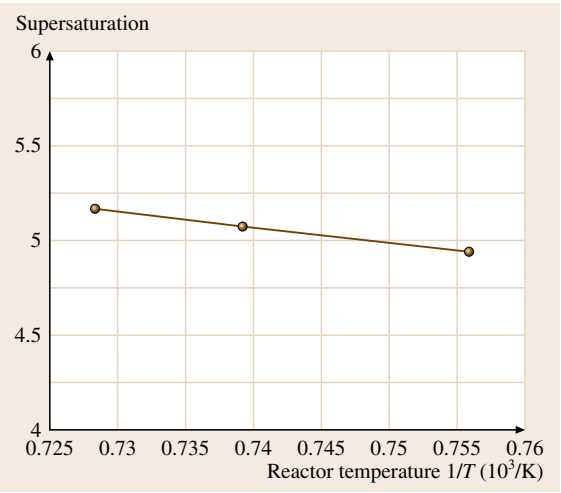


Fig. 37.24 Supersaturation of GaN growth under different reactor temperatures

and (37.74). The contribution of reaction (37.73) to the total GaN deposition rate increases with the V/III ratio, while for reaction (37.74) it decreases with the V/III ratio.

The supersaturation of GaN growth on the substrate can be calculated once the partial pressure of each species is obtained. Figure 37.24 shows the calculation results for the supersaturation with different reactor temperatures. The supersaturation is around 5. This is typical for an epitaxy growth system that needs a relatively low supersaturation to prevent the reactant partial pressures from deviating excessively from their equilibrium values. The supersaturation decreases only slightly with temperature, which means that the GaN deposition is diffusion controlled under the current operating conditions.

Table 37.15 Sticking probability of reactant i with different V/III ratios

V/III ratio	Sticking probability		
	GaI ($\times 10^{-3}$)	NH_3($\times 10^{-6}$)	GaI_3($\times 10^{-3}$)
75	2.66	6.50	4.15
105	2.07	4.10	3.12
132	1.75	3.02	2.69
172	1.43	2.08	2.07
210	1.22	1.55	1.86
230	1.13	1.38	1.77

Table 37.15 summarizes the sticking probability of the reacting species deduced from various simulations. The calculated sticking probabilities of the order of 10^{-3}–10^{-6} are reasonable for the saturated compound. The sticking probability of species i is defined by

$$S_i = f(\theta^s)_i \exp\left[\frac{-E_{ai}(\theta^s)}{RT}\right] = \frac{R_i}{F_i}\,, \qquad (37.85)$$

where $f(\theta^s)_i$ is the function of the existing surface coverage of adsorbed species i, θ^s is the Langmuir definition of surface coverage of certain species, or the fraction of sites which are occupied on the substrate surface, E_{ai} is the activation energy for adsorption of species i,

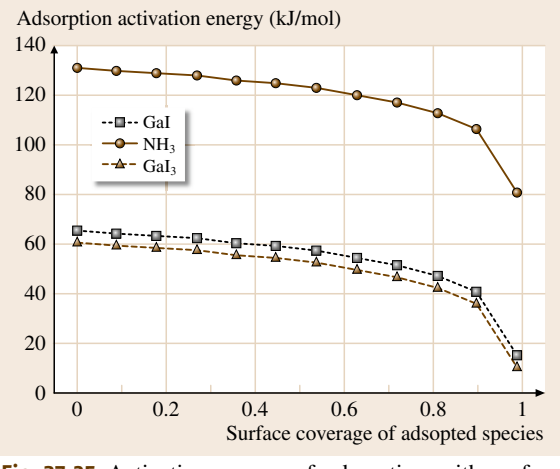


Fig. 37.25 Activation energy of adsorption with surface coverage of 0–0.99

R_i (kmol/m^{2}s) is the rate of adsorption of species i chemically, and $F_i = P_i^{\mathrm{w}}(2\pi RTM_i)^{1/2}$ [kmol/m^{2}s] is the incident flux rate of species i onto the substrate.

It is observed that the sticking probability decreases with increasing V/III ratio. When the I_2 concentration decreases in the source, the species reaction rate decreases more significantly compared with its incident flux onto the substrate. The sticking probability of ammonia is three orders of magnitude smaller than that of GaI or GaI_3, because the ammonia concentration is kept much higher than that of GaI on the substrate.

To determine the activation energy for adsorption of species i, it is assumed that the sticking probability is directly proportional to the concentration of vacant surface sites $(1-\theta^{\mathrm{s}})$. This assumption is a reasonable first approximation for nondissociative adsorption. The activation energy of adsorption is assumed to be independent of the surface coverage. The sticking probability is, thus, revised as

$$S_i = (1-\theta^{\mathrm{s}})\exp\left(\frac{-E_{ai}}{\mathrm{RT}}\right). \tag{37.86}$$

Figure 37.25 shows the calculated activation energy of adsorption for the three species for surface coverage of 0–0.99. The activation energy drops as the surface coverage increases. The averaged activation energies of adsorption for GaI, NH_3, and GaI_3 by calculation are 5.72×10^4, 1.24×10^5, and 5.28×10^4 J/mol, respectively.

37.5 Surface Evolution of GaN/AlN Film Growth from Vapor

In this section, sublimation and HVPE growth of AlN will be used as an example to investigate the surface evolution of thin film during growth. Since large-diameter AlN single crystal is not available in nature, the goal of research is to find a way to produce large, thick AlN films/ingots, which can then be sliced for use as substrates. The growth conditions of two systems are compared in Table 37.16. Figure 37.26 shows the AlN crystals grown by sublimation and HVPE techniques.

Srolovitz et al. [37.128] studied the surface evolution of film growth in which surface diffusion determines the crystal grain size and morphology. Figure 37.27 shows the deposition geometry in which a one-dimensional partial differential equation describes the evolution of an arbitrary initial surface profile $h(x, 0)$ under the joint influence of constant uniform deposition flux rate J of finite-size atoms of radius r and surface diffusion. C_{k} is the surface curvature and V_{s} is the velocity of surface diffusion.

By assuming a sinusoidal perturbation, i. e., $h(x, t) = \sin(x, t)$, the surface profile can be described as

$$h(x,t) = \mathrm{e}^{\left(rJk^2 - D_{\mathrm{e}}k^4\right)t}\sin(kx)\,, \tag{37.87}$$

where $D_{\mathrm{e}} = (D_{\mathrm{s}}\sigma_{\mathrm{s}}\Omega^2\varepsilon)/(k_{\mathrm{B}}T)$, D_{s} is the surface diffusivity of the atoms, σ_{s} is the isotropic surface energy density, Ω is the atomic volume, ε is the number of atoms per unit area, $k_{\mathrm{B}}T$ is the thermal energy, and k is the wavenumber, relating to the wavelength λ by $\lambda = 2\pi/k$. It is revealed by (37.88) that, if $\delta Jk^2 - D_{\mathrm{e}}k^4 > 0$, $\mathrm{e}^{\left(rJk^2 - D_{\mathrm{e}}k^4\right)t} > 1$ is obtained, and the initial sinusoidal perturbation will be enlarged, which means that the growth will be unstable and growth quality will be deteriorated. For stable growth, the wavelength of the initial perturbation λ has to be smaller than the effective diffusion length λ_{e}, which is defined as

$$\lambda_{\mathrm{e}} = \sqrt{\frac{4\pi D_{\mathrm{e}}}{rJ}}\,. \tag{37.88}$$

Table 37.16 Growth comparison for sublimation and HVPE systems

Growth system	Sublimation	HVPE
Surface reaction	$\mathrm{Al(g)} + \frac{1}{2}\mathrm{N_2(g)} \rightleftharpoons \mathrm{AlN(s)}$	$\mathrm{AlCl_3(g)} + \mathrm{NH_3(g)} \rightleftharpoons \mathrm{AlN(s)} + 3\mathrm{HCl(g)}$
Substrate	AlN seed	Sapphire/SiC
Growth temperature	2240 °C	900–1200 °C
Growth rate	0.2–0.5 mm/h along c axis	5–75 μm/h
Crystal property	Transparent, DCRC 20 arcsec	Opaque, DCRC 1000 arcsec

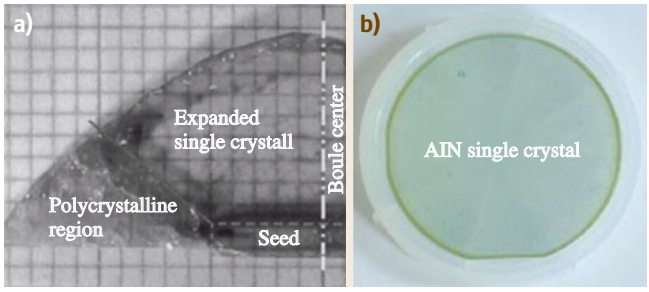


Fig. 37.26 **(a)** An AlN crystal grown by sublimation method (after [37.70]). **(b)** An AlN crystal grown by HVPE method (after [37.130]) (with permission from Elsevier) ▶

To obtain stable growth, one can either reduce the initial perturbation wavelength or increase the effective diffusion wavelength. The initial perturbation wavelength is related to the substrate and supersaturation. The effective diffusion length is proportional to the effective diffusion coefficient D_e and inversely proportional to the atom radius r and the growth flux rate. *Collazo* et al. [37.129] studied the effect of process conditions such as surface temperature T and the growth flux rate J on the effective diffusion length λ_e and the effect of process conditions such as supersaturation on the initial perturbation wavelength λ. They found that changing the process conditions cannot significantly affect the values of λ_e and λ.

It is concluded here that the substrate plays a critical role in determining the quality of AlN crystal when the crystal is grown large and thick. For HVPE growth using a foreign substrate such as sapphire or silicon carbide, the wavelength of the surface perturbation will be larger than the effective surface diffusion length due to the lattice mismatch between the AlN crystal and substrate when the thickness of the crystal is large.

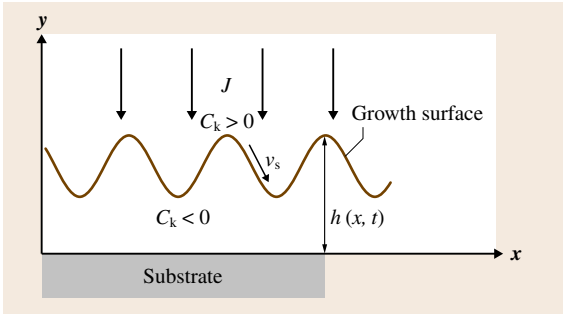


Fig. 37.27 Deposition geometry (after [37.128])

Therefore the film growth will be unstable, and consequently, low-quality crystals will be obtained.

37.6 Concluding Remarks

Transport phenomena in chemical vapor deposition of III nitrides are complicated due to gas flow, coupled convection and radiation, conjugate heat and mass transfer, homogeneous gas-phase and heterogeneous surface reactions, and the coexistence of multiple length and time scales. The multiplicity of governing parameters, complex geometric configurations, complicated boundary conditions, and the lack of information on gas-phase and surface reaction mechanisms make this process difficult to predict and control.

It is the authors' intention to provide a systematical procedure for those who are new to the field of vapor-growth process modeling. In this chapter an attempt has been made to present a comprehensive description of the fundamental theory as well as an extensive overview of the state of the art in vapor-growth process modeling. The main focus has been on continuum modeling of macroscopic gas flow, homogeneous gas-phase reactions inside the reactor chamber, heterogeneous surface reactions on the substrate surface, heat transfer, and species transport in reactors for GaN/AlN vapor crystal growth.

Gas velocity, temperature and pressure, and species concentration distributions in a CVD reactor are correlated. The main difficulty in accurately predicting species concentration lies in the lack of detailed information for gas-phase and surface reactions. By using GaN IVPE growth as an example, mathematic models for thermodynamic and kinetic gas-phase and surface reactions have been described. Thermodynamic models can be used to predict gas-phase and surface reactions with or without detailed reactions steps. The results, however, only give the upper limit of what one can expect. By comparing the results with available experiment observations, one may determine the extent to which the gas-phase and surface reactions deviate from the equilibrium assumption. Empirical models were used in this chapter to simulate gas-phase and surface reaction kinetically. The detailed multistep gas-phase and surface reaction

mechanisms can only be obtained from the experiments. The current models can be used to predict species transport in the reactor in order to identify the critical controlling parameters for fast and uniform deposition.

Film instability during GaN/AlN growth has been analyzed, in which surface diffusion determines the crystal grain size and morphology. The key parameters, such as the lattice mismatch between the substrate and deposited film, the effective diffusion length, and the supersaturation above the substrate, have been identified and their effects on film instability during growth quantified. It is found that, for growth of GaN/AlN bulk crystals, use of native substrate is the key for high deposition quality when the deposited film thickness becomes very large.

References

37.1 E.J. Tarsa, B. Heying, X.H. Wu, P. Fini, S.P. DenBaars, J.S. Speck: Homoepitaxial growth of GaN under Ga-stable and N-stable conditions by plasma-assisted molecular beam epitaxy, J. Appl. Phys. **82**, 5472–5479 (1997)

37.2 C.R. Elsass, I.P. Smorchkova, H.Y. Ben, E. Haus, C. Poblenz, P. Fini, K. Maranowski, P.M. Petroff, S.P. DenBaars, U.K. Mishra, J.S. Speck, A. Saxler, S. Elhamri, W.C. Mitchel: Electron transport in AlGaN/GaN heterostructures grown by plasma-assisted molecular beam epitaxy, Jpn. J. Appl. Phys. Part 2–Letters **39**, L1023–L1025 (2000)

37.3 A. Hierro, A.R. Arehart, B. Heying, M. Hansen, J.S. Speck, U.K. Mishra, S.P. DenBaars, S.A. Ringel: Capture kinetics of electron traps in MBE-grown n-GaN, Phys. Status Solidi b **228**, 309–313 (2001)

37.4 R.F. Davis, A.M. Roskowski, E.A. Preble, J.S. Speck, B. Heying, J.A. Freitas, E.R. Glaser, W.E. Carlos: Gallium nitride materials – progress, status, and potential roadblocks, Proc. IEEE **90**, 993–1005 (2002)

37.5 P. Waltereit, H. Sato, C. Poblenz, D.S. Green, J.S. Brown, M. McLaurin, T. Katona, S.P. DenBaars, J.S. Speck, J.H. Liang, M. Kato, H. Tamura, S. Omori, C. Funaoka: Blue GaN-based light-emitting diodes grown by molecular-beam epitaxy with external quantum efficiency greater than 1.5%, Appl. Phys. Lett. **84**, 2748–2750 (2004)

37.6 C. Poblenz, P. Waltereit, J.S. Speck: Uniformity and control of surface morphology during growth of GaN by molecular beam epitaxy, J. Vac. Sci. Technol. B **23**, 1379–1385 (2005)

37.7 A. Corrion, F. Wu, T. Mates, C.S. Gallinat, C. Poblenz, J.S. Speck: Growth of Fe-doped GaN by RF plasma-assisted molecular beam epitaxy, J. Cryst. Growth **289**, 587–595 (2006)

37.8 A. Corrion, C. Poblenz, P. Waltereit, T. Palacios, S. Rajan, U.K. Mishra, J.S. Speck: Review of recent developments in growth of AlGaN/GaN high-electron mobility transistors on 4H-SiC by plasma-assisted molecular beam epitaxy, IEICE Trans. Electron. **E89C**, 906–912 (2006)

37.9 A.S. Segal, A.V. Kondratyev, S.Y. Karpov, D. Martin, V. Wagner, M. Ilegems: Surface chemistry and transport effects in GaN hydride vapor phase epitaxy, J. Cryst. Growth **270**, 384–395 (2004)

37.10 S.Y. Karpov, D.V. Zimina, Y.N. Makarov, B. Beaumont, G. Nataf, P. Gibart, M. Heuken, H. Jurgensen, A. Krishnan: Modeling study of hydride vapor phase epitaxy of GaN, Phys. Status Solidi a **176**, 439–442 (1999)

37.11 A.S. Segal, S.Y. Karpov, Y.N. Makarov, E.N. Mokhov, A.D. Roenkov, M.G. Ramm, Y.A. Vodakov: On mechanisms of sublimation growth of AlN bulk crystals, J. Cryst. Growth **211**, 68–72 (2000)

37.12 M.V. Bogdanov, S.E. Demina, S.Y. Karpov, A.V. Kulik, M.S. Ramm, Y.N. Makarov: Advances in modeling of wide-bandgap bulk crystal growth, Cryst. Res. Technol. **38**, 237–249 (2003)

37.13 E.N. Mokhov, A.D. Roenkov, Y.A. Vodakov, S.Y. Karpov, M.S. Ramm, A.S. Segal, Y.A. Makarov, H. Helava: Growth of AlN bulk crystals by sublimation sandwich method. In: *Silicon Carbide and Related Materials*, ed. by P. Bergman, E. Janzén (Trans Tech, Zurich 2002) pp. 979–982

37.14 S.Y. Karpov, D.V. Zimina, Y.N. Makarov, E.N. Mokhov, A.D. Roenkov, M.G. Ramm, Y.A. Vodakov: Sublimation growth of AlN in vacuum and in a gas atmosphere, Phys. Status Solidi a **176**, 435–438 (1999)

37.15 P.G. Baranov, E.N. Mokhov, A.O. Ostroumov, M.G. Ramm, M.S. Ramm, V.V. Ratnikov, A.D. Roenkov, Y.A. Vodakov, A.A. Wolfson, G.V. Saparin, S.Y. Karpov, D.V. Zimina, Y.N. Makarov, H. Juergensen: Current status of GaN crystal growth by sublimation sandwich technique, MRS Internet J. Nitride Semicond. Res. **3**, 50 (1998)

37.16 C.H. Wei, Z.Y. Xie, L.Y. Li, Q.M. Yu, J.H. Edgar: MOCVD growth of cubic GaN on 3C-SiC deposited on Si(100) substrates, J. Electron. Mater. **29**, 317–321 (2000)

37.17 Z.Y. Xie, C.H. Wei, L.Y. Li, J.H. Edgar, J. Chaudhuri, C. Ignatiev: Effects of surface preparation on epitaxial GaN on 6H-SiC deposited via MOCVD, MRS Internet J. Nitride Semicond. Res. **4S1**, G3.39 (1999)

37.18 B.S. Sywe, J.R. Schlup, J.H. Edgar: Fourier-transform infrared spectroscopic study of predeposition reactions in metalloorganic chemical

vapor-deposition of gallium nitride – Part II, Chem. Mater. **3**, 1093–1097 (1991)
37.19 J.H. Edgar, L.H. Robins, S.E. Coatney, L. Liu, J. Chaudhuri, K. Ignatiev, Z. Rek: A comparison of aluminum nitride freely nucleated and seeded on 6H-silicon carbide, Silicon Carbide Relat. Mater. **338–3**, 1599–1602 (2000)
37.20 Y. Shi, B. Liu, L.H. Liu, J.H. Edgar, E.A. Payzant, J.M. Hayes, M. Kuball: New technique for sublimation growth of AlN single crystals, MRS Internet J. Nitride Semicond. Res. **6**, 1–10 (2001)
37.21 J.H. Edgar, L. Liu, B. Liu, D. Zhuang, J. Chaudhuri, M. Kuball, S. Rajasingam: Bulk AlN crystal growth: self-seeding and seeding on 6H-SiC substrates, J. Cryst. Growth **246**, 187–193 (2002)
37.22 B. Liu, J.H. Edgar, Z. Gu, D. Zhuang, B. Raghothamachar, M. Dudley, A. Sarua, M. Kuball, H.M. Meyer: The durability of various crucible materials for aluminum nitride crystal growth by sublimation, MRS Internet J. Nitride Semicond. Res. **9**, 6 (2004)
37.23 Z. Gu, L. Du, J.H. Edgar, E.A. Payzant, L. Walker, R. Liu, M.H. Engelhard: Aluminum nitride-silicon carbide alloy crystals grown on SiC substrates by sublimation, MRS Internet J. Nitride Semicond. Res. **10**, 5 (2005)
37.24 Z. Gu, J.H. Edgar, D.W. Coffey, J. Chaudhuri, L. Nyakiti, R.G. Lee, J.G. Wen: Defect-selective etching of scandium nitride crystals, J. Cryst. Growth **293**, 242–246 (2006)
37.25 M. Callahan, M. Harris, M. Suscavage, D.F. Bliss, J. Bailey: Synthesis and growth of gallium nitride by the chemical vapor reaction process (CVRP), MRS Internet J. Nitride Semicond. Res. **4**, 10 (1999)
37.26 V. Tassev, D.F. Bliss, M. Suscavage, Q.S. Paduano, S.Q. Wang, L. Bouthillette: Iodine vapor phase growth of GaN: dependence of epitaxial growth rate on process parameters, J. Cryst. Growth **235**, 140–148 (2002)
37.27 M. Suscavage, L. Bouthillette, D.F. Bliss, S.Q. Wang, C. Sung: New iodide method for growth of GaN, Phys. Status Solidi a **188**, 477–480 (2001)
37.28 D.F. Bliss, V.L. Tassev, D. Weyburne, J.S. Bailey: Aluminum nitride substrate growth by halide vapor transport epitaxy, J. Cryst. Growth **250**, 1–6 (2003)
37.29 D. Cai, L.L. Zheng, H. Zhang, V.L. Tassev, D.F. Bliss: Modeling of aluminum nitride growth by halide vapor transport epitaxy method, J. Cryst. Growth **276**, 182–193 (2005)
37.30 W.J. Mecouch, B.J. Rodriguez, Z.J. Reitmeier, J.S. Park, R.F. Davis, Z. Sitar: Initial stages of growth of gallium nitride via iodine vapor phase epitaxy, Mater. Res. Soc. Symp. Proc. **831**, E3.23.1. (2005)
37.31 R. Schlesser, Z. Sitar: Growth of bulk AlN crystals by vaporization of aluminum in a nitrogen atmosphere, J. Cryst. Growth **234**, 349–353 (2002)
37.32 D. Zhuang, Z.G. Herro, R. Schlesser, B. Raghothamachar, M. Dudley, Z. Sitar: Seeded growth of AlN crystals on nonpolar seeds via physical vapor transport, J. Electron. Mater. **35**, 1513–1517 (2006)
37.33 D. Zhuang, Z.G. Herro, R. Schlesser, Z. Sitar: Seeded growth of AlN single crystals by physical vapor transport, J. Cryst. Growth **287**, 372–375 (2006)
37.34 R. Dalmau, R. Schlesser, B.J. Rodriguez, R.J. Nemanich, Z. Sitar: AlN bulk crystals grown on SiC seeds, J. Cryst. Growth **281**, 68–74 (2005)
37.35 V. Noveski, R. Schlesser, B. Raghothamachar, M. Dudley, S. Mahajan, S. Beaudoin, Z. Sitar: Seeded growth of bulk AlN crystals and grain evolution in polycrystalline AlN boules, J. Cryst. Growth **279**, 13–19 (2005)
37.36 V. Noveski, R. Schlesser, S. Mahajan, S. Beaudoin, Z. Sitar: Growth of AlN crystals on AlN/SiC seeds by AlN powder sublimation in nitrogen atmosphere, MRS Internet J. Nitride Semicond. Res. **9**, 2 (2004)
37.37 B. Wu, R.H. Ma, H. Zhang, M. Dudley, R. Schlesser, Z. Sitar: Growth kinetics and thermal stress in AlN bulk crystal growth, J. Cryst. Growth **253**, 326–339 (2003)
37.38 E. Silveira, J.A. Freitas, G.A. Slack, L.J. Schowalter, M. Kneissl, D.W. Treat, N.M. Johnson: Depth-resolved cathodoluminescence of a homoepitaxial AlN thin film, J. Cryst. Growth **281**, 188–193 (2005)
37.39 J.C. Rojo, G.A. Slack, K. Morgan, B. Raghothamachar, M. Dudley, L.J. Schowalter: Report on the growth of bulk aluminum nitride and subsequent substrate preparation, J. Cryst. Growth **231**, 317–321 (2001)
37.40 L.J. Schowalter, J.C. Rojo, N. Yakolev, Y. Shusterman, K. Dovidenko, R.J. Wang, I. Bhat, G.A. Slack: Preparation and characterization of single-crystal aluminum nitride substrates, MRS Internet J. Nitride Semicond. Res. **5**, W6.7 (2000)
37.41 G. Koley, M.G. Spencer: Scanning Kelvin probe microscopy characterization of dislocations in III-nitrides grown by metalorganic chemical vapor deposition, Appl. Phys. Lett. **78**, 2873–2875 (2001)
37.42 I. Jenkins, K.G. Irvine, M.G. Spencer, V. Dmitriev, N. Chen: Growth of solid-solutions of aluminum nitride and silicon-carbide by metalorganic chemical vapor-deposition, J. Cryst. Growth **128**, 375–378 (1993)
37.43 K. Wongchotigul, N. Chen, D.P. Zhang, X. Tang, M.G. Spencer: Low resistivity aluminum nitride: Carbon (AlN:C) films grown by metal organic chemical vapor deposition, Mater. Lett. **26**, 223–226 (1996)
37.44 C.M. Zetterling, M. Ostling, K. Wongchotigul, M.G. Spencer, X. Tang, C.I. Harris, N. Nordell, S.S. Wong: Investigation of aluminum nitride grown by metal-organic chemical-vapor deposition on silicon, J. Appl. Phys. **82**, 2990–2995 (1997)
37.45 K. Wongchotigul, S. Wilson, C. Dickens, J. Griffin, X. Tang, M.G. Spencer: Growth of aluminum nitride with superior optical and morphological properties. In: *Silicon Carbide, III–Nitrides and Related*

Materials, Parts 1 and 2, ed. by G. Pensl, H. Morkoç, B. Monemar, E. Janzén (Trans Tech, Zurich 1998) pp. 1137–1140

37.46 H. Hirayama: Quaternary InAlGaN-based high-efficiency ultraviolet light-emitting diodes, J. Appl. Phys. **97**, 091101 (2005)

37.47 S. Krukowski, C. Skierbiszewski, P. Perlin, M. Leszczynski, M. Bockowski, S. Porowski: Blue and UV semiconductor lasers, Acta Phys. Polonica B **37**, 1265–1312 (2006)

37.48 T.P. Chow, R. Tyagi: Wide bandgap compound semiconductors for superior high-voltage unipolar power devices, IEEE Trans. Electron. Dev. **41**, 1481–1483 (1994)

37.49 S.J. Pearton, C.R. Abernathy, B.P. Gila, A.H. Onstine, M.E. Overberg, G.T. Thaler, J. Kim, B. Luo, R. Mehandru, F. Ren, Y.D. Park: Recent advances in gate dielectrics and polarised light emission from GaN, Optoelectron. Rev. **10**, 231–236 (2002)

37.50 S.H. Kim, J.H. Ko, S.H. Ji, Y.S. Yoon: Crystallinity effect of AlN thin films on the frequency response of an AlN/IDT/Si surface acoustic wave device, J. Korean Phys. Soc. **49**, 199–202 (2006)

37.51 K. Katahira, H. Ohmori, Y. Uehara, M. Azuma: ELID grinding characteristics and surface modifying effects of aluminum nitride (AlN) ceramics, Int. J. Mach. Tool. Manuf. **45**, 891–896 (2005)

37.52 Y.G. Gao, D.A. Gulino, R. Higgins: Effects of susceptor geometry on GaN growth on Si(111) with a new MOCVD reactor, MRS Internet J. Nitride Semicond. Res. **4S1**, G3.53 (1999)

37.53 M. Morita, S. Isogai, N. Shimizu, K. Tsubouchi, N. Mikoshiba: Aluminum nitride epitaxially grown on silicon – orientation relationships, Jpn. J. Appl. Phys. **20**, L173–L175 (1981)

37.54 J.D. Brown, R. Borges, E. Piner, A. Vescan, S. Singhal, R. Therrien: AlGaN/GaN HFETs fabricated on 100 mm GaN on silicon (111) substrates, Solid-State Electron. **46**, 1535–1539 (2002)

37.55 K. Kim, S.K. Noh: Reactor design rules for GaN epitaxial layer growths on sapphire in metal-organic chemical vapour deposition, Semicond. Sci. Technol. **15**, 868–874 (2000)

37.56 R.P. Parikh, R.A. Adomaitis: An overview of gallium nitride growth chemistry and its effect on reactor design: application to a planetary radial-flow CVD system, J. Cryst. Growth **286**, 259–278 (2006)

37.57 W.E. Hoke, A. Torabi, R.B. Hallock, J.J. Mosca, T.D. Kennedy: Reaction of molecular beam epitaxial grown AlN nucleation layers with SiC substrates, J. Vac. Sci. Technol. B **24**, 1500–1504 (2006)

37.58 R.M. Feenstra, Y. Dong, C.D. Lee, J.E. Northrup: Recent developments in surface studies of GaN and AlN, J. Vac. Sci. Technol. B **23**, 1174–1180 (2005)

37.59 S.H. Cho, K. Hata, T. Maruyama, K. Akimoto: Optical and structural properties of GaN films grown on c-plane sapphire by ECR-MBE, J. Cryst. Growth **173**, 260–265 (1997)

37.60 A. Usui, H. Sunakawa, A. Sakai, A.A. Yamaguchi: Thick GaN epitaxial growth with low dislocation density by hydride vapor phase epitaxy, Jpn. J. Appl. Phys. Part 2 – Letters **36**, L899–L902 (1997)

37.61 A. Dollet, Y. Casaux, M. Matecki, R. Rodriguez-Clemente: Chemical vapour deposition of polycrystalline AlN films from $AlCl_3 - NH_3$ mixtures: II – surface morphology and mechanisms of preferential orientation at low-pressure, Thin Solid Films **406**, 118–131 (2002)

37.62 M. Callahan, B.G. Wang, K. Rakes, D.F. Bliss, L. Bouthillette, M. Suscavage, S.Q. Wang: GaN single crystals grown on HVPE seeds in alkaline supercritical ammonia, J. Mater. Sci. **41**, 1399–1407 (2006)

37.63 Y. Kumagai, T. Yamane, A. Koukitu: Growth of thick AlN layers by hydride vapor-phase epitaxy, J. Cryst. Growth **281**, 62–67 (2005)

37.64 M. Tanaka, S. Nakahata, K. Sogabe, H. Nakata, M. Tobioka: Morphology and x-ray diffraction peak widths of aluminum nitride single crystals prepared by the sublimation method, Jpn. J. Appl. Phys. Part 2 – Letters **36**, L1062–L1064 (1997)

37.65 C.M. Balkas, Z. Sitar, T. Zheleva, L. Bergman, R. Nemanich, R.F. Davis: Sublimation growth and characterization of bulk aluminum nitride single crystals, J. Cryst. Growth **179**, 363–370 (1997)

37.66 S. Kurai, K. Nishino, S. Sakai: Nucleation control in the growth of bulk GaN by sublimation method, Jpn. J. Appl. Phys. Part 2–Letters **36**, L184–L186 (1997)

37.67 Y. Naoi, K. Kobatake, S. Kurai, K. Nishino, H. Sato, M. Nozaki, S. Sakai, Y. Shintani: Characterization of bulk GaN grown by sublimation technique, J. Cryst. Growth **190**, 163–166 (1998)

37.68 I.K. Shmagin, J.F. Muth, J.H. Lee, R.M. Kolbas, C.M. Balkas, Z. Sitar, R.F. Davis: Optical metastability in bulk GaN single crystals, Appl. Phys. Lett. **71**, 455–457 (1997)

37.69 L.J. Schowalter, J.C. Rojo, G.A. Slack, Y. Shusterman, R. Wang, I. Bhat, G. Arunmozhi: Epitaxial growth of AlN and $Al_{0.5}Ca_{0.5}N$ layers on aluminum nitride substrates, J. Cryst. Growth **211**, 78–81 (2000)

37.70 D. Zhuang, Z.G. Herro, R. Schlesser, Z. Sitar: Seeded growth of AlN single crystals by physical vapor transport, J. Cryst. Growth **287**, 372–375 (2006)

37.71 W.J. Mecouch: Preparation and characterization of thin, tomically clean GaN(0001) and AlN(0001) films and the depostion of thick GaN films via iodine vapor phase growth, Mater. Sci. Eng. Dep. (North Carolina State Univ. 2005)

37.72 D. Cai, W.J. Mecouch, L.L. Zheng, H. Zhang, Z. Sitar: Thermodynamic and kinetic study of transport and reaction phenomena in gallium nitride epitaxy growth, Int. J. Heat Mass Transf. **51**, 1264–1280 (2008)

37.73 R.J. Molnar, W. Gotz, L.T. Romano, N.M. Johnson: Growth of gallium nitride by hydride vapor-phase epitaxy, J. Cryst. Growth **178**, 147–156 (1997)
37.74 T. Paskova, E.M. Goldys, B. Monemar: Hydride vapour-phase epitaxy growth and cathodoluminescence characterisation of thick GaN films, J. Cryst. Growth **203**, 1–11 (1999)
37.75 K. Ohno, K. Esfarjani, Y. Kawazoe: *Computational Materials Science: From ab Initio to Monte Carlo Methods* (Springer, Berlin 1999)
37.76 W. Hergert, A. Ernst, M. Däne: *Computational Materials Science: From Basic Principles to Material Properties* (Springer, Berlin 2004)
37.77 D. Raabe: *Computational Materials Science: The Simulation of Materials Microstructures and Properties* (Wiley-VCH, Weinhein 1998)
37.78 C.R.A. Catlow: *Computational Materials Science* (NATO Science, Amsterdam 2003)
37.79 M.A. Gosálvez, R.M. Nieminen: Target-rate kinetic Monte Carlo method for simulation of nanoscale structural evolution, J. Comput. Theor. Nanosci. **1**, 303–308 (2004)
37.80 C.C. Battaile, D.J. Srolovitz, J.E. Butler: A kinetic Monte Carlo method for the atomic-scale simulation of chemical vapor deposition: application to diamond, J. Appl. Phys. **82**, 6293–6300 (1997)
37.81 C.C. Battaile, D.J. Srolovitz: Kinetic Monte Carlo simulation of chemical vapor deposition, Annu. Rev. Mater. Res. **32**, 297–319 (2002)
37.82 C. Cavallotti, A. Barbato, A. Veneroni: A combined three-dimensional kinetic Monte Carlo and quantum chemistry study of the CVD of Si on Si(100) surfaces, J. Cryst. Growth **266**, 371–380 (2004)
37.83 C. Cavallotti, M. DiStanislao, D. Moscatelli, A. Veneroni: Materials computation towards technological impact: The multiscale approach to thin films deposition, Electrochem. Acta **50**, 4566–4575 (2005)
37.84 Y. Akiyama: Modeling thermal CVD, J. Chem. Eng. Jpn. **35**, 701–713 (2002)
37.85 L.C. Musson, P. Ho, S.J. Plimpton, R.C. Schmidt: Feature length-scale modeling of LPCVD and PECVD MEMS fabrication processes, Microsystem Technol. Micro. Nanosyst. Info. Storage Proc. Syst. **12**, 137–142 (2005)
37.86 H.A. Al-Mohssen, N.G. Hadjiconstantinou: Arbitrary-pressure chemical vapor deposition modeling using direct simulation Monte Carlo with nonlinear surface chemistry, J. Comput. Phys. **198**, 617–627 (2004)
37.87 K.F. Jensen: Transport phenomena in vapor phase epitaxy reactors. In: *Handbook of Crystal Growth*, ed. by F.J.T. Hurle (Elsevier, Amsterdam 1994) pp. 541–
37.88 C.R. Kleijn: Chemical vapor deposition processes. In: *Computational Modeling in Semiconductor Processing*, ed. by M. Meyyappan (Artech House, Northwood 1995) pp. 97–216
37.89 A.A. Schmidt, V.S. Kharlamov, K.L. Safonov, Y.V. Trushin, E.E. Zhurkin, V. Cimalla, O. Ambacher, J. Pezoldt: Growth of three-dimensional SiC clusters on Si modelled by KMC, Comput. Mater. Sci. **33**, 375–381 (2005)
37.90 L.H. Liu, J.H. Edgar: Transport effects in the sublimation growth of aluminum nitride, J. Cryst. Growth **220**, 243–253 (2000)
37.91 L.H. Liu, J.H. Edgar: A global growth rate model for aluminum nitride sublimation, J. Electrochem. Soc. **149**, G12–G15 (2002)
37.92 S.Y. Karpov, V.G. Prokofyev, E.V. Yakovlev, R.A. Talalaev, Y.N. Makarov: Novel approach to simulation of group-III nitrides growth by MOVPE, MRS Internet J. Nitride Semicond. Res. **4**, 4 (1999)
37.93 S.Y. Karpov, R.A. Talalaev, Y.N. Makarov, N. Grandjean, J. Massies, B. Damilano: Surface kinetics of GaN evaporation and growth by molecular-beam epitaxy, Surf. Sci. **450**, 191–203 (2000)
37.94 S.A. Safvi, N.R. Perkins, M.N. Horton, R. Matyi, T.F. Kuech: Effect of reactor geometry and growth parameters on the uniformity and material properties of GaN/sapphire grown by hydride vapor-phase epitaxy, J. Cryst. Growth **182**, 233–240 (1997)
37.95 S.A. Safvi, J.M. Redwing, M.A. Tischler, T.F. Kuech: GaN growth by metallorganic vapor phase epitaxy – a comparison of modeling and experimental measurements, J. Electrochem. Soc. **144**, 1789–1796 (1997)
37.96 C. Theodoropoulos, T.J. Mountziaris, H.K. Moffat, J. Han: Design of gas inlets for the growth of gallium nitride by metalorganic vapor phase epitaxy, J. Cryst. Growth **217**, 65–81 (2000)
37.97 E. Aujol, J. Napierala, A. Trassoudaine, E. Gil-Lafon, R. Cadoret: Thermodynamical and kinetic study of the GaN growth by HVPE under nitrogen, J. Cryst. Growth **222**, 538–548 (2001)
37.98 A. Dollet, Y. Casaux, G. Chaix, C. Dupuy: Chemical vapour deposition of polycrystalline A1N films from $AlCl_3 - NH_3$ mixtures: analysis and modelling of transport phenomena, Thin Solid Films **406**, 1–16 (2002)
37.99 B. Wu, R.H. Ma, H. Zhang, V. Prasad: Modeling and simulation of AlN bulk sublimation growth systems, J. Cryst. Growth **266**, 303–312 (2004)
37.100 B. Wu, H. Zhang: Isotropic and anisotropic growth models for the sublimation vapour transport process, Model. Simul. Mater. Sci. Eng. **13**, 861–873 (2005)
37.101 D. Cai, L.L. Zheng, H. Zhang, V.L. Tassev, D.F. Bliss: Modeling of gas phase and surface reactions in an aluminum nitride growth system, J. Cryst. Growth **293**, 136–145 (2006)
37.102 A. Dollet: Multiscale modeling of CVD film growth – a review of recent works, Surf. Coat. Technol. **177**, 245–251 (2004)
37.103 A.V. Vasenkov, A.I. Fedoseyev, V.I. Kolobov, H.S. Choi, K.H. Hong, K. Kim, J. Kim, H.S. Lee,

J.K. Shin: Computational framework for modeling of multi-scale processes, J. Comput. Theor. Nanosci. **3**, 453–458 (2006)
37.104 S.T. Rodgers, K.F. Jensen: Multiscale modeling of chemical vapor deposition, J. Appl. Phys. **83**, 524–530 (1998)
37.105 K.F. Jensen, S.T. Rodgers, R. Venkataramani: Multiscale modeling of thin film growth, Curr. Opin. Solid State Mater. Sci. **3**, 562–569 (1998)
37.106 H.N.G. Wadley, A.X. Zhou, R.A. Johnson, M. Neurock: Mechanisms, models and methods of vapor deposition, Prog. Mater. Sci. **46**, 329–377 (2001)
37.107 W.M. Roshsenow, J.P. Hartnett: *Handbook of Heat Tranfer* (McGraw-Hill, New York 1975)
37.108 P. Cheng: 2-dimensional radiating gas flow by a moment method, AIAA Journal **2**, 1662–1664 (1964)
37.109 R. Siegel, J.R. Howell: *Thermal Radiation Heat Transfer* (Hemisphere, Washington 1992)
37.110 M.F. Modest: *Radiative Heat Transfer* (McGraw-Hill, New York 1993)
37.111 M.H.N. Naraghi, J.C. Huan: An N-bounce method for analysis of radiative-transfer in enclosures with anisotropically scattering media, J. Heat Transf. Trans. ASME **113**, 774–777 (1991)
37.112 G.D. Raithby, E.H. Chui: A finite-volume method for predicting a radiant-heat transfer in enclosures with participating media, J. Heat Transf. Trans. ASME **112**, 415–423 (1990)
37.113 M.G. Carvalho, T. Farias, P. Fotes: Predicting radiative heat transfer in absorbing, emmiting, scattering media using the discrete transfer method. In: *Fundamentals of Radiation Heat Transfer*, ed. by W.A. Fiveland (ASME HTD, New York 1991) pp. 17–26
37.114 L.S. Yao: Free and forced-convection in the entry region of a heated vertical channel, Int. J. Heat Mass Transf. **26**, 65–72 (1983)
37.115 T.K. Sherwood, R.J. Pigford, C.R. Wilke: *Mass Transfer* (McGraw-Hill, New York 1975)
37.116 A. Bejan: *Convective Heat Transfer* (Wiley-Interscience, Hoboken 2004)
37.117 O. Knacke, O. Kubaschewski, K. Hesselmann: *Thermochemical Properties of Inorganic Substances I* (Springer, Berlin 1991)
37.118 I. Barin: *Thermochemical Data of Pure Substances* (VCH, Weinheim 1989)
37.119 A.A. Chernov: *Modern Crystallography III. Crystal Growth* (Springer, Berlin 1984)
37.120 W.J. Mecouch: Preparation and characterization of thin, atomically clean GaN(0001) and AlN(0001) films and the depostion of thick GaN films via iodine vapor phase growth, Ph. D. Thesis (North Carolina State Univ. 2005)
37.121 R. Cadoret, E. Gil-Lafon: GaAs growth mechanisms of exact and isoriented {001} faces by the chloride method in H_2: Surface diffusion, spiral growth, HCl and $GaCl_3$ desorption mechanisms, J. Phys. I France **7**, 889–907 (1997)
37.122 D.F. Davidson, K. Kohse-Hoinghaus, A.Y. Chang, R.K. Hanson: A pyrolysis mechanism for ammonia, Int. J. Chem. Kinetics **22**, 513–535 (1990)
37.123 D.R. Stull: Vapor pressure of pure substances – organic compounds, Ind. Eng. Chem. **39**, 517–540 (1947)
37.124 R.F. Rolsten: *Iodide Metals and Metal Iodides* (Wiley, New York 1961)
37.125 A. Shintani, S. Minagawa: Kinetics of epitaxial-growth of GaN using Ga, HCl and NH_3, J. Cryst. Growth **22**, 1–5 (1974)
37.126 D.W. Shaw: Influence of substrate temperature on gaas epitaxial deposition rates, J. Electrochem. Soc. **115**, 405 (1968)
37.127 M. Mesrine, N. Grandjean, J. Massies: Efficiency of NH_3 as nitrogen source for GaN molecular beam epitaxy, Appl. Phys. Lett. **72**, 350–352 (1998)
37.128 D.J. Srolovitz, A. Mazor, B.G. Bukiet: Analytical and numerical modeling of columnar evolution in thin-films, J. Vac. Sci. Technol. A **6**, 2371–2380 (1988)
37.129 R. Collazo, R. Dalmau, Z. Herro, D. Zhuang, Z. Sitar: Is HVPE fundamentally inferior to PVT of AlN, ONR MURI: III–nitride crystal growth and wafering meeting (Arizona 2006)
37.130 O. Kovalenkov, V. Soukhoveev, V. Ivantsov, A. Usikov, V. Dmitriev: Thick AlN layers grown by HVPE, J. Cryst. Growth **281**, 87–92 (2005)

38. Continuum-Scale Quantitative Defect Dynamics in Growing Czochralski Silicon Crystals

Milind S. Kulkarni

The vast majority of modern microelectronic devices are built on monocrystalline silicon substrates produced from crystals grown by the Czochralski (CZ) and float-zone (FZ) processes. Silicon crystals inherently contain various crystallographic imperfections known as microdefects that often affect the yield and performance of many devices. Hence, quantitative understanding and control of the formation and distribution of microdefects in silicon crystals play a central role in determining the quality of silicon substrates. These microdefects are primarily aggregates of intrinsic point defects of silicon (vacancies and self-interstitials) and oxygen (silicon dioxide). The distribution of microdefects in a CZ crystal is determined by the complex dynamics, influenced by various reactions involving the intrinsic point defects and oxygen, and their transport. The distribution of these microdefects can also be strongly influenced and controlled by the addition of impurities such as nitrogen to the crystal. In this chapter, significant developments in the field of defect dynamics in growing CZ and FZ crystals are reviewed. The breakthrough discovery of the *initial point defect incorporation* in the vicinity of the melt–crystal interface, made in the early 1980s, allows a simplified quantification of CZ and FZ defect dynamics. Deeper insight into the formation and growth of microdefects was provided in the last decade by various treatments of the aggregation of oxygen and the intrinsic point defects of silicon. In particular, rigorous quantification of the aggregation of intrinsic point defects using the classical nucleation theory, a recently developed lumped model that captures the microdefect distribution by representing the actual population of microdefects by an equivalent population of identical microdefects, and another rigorous treatment involving the Fokker–Planck equations are discussed in detail.

The industrially significant dynamics of growing CZ crystals free of large microdefects is also reviewed. Under the conditions of large microdefect-free growth, a moderate vacancy supersaturation develops in the vicinity of the lateral surface of a growing crystal, leading to the formation of oxygen clusters and small voids, at lower temperatures. The vacancy incorporation near the lateral surface of a crystal, or the *lateral incorporation of vacancies*, is driven by the interplay among the Frenkel reaction, the diffusion of the intrinsic point defects, and their convection.

A review of CZ defect dynamics with a particular focus on the growth of large microdefect-free crystals is presented and discussed.

38.1 **The Discovery of Microdefects** 1283

38.2 **Defect Dynamics in the Absence of Impurities** 1284
38.2.1 The Theory of the Initial Incorporation of Intrinsic Point Defects 1284
38.2.2 The Quantification of the Microdefect Formation 1290

38.3 **Czochralski Defect Dynamics in the Presence of Oxygen** 1304
38.3.1 Reactions in Growing CZ Crystals... 1304
38.3.2 The Model 1305
38.3.3 Defect Dynamics in One-Dimensional Crystal Growth.. 1308
38.3.4 Defect Dynamics in Two-Dimensional Crystal Growth . 1310

38.4 **Czochralski Defect Dynamics in the Presence of Nitrogen** 1313
38.4.1 The Model 1313
38.4.2 CZ Defect Dynamics in One-Dimensional Crystal Growth.. 1316

38.4.3 CZ Defect Dynamics in Two-Dimensional Crystal Growth . 1318
38.5 The Lateral Incorporation of Vacancies in Czochralski Silicon Crystals 1321
38.5.1 General Defect Dynamics: A Brief Revisit 1322
38.5.2 Defect Dynamics Under Highly Vacancy-Rich Conditions 1323
38.5.3 Defect Dynamics Near the Critical Condition 1324
38.6 Conclusions 1328
38.6.1 CZ Defect Dynamics in the Absence of Impurities 1329
38.6.2 CZ Defect Dynamics in the Presence of Oxygen 1330
38.6.3 CZ Defect Dynamics in the Presence of Nitrogen 1330
38.6.4 The Lateral Incorporation of Vacancies 1331
References 1332

In modern microelectronics industry, the majority of devices are fabricated on silicon substrates produced from silicon crystals grown by the Czochralski (CZ) process. The float-zone (FZ) process is also used to produce a small fraction of the silicon substrates used in the modern industry. In the CZ process, a silicon crystal is continuously pulled from a silicon melt placed in a quartz crucible, as shown in Fig. 38.1. A CZ crystal has a conical top called the *crown*, a conical bottom called the *endcone*, and a cylindrical section called the *body*, which provides substrates for the fabrication of devices. In the FZ process, a molten silicon zone is allowed to form and solidify along the entire length of a polycrystalline silicon body to form a silicon single crystal. Silicon crystals grown prior to the late 1950s contained thermomechanically induced dislocations. A breakthrough discovery by *Dash* in the late 1950s allowed crystal growth free of these dislocations [38.2, 3]. In modern crystals, free of thermomechanically induced dislocations, various crystallographic imperfections known as microdefects, aggregated defects, and grown-in bulk defects can form and grow [38.4–15]. Microdefects vary in size from a few nanometers to > 200 nm. Microdefects of this size significantly affect the performance of the continuously shrinking modern microelectronic devices. This chapter addresses key developments in the crystal growth industry in the understanding and quantification of the physics of the formation of microdefects, popularly termed *defect dynamics*.

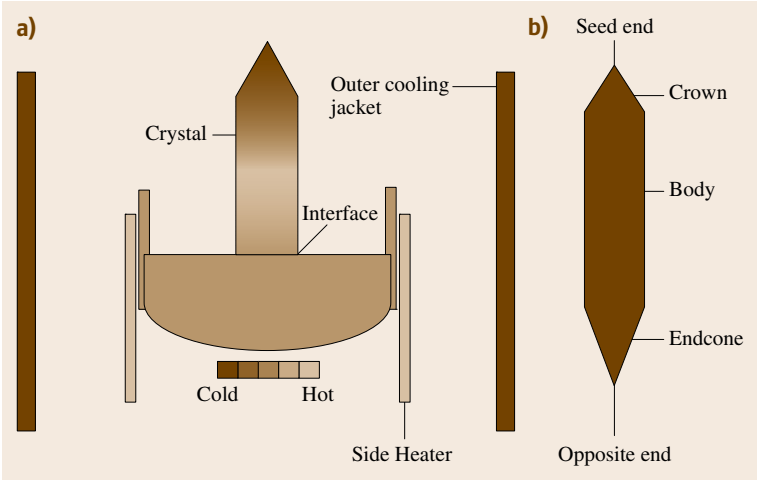


Fig. 38.1 (a) A schematic representation of the Czochralski (CZ) crystal growth process, (b) a grown CZ crystal (after [38.1])

38.1 The Discovery of Microdefects

Various types of microdefects are formed in CZ and FZ crystals. However, two types of microdefects commonly occur and hence are of critical significance. These are the aggregates of the intrinsic point defects of silicon: self-interstitials and vacancies. Vacancies are formed by silicon atoms missing from the silicon crystal lattice, whereas self-interstitials are silicon atoms that are not part of the lattice. Self-interstitial aggregates in FZ crystals were first observed in the 1960s and were termed *A defect swirls* (*A defects*) and *B defect swirls* (*B defects*), although the origin of these defects was not clearly known [38.4, 6–9]. Later, A defects were identified as dislocation loops and B defects were presumed to be globular self-interstitial aggregates [38.10–13]. In the 1980s, vacancy aggregates in FZ crystals were first reported and later termed *D defects* [38.14, 15]. More recent studies have identified D defects as octahedral voids [38.17, 18].

The discovery of the described microdefects was followed by various attempts to explain their formation. The dependence of the type of microdefects formed on the crystal growth rate was reported in various papers [38.6, 8, 14, 15]. A and B defects were observed at lower growth rates while D defects were observed at higher growth rates (Fig. 38.2). A unifying and acceptable analysis of the microdefect distribution in silicon crystals, however, was not available until 1982, when *Voronkov* provided the first groundbreaking explanation of defect dynamics in both FZ and CZ growth [38.19].

Various impurities present in a silicon crystal can influence the defect dynamics. CZ crystals are grown

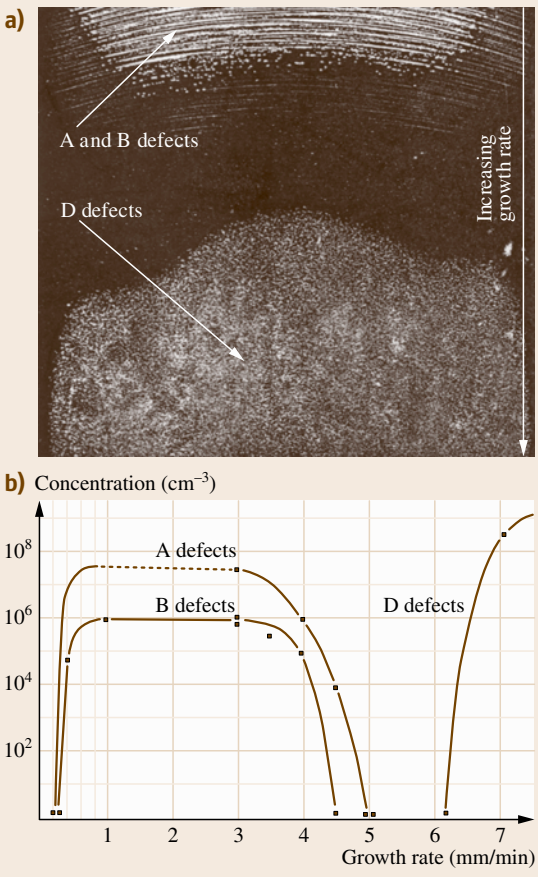


Fig. 38.2 **(a)** X-ray transmission topograph of a copper-decorated longitudinal section of a pedestal-pulled crystal, showing the dependence of the microdefect type on the growth rate. The growth rate varies from 4 to 7 mm/min from top to bottom. Note that the morphological details of microdefects are not revealed by the decoration of microdefects by the precipitation of copper. **(b)** The dependence of the type and density of microdefects on the crystal growth rate (after [38.14], © Elsevier 1981)

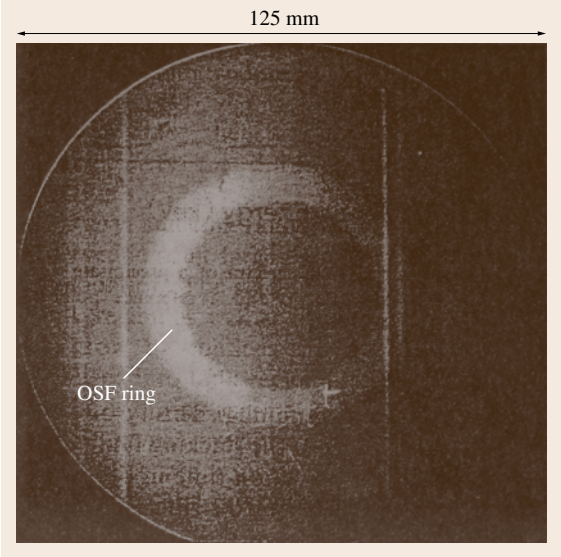


Fig. 38.3 An x-ray topograph (positive image) of a wafer showing the annular region containing OSFs. The wafer was sliced from a CZ crystal and treated at 1100 °C for 1 h in steam (after [38.16] © Jpn. J. Appl. Phys.)

from a melt contained in a quartz crucible, which is a source of oxygen. Hence, CZ crystals contain oxygen, unlike FZ crystals. Oxygen in CZ crystals typically precipitates as particles of silicon dioxide, which are popularly termed *oxygen precipitates* or *oxygen clusters*. Oxygen clusters in a growing CZ crystal are typically very small. These clusters facilitate the formation of stacking faults in crystals subjected to selective heat treatments that generate self-interstitials. Hence, oxygen clusters are typically identified by these stacking faults, known as oxidation-induced stacking faults (OSFs) (Fig. 38.3) [38.16]. As microdefects in silicon substrates produced from CZ crystals can adversely affect the performance of devices built on them, the development of crystal growth processes that reduce the size of microdefects in growing crystals is of industrial significance. Various studies have shown that the microdefect distribution in CZ crystals can be influenced in the presence of nitrogen, in general, and that the size of D defects (voids) can be reduced, in particular [38.20]. Dopants such as boron also affect CZ defect dynamics [38.21–24].

38.2 Defect Dynamics in the Absence of Impurities

Considering the complexity of CZ defect dynamics, first CZ and FZ defect dynamics in the absence of impurities are discussed in this chapter, followed by CZ defect dynamics in the presence of nitrogen and oxygen.

38.2.1 The Theory of the Initial Incorporation of Intrinsic Point Defects

Voronkov was the first to provide a satisfactory and well-accepted explanation for the quality of the microdefect distributions observed in silicon crystals [38.19]. According to Voronkov's theory, an interplay between the transport of the intrinsic point defects in a growing crystal in the vicinity of the melt–crystal interface and the Frenkel reaction involving the mutual annihilation or the recombination of vacancies and self-interstitials plays a key role in determining the final microdefect distribution. Voronkov analyzed this interplay by quantifying the Frenkel reaction dynamics and the intrinsic point defect balances in a growing crystal. His groundbreaking work is briefly discussed below.

The Frenkel Reaction Dynamics in a Growing Crystal

The Frenkel reaction involves the reversible annihilation of vacancies and self-interstitials by their recombination and the spontaneous generation of vacancies and self-interstitials from silicon lattice atoms

$$\mathrm{i} + \mathrm{v} \rightleftharpoons \mathrm{Si}\,, \tag{38.1}$$

where i is a self-interstitial, v is a vacancy, and Si is a silicon lattice atom. The net rate of recombination of vacancy and self-interstitial pairs is equal and opposite to the net rate of formation of self-interstitials or vacancies, which is given by

$$-r_\mathrm{i} = -r_\mathrm{v} = k_{\mathrm{i}\leftrightarrow\mathrm{v}}(C_\mathrm{i}C_\mathrm{v} - C_{\mathrm{i,e}}C_{\mathrm{v,e}})\,, \tag{38.2}$$

where r is the net rate of formation of any species per unit volume, C is the concentration of any species, and $k_{\mathrm{i}\leftrightarrow\mathrm{v}}$ is the reaction rate constant known as the recombination constant. The subscript "i" denotes self-interstitials, "v" vacancies, "e" equilibrium conditions, and "i ↔ v" the interaction between self-interstitials and vacancies. The intrinsic point defects are assumed to exist at equilibrium at the melt–crystal interface. The concentration of the intrinsic point defects, however, drops significantly in a crystal near the interface, as the decreasing temperature in the crystal facilitates rapid recombination of self-interstitials and vacancies. Hence, the intrinsic point defects diffuse from the interface into the crystal. In addition, the continuous growth of the crystal, or the crystal pulling, facilitates the transport of the intrinsic point defects by their convection. Compared with the timescale of the transport, the Frenkel reaction can be assumed to be instantaneous and, hence, in equilibrium. Thus, the following relationship remains valid

$$C_\mathrm{i}C_\mathrm{v} = C_{\mathrm{i,e}}C_{\mathrm{v,e}}\,. \tag{38.3}$$

The equilibrium concentrations of the intrinsic point defects are functions of the crystal temperature and are expressed as

$$C_{\mathrm{i,e}} = C_{\mathrm{i},0}\exp\left(-\frac{E_{\mathrm{form,i}}}{k_\mathrm{B}T}\right)\,, \tag{38.4}$$

$$C_{\mathrm{v,e}} = C_{\mathrm{v},0}\exp\left(-\frac{E_{\mathrm{form,v}}}{k_\mathrm{B}T}\right)\,, \tag{38.5}$$

where E_form is the formation energy (the subscript "form" denotes the formation energy) and k_B is the

Boltzmann constant. The temperature profile in a crystal, in the vicinity of the melt–crystal interface and at a fixed radial location, is reasonably represented by a linear approximation of the inverse of the temperature

$$\frac{1}{T} = \frac{1}{T_{s/l}} + \frac{1}{T_{s/l}^2} G z \,, \tag{38.6}$$

where T is the temperature, G is the magnitude of the axial temperature gradient at the interface at any radial location, and z is the distance from the interface. The subscript "s/l" represents the conditions at the interface. Using (38.6), (38.4) and (38.5) can be written as

$$C_{i,e} = C_{i,e}\,(T_{s/l}) \exp\left(-\frac{E_{form,i} G z}{k_B T_{s/l}^2}\right) , \tag{38.7}$$

$$C_{v,e} = C_{v,e}\,(T_{s/l}) \exp\left(-\frac{E_{form,v} G z}{k_B T_{s/l}^2}\right) . \tag{38.8}$$

Thus, (38.3) describing the Frenkel reaction equilibrium takes the form

$$C_i C_v = C_{i,e}\,(T_{s/l})\, C_{v,e}\,(T_{s/l}) \exp\left(\frac{-2z}{l}\right) , \tag{38.9}$$

where the *characteristic recombination length* l is expressed as

$$l = \frac{2 k_B T_{s/l}^2}{(E_{form,i} + E_{form,v}) G} . \tag{38.10}$$

The discussed Frenkel reaction dynamics can be coupled with the overall intrinsic point defect balance in a growing crystal to describe the microdefect distributions observed in CZ and FZ crystals.

The Intrinsic Point Defect Balance and the Initial Incorporation

Driven by the Frenkel reaction, the concentrations of both the intrinsic point defect species dramatically drop in the vicinity of the melt–crystal interface. This intrinsic point defect concentration drop, in turn, drives the diffusion of both species from the interface, where they exist at equilibrium, into the crystal. In addition, the intrinsic point defects are also transported by the physical growth of the crystal, with respect to a fixed coordinate system. Crystal growth at a fixed rate and through a fixed temperature field with respect to a fixed coordinate system can be assumed to take place under a steady state. The excess intrinsic point defect flux, defined as the difference between the flux of vacancies and the flux of self-interstitials, is not explicitly affected by the Frenkel reaction. This excess flux is fixed for a given crystal growth condition

$$j_{iv} = \left(-D_v \frac{dC_v}{dz} + V C_v\right) - \left(-D_i \frac{dC_i}{dz} + V C_i\right) , \tag{38.11}$$

where j_{iv} is the excess intrinsic point defect flux, D is the diffusivity, and V is the magnitude of the axial crystal pull rate. The effects of radial diffusion are ignored in (38.11).

When the vacancy flux is greater than the self-interstitial flux or when the excess intrinsic point defect flux is positive, vacancies are the surviving dominant species in the growing crystal and self-interstitials are annihilated to very low concentrations, within a short distance from the melt–crystal interface. Vacancy supersaturation increases at lower temperatures to drive the formation of D defects. Vacancies are termed the *incorporated dominant intrinsic point defects* under this condition. When the self-interstitial flux is greater than the vacancy flux or when the excess intrinsic point defect flux is negative, self-interstitials are the dominant species in the crystal and vacancies are annihilated to very low concentrations, within a short distance from the interface. The crystal becomes supersaturated with self-interstitials, leading to the formation of B and A defects by the aggregation of self-interstitials at lower temperatures. Self-interstitials are termed the *incorporated dominant intrinsic point defects* under this condition.

When the vacancy flux is equal to the self-interstitial flux or when the excess intrinsic point defect flux is equal to zero, both the intrinsic point defect species remain in comparable concentrations in the growing crystal and annihilate each other to very low concentrations. Under this condition, defined as the *critical condition*, no detectable microdefects are formed in the crystal at any temperature. At the critical condition, (38.11) is satisfied when both the intrinsic point defect concentrations show the same dependence on z, if the diffusivities of both the species are fixed and temperature independent. This assumption is approximate but acceptable within a narrow range of temperatures close to the interface. Then, using (38.9), the intrinsic point defect concentrations in a crystal growing under the critical condition are defined as

$$C_i = C_{i,0} \exp\left(\frac{E_{form,i} + E_{form,v}}{2 k_B T_{s/l}} - \frac{E_{form,i}}{k_B T_{s/l}} - \frac{E_{form,i} + E_{form,v}}{2 k_B T}\right) , \tag{38.12}$$

$$C_v = C_{v,0} \exp\left(\frac{E_{form,i} + E_{form,v}}{2k_B T_{s/l}} - \frac{E_{form,v}}{k_B T_{s/l}} - \frac{E_{form,i} + E_{form,v}}{2k_B T}\right) . \quad (38.13)$$

Using (38.11–38.13), the critical condition is analytically derived as

$$(V/G)_c = \frac{E_{form,i} + E_{form,v}}{2k_B T_{s/l}^2} \times \frac{D_i(T_{s/l})C_{i,e}(T_{s/l}) - D_v(T_{s/l})C_{v,e}(T_{s/l})}{C_{v,e}(T_{s/l}) - C_{i,e}(T_{s/l})} , \quad (38.14)$$

where the subscript "c" denotes the critical value at zero excess intrinsic point defect flux.

Using this analysis, Voronkov explained the dependence of the microdefect distribution in a crystal on its growth rate (Fig. 38.2). At very high V, the convection dominates the diffusion. He hypothesized that the flux of vacancies into the crystal is higher than the flux of self-interstitials when the species convection is relatively appreciable, because the concentration of vacancies is higher than the concentration of self-interstitials at the interface. The crystal remains vacancy rich as the temperature drops. At very low V, the diffusion dominates the convection; Voronkov hypothesized that the flux of self-interstitials is greater than the flux of vacancies when the species diffusion is relatively appreciable, because self-interstitials diffuse faster than vacancies at higher temperatures. Thus, self-interstitials become the dominant incorporated species within a short distance from the interface, while vacancies are effectively annihilated. The dominant incorporated species eventually nucleates to form the appropriate microdefects at lower temperatures.

The competition between the intrinsic point defect convection and the intrinsic point defect diffusion is quantified not just by the crystal pull rate, but by the ratio of the crystal pull rate to the magnitude of the axial temperature gradient V/G. The convection of the intrinsic point defects increases with the crystal pull rate. The diffusion flux of an intrinsic point defect species increases with the increasing magnitude of its concentration gradient, which is driven by the temperature gradient near the interface. Thus, vacancies become the dominant incorporated intrinsic point defect species at higher V/G and self-interstitials become the dominant incorporated species at lower V/G. At the *critical* V/G, the flux of vacancies is equal to the flux of self-interstitials and there is no dominance of either intrinsic point defect species as both species mutually annihilate each other to very low concentrations.

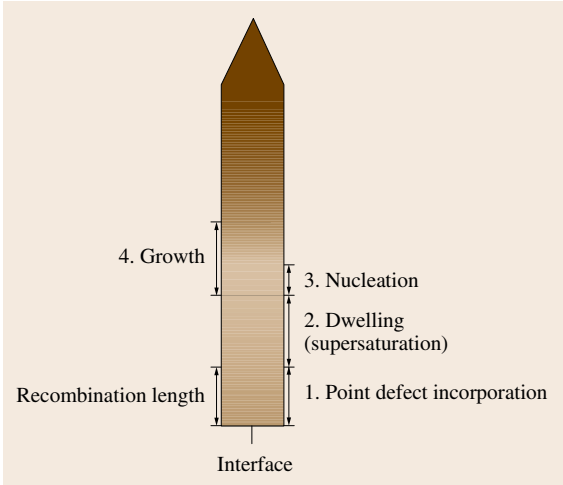


Fig. 38.4 The phases of defect dynamics in a growing crystal according to Voronkov's theory (not to scale) (after [38.25])

Table 38.1 Key properties of various species participating in reactions in growing CZ crystals and the recombination rate constant (after [38.1])

Property set I
$D_i\,(\mathrm{cm^2/s}) = 0.19497 \exp\left(\frac{-0.9\,(\mathrm{eV})}{k_B T}\right)$
$D_v\,(\mathrm{cm^2/s}) = 6.2617 \times 10^{-4} \exp\left(\frac{-0.4\,(\mathrm{eV})}{k_B T}\right)$
$C_{i,e}\,(\mathrm{cm^{-3}}) = 6.1859 \times 10^{26} \exp\left(\frac{-4.0\,(\mathrm{eV})}{k_B T}\right)$
$C_{v,e}\,(\mathrm{cm^{-3}}) = 7.59982 \times 10^{26} \exp\left(\frac{-4.0\,(\mathrm{eV})}{k_B T}\right)$
$k_{i\leftrightarrow v}\,(\mathrm{cm^3/s}) = 1.2 \times 10^{-6}\,[D_i(T) + D_v(T)] \times \exp\left[-\frac{0.61 + \left(-2.30 + 7.38 \times 10^{-3} T\right) k_B T}{k_B T}\right]^*$
$\lambda_i\,(\mathrm{eV}) = 2.95^{**}$
$\lambda_v\,(\mathrm{eV}) = 1.85$

* Reported by *Sinno* et al. [38.26]; the enthalpic barrier (0.61) is ignored in the simulations discussed in this chapter [38.1, 27]. An accurate estimation of this parameter is not necessary, because the Frenkel reaction dynamics is fast

** For two-dimensional (2-D) simulations, a value of 2.85 eV is used

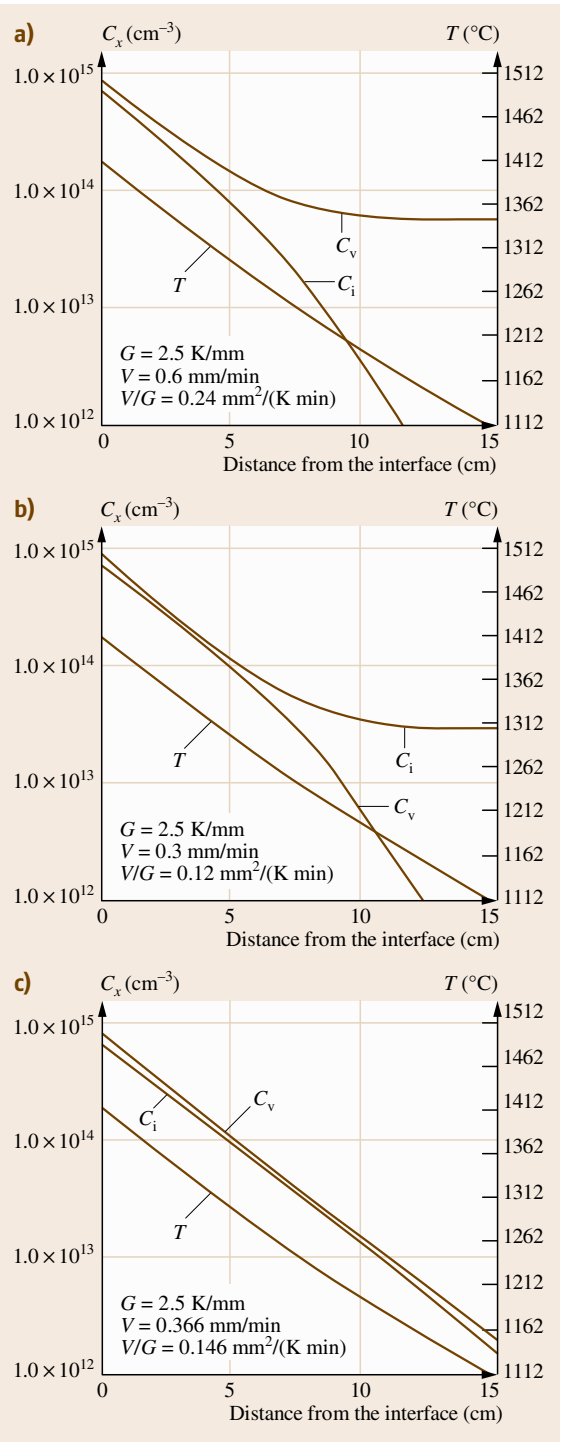


Fig. 38.5a–c The profiles of the vacancy concentration, the self-interstitial concentration, and the temperature in a Czochralski crystal growing at **(a)** a higher V/G, **(b)** a lower V/G, **(c)** close to the critical V/G (after [38.25]) ◀

The process of establishing the key intrinsic point defect concentration field in a growing CZ crystal in the vicinity of the melt–crystal interface is termed the *initial incorporation*. The initial incorporation takes place within a short distance from the interface, which scales with the characteristic recombination length l. When the operating V/G is either higher or lower than the critical V/G, the dominance of one intrinsic point defect species is established within the recombination length; when the operating V/G is closer to the critical V/G, however, both the intrinsic point defect species are incorporated in comparable concentrations and continue to recombine and annihilate each other beyond the recombination length, without forming large microdefects. A schematic representation of Voronkov's theory is shown in Fig. 38.4. Voronkov's theory can be verified by quantifying the intrinsic point defect concentration profiles ($C_x(z)$, x denoting either i or v) in a CZ crystal growing at various V/G, as shown in Fig. 38.5a–c. These computations were accomplished by assuming that both the diffusivities and the equilibrium concentrations of the intrinsic point defects vary with the temperature and by accounting for the Frenkel reaction kinetics [38.25]. The properties of the intrinsic point defects used in these simulations are listed in Table 38.1. There is significant uncertainty in the estimation of the properties of the intrinsic point defects. The parameters describing the Frenkel reaction kinetics, in particular, are not well known. An accurate estimation of the kinetic parameters is not necessary because the Frenkel reaction dynamics is very fast and, hence, reaction equilibrium prevails. It must be noted that the intrinsic point defect concentration profiles computed assuming the constant (fixed) intrinsic point defect diffusivities give slightly different results.

The evolution of the concentrations of the intrinsic point defects in a growing crystal strongly depends on the properties of the intrinsic point defects, especially the formation and migration energies. As shown by (38.12) and (38.13), at the critical V/G, the concentration of each intrinsic point defect species is equal to its equilibrium concentration at any axial location z, when the formation energy of a vacancy is equal to

the formation energy of a self-interstitial, the migration energies of both species are zero, and the temperature profile in the crystal is given by (38.6). Deviations from these conditions establish the nonequilibrium intrinsic point defect concentrations in a crystal growing at the critical V/G. The dynamics of crystal growth near the critical condition and the effects of the properties of the intrinsic point defects on this dynamics are discussed in detail in Sect. 38.5.

The significance of the discussed discovery of the initial intrinsic point defect incorporation in growing silicon crystals extends beyond the field of CZ and FZ defect dynamics; it remains valuable in the prediction of the properties of the intrinsic point defects of silicon by requiring that the difference $C_{v,e}(T_{s/l}) - C_{i,e}(T_{s/l})$ and the difference $D_i(T_{s/l})C_{i,e}(T_{s/l}) - D_v(T_{s/l})C_{v,e}(T_{s/l})$ be positive.

Validation and Limits of the Theory of One-Dimensional Initial Incorporation

Voronkov's theory of the intrinsic point defect incorporation has been validated by many experimental observations reported before and after its publication, in spite of a few initial questions raised by *Tan* and *Gösele*, among others [38.28]. The dependence of the microdefect type in FZ crystals on the crystal growth rate was reported in [38.6, 8, 14, 15]. By varying the pull rate of various CZ crystals, *Sadamitsu* et al. showed that the microdefect quality in the crystals shifted from vacancy type (vacancy aggregates) to self-interstitial type (self-interstitial aggregates) [38.29]. This study also revealed a radial variation in the microdefect distribution. This was explained by the radial variation in the temperature field, or the radial variation of G. In a typical CZ crystal, G monotonically increases and V/G monotonically decreases along the radial position. Thus, the central regions in many CZ crystals exhibit D defects and the peripheral regions exhibit A defects, as determined by V/G. The narrow microdefect-free region between the region of vacancy aggregates and the region of self-interstitial aggregates is known as the v/i boundary. The microdefect distributions in CZ crystals showing the v/i boundary were reported by many in the last decade [38.30–33]

The surface of a growing CZ crystal acts as a source or a sink of the intrinsic point defects and, hence, induces their radial diffusion; in addition, the radial variation in the temperature field also causes radial diffusion of the intrinsic point defects. The microdefect distributions away from the surfaces in rapidly pulled crystals in which the radial diffusion effects can be ignored, however, are very well explained by the one-dimensional initial incorporation theory. In a silicon crystal, the region containing abundant vacancies or vacancy aggregates is termed the *v-rich region* and the region containing abundant self-interstitials or the self-interstitial aggregates is termed the *i-rich region*. The position of the boundary between the v-rich region and the i-rich region established after the initial incorporation remains essentially the same even after the aggregation of the intrinsic point defects in a rapidly pulled CZ crystal. Hence, by following the v/i boundary separating the vacancy aggregates and the self-interstitial aggregates, in such crystals, the critical V/G can be experimentally determined. Typically, a narrow annular region inside the v-rich region of a CZ crystal near the v/i boundary exhibits another type of microdefects known as oxidation-induced stacking faults (OSFs), after a treatment with selective heat cycles. These OSFs are formed by the oxides of silicon, formed during the crystal growth, in the regions grown at V/G slightly above the critical V/G (Figs. 38.3 and 38.6). The incorporated vacancies in this region are too low in concentration to nucleate at higher temperatures ($\approx 1100\,^{\circ}\mathrm{C}$) to form D defects during crystal growth; they survive at lower temperatures (≈ 1000–$850\,^{\circ}\mathrm{C}$) to facilitate the formation of silicon oxide (primarily silicon dioxide) particles. Oxygen required for this oxidation in CZ crystals comes from the quartz crucible used in the CZ process. The specific volume of an oxide particle is greater than the specific volume of silicon. Hence, the formation and growth of the oxide particles generates compressive stresses. Relief of this stress can take place by the consumption of vacancies and the ejection of self-interstitials from the silicon lattice. During crystal growth, the oxide particles in the OSF region are formed essentially by facilitation by vacancies. The growth of the oxide particles by the ejection of self-interstitials can take place after the depletion of free vacancies at lower temperatures. These compressed oxide particles in a silicon wafer facilitate the growth of stacking faults or OSFs after selective heat treatments that inject self-interstitials into the silicon wafer. *Hasebe* et al. were among the first to report the presence of the annular region of OSFs termed the *OSF ring* in CZ crystals (Fig. 38.3) [38.16]. Often the location of the *OSF ring*, because of its vicinity to the v/i boundary, is used to mark the critical V/G. The CZ defect dynamics in the presence of oxygen describing the formation and growth of oxygen clusters is discussed later in this chapter.

The radial diffusion of the intrinsic point defects in a CZ crystal, both driven by the variation in the crys-

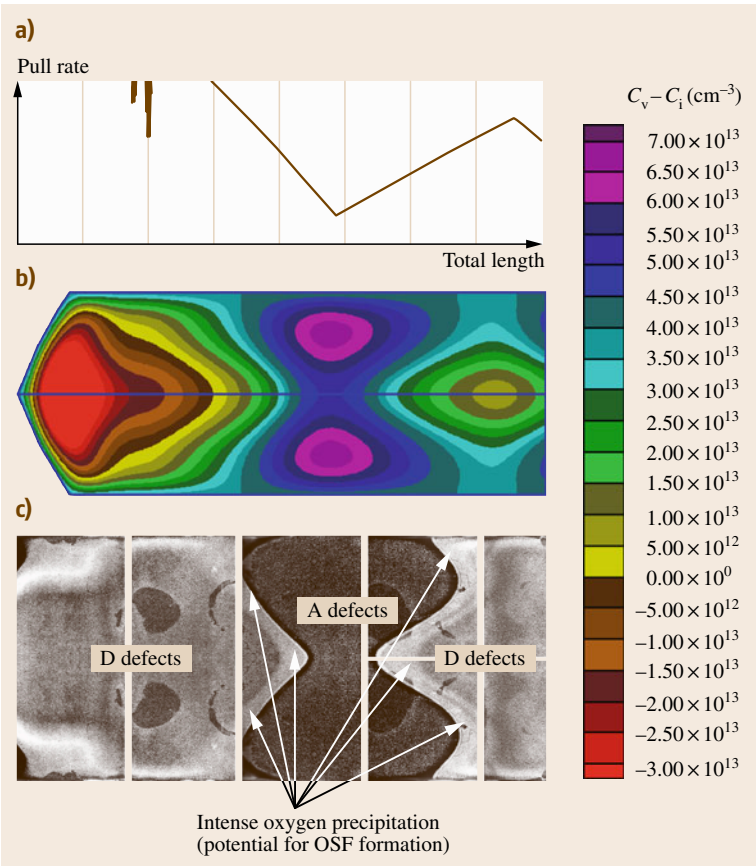


Fig. 38.6 **(a)** The pull-rate profile used to pull the experimental crystal. **(b)** The predicted excess intrinsic point defect concentration field in the crystal. **(c)** The experimentally observed microdefect distribution. The radial variation of G at the interface was more than 140%. Note: The simulation results are not completely mesh insensitive (after [38.25])

tal temperature field and induced by the lateral surface of the crystal, cannot be ignored in modern CZ growth. Thus, a more complete picture of the intrinsic point defect dynamics is described by considering the transport of the intrinsic point defects in an axisymmetric crystal growing through a temperature field at pseudo-steady state

$$\frac{\partial C_i}{\partial t} + V\frac{\partial C_i}{\partial z} = \nabla\left(D_i \nabla C_i\right) - k_{i\leftrightarrow v}\left(C_i C_v - C_{i,e} C_{v,e}\right), \quad (38.15)$$

$$\frac{\partial C_v}{\partial t} + V\frac{\partial C_v}{\partial z} = \nabla\left(D_v \nabla C_v\right) - k_{i\leftrightarrow v}\left(C_i C_v - C_{i,e} C_{v,e}\right). \quad (38.16)$$

Nakamura et al. and *Sinno* et al. were among the first to solve the two-dimensional intrinsic point defect dynamics in growing CZ crystals to clearly define the final microdefect distributions [38.26, 34].

The equations describing the intrinsic point defect distributions in a CZ crystal can be further simplified by assuming the Frenkel reaction equilibrium. The evolution of the excess intrinsic point defect concentration $C_v - C_i$ is described without an explicit use of the Frenkel reaction kinetics as

$$\frac{\partial\left(C_v - C_i\right)}{\partial t} + V\frac{\partial C_v}{\partial z} - V\frac{\partial C_i}{\partial z} = \nabla\left(D_v \nabla C_v\right) - \nabla\left(D_i \nabla C_i\right). \quad (38.17)$$

The intrinsic point defect concentration field is then predicted by the solution of (38.17) and (38.3). The concentrations of the intrinsic point defects at the boundary of the crystal including the melt–crystal interface can be assumed to be at equilibrium. The excess intrinsic point defect concentration field can be mapped to the final microdefect distribution. A comparison between the predicted excess intrinsic point defect concentration field and the experimentally observed microdefect distribution in a crystal pulled by a varying rate in MEMC Electronic Materials is shown in Fig. 38.6 [38.25]. The crystal temperature field used in the simulation was

first predicted by the commercial software MARC and then corrected using the experimentally measured interface [38.35–37]. The heat transport dynamics was assumed to be very fast, and the melt was assumed to be an effective solid [38.35–37]. The shape of the interface changes with the crystal pull rate. It is reasonable, however, to assume that the interface remains fixed when the pull rate variation is moderate or that there exists a locally fixed effective interface when the pull rate variation is significant. Typically, a representative interface measured in the center of the region of interest suffices. The simulation was performed using one such experimentally measured interface. The properties of the intrinsic point defects used in the simulation are listed in Table 38.1. Figure 38.6a shows the actual crystal pull rate profile, and Fig. 38.6b shows the predicted excess intrinsic point defect concentration field in the crystal. The positive excess intrinsic point defect concentration represents vacancy aggregates or D defects, and the negative excess intrinsic point defect concentration represents self-interstitial aggregates (A and B defects). The experimental microdefect distribution is determined by the method of copper decoration and Secco etching and is shown in Fig. 38.6c. The reader is referred to *Kulkarni* et al. for the details of the microdefect decoration by the employed characterization technique [38.38–40]. The excess intrinsic point defect concentration field shown in Fig. 38.6 is influenced by the varying pull rate, the temperature field in the crystal, and the radial intrinsic point defect diffusion, to some extent. The v/i boundary in the crystal is clearly identified. The intensity of the oxygen precipitation is very high very close to the v/i boundary, indicating the potential for OSF formation. In rapidly pulled crystals, the observed microdefect distribution, the location of the OSF region, and the location of the v/i boundary can be quantified by the radial variation of G and the axial variation of the pull rate. Several studies [38.30–33] have directly or indirectly quantified the location of the v/i boundary in CZ crystals. It must be noted that the simulation results shown in Fig. 38.6 are not completely mesh insensitive at the level of discretization used. This inaccuracy is corrected in the computations discussed in the later sections of this chapter.

38.2.2 The Quantification of the Microdefect Formation

The quantification of the intrinsic point defect concentration field in a CZ crystal provides only a qualitative information of the final microdefect distribution [38.19, 26, 34, 41–45]. As microelectronic devices are very sensitive to the size of microdefects, an accurate quantification of the microdefect distribution in CZ crystals is essential. Capturing the distribution of microdefects in a CZ crystal is quite rigorous and involves the solution of a set of complex equations. The formation and growth of microdefects in an element of silicon takes place over a finite period of time; thus, a population of the microdefects formed at various moments of the elapsed time period exists in the element. Hence, various approximations are made in capturing the entire microdefect distribution in a growing CZ crystal [38.1, 27, 46–54].

Based on the research reported in the literature, three broad approaches for the quantification of the microdefect distribution in a CZ crystal can be identified. The first approach involves the application of the classical nucleation theory for the formation of stable nuclei of microdefects followed by their diffusion-limited growth, the second approach involves representing the population of microdefects present at any location by an equivalent population of identical microdefects, and the third approach involves the application of the Fokker–Planck equation to create a continuum of the microdefect size. *Voronkov* and *Falster* were the first to publish the first approach [38.46]. For the sake of simplicity, however, they ignored axial and radial diffusion of the intrinsic point defects during the formation and growth of microdefects. Thus, in effect, the formation of microdefects was treated in an isolated element of silicon following a predetermined decrease in the temperature. *Kulkarni* et al. [38.1] augmented this model and quantified the CZ defect dynamics including the axial diffusion effects [38.1]. A detailed two-dimensional treatment of the CZ defect dynamics to capture the microdefect distribution by the approach of *Kulkarni* et al. [38.1] remains computationally unattractive. To address this shortcoming, *Kulkarni* and *Voronkov* [38.53] developed the *lumped model*, which represents a population of microdefects of varying sizes at any location in a CZ crystal by an equivalent population of identical microdefects and captures the two-dimensional distribution of microdefects.

The quantification of the CZ defect dynamics by a more rigorous approach can be accomplished by a direct treatment of the reactions between the nucleating monomers and microdefects of various sizes. This approach is computationally impractical, considering the size of mature microdefects. The Fokker–Planck equations (FPE) are popularly used to address the

quantification of such systems. By the Fokker–Planck approach, a microdefect population at any location in a CZ crystal is quantified by a *size* coordinate. *Sinno* and *Brown* [38.47], *Mori* [38.27], *Wang* and *Brown* [38.48], and *Brown* et al. [38.49] accomplished the quantification of the defect dynamics in CZ crystals using the Fokker–Planck formulation.

For the sake of continuity, the quantifications of the defect dynamics in CZ crystals by the first and the second approach are discussed first in this chapter. The third approach, employing the Fokker–Planck equations, is discussed last.

First Approach: CZ Defect Dynamics Using the Classical Nucleation Theory

All microdefects are treated as clusters of the intrinsic point defects. The formation of microdefects takes place by a series of reactions of the following type [38.55–57]:

$$P_{m_x} + x \rightleftharpoons P_{(m+1)_x} , \quad \{x = \mathrm{i}, \mathrm{v}\} . \tag{38.18}$$

In reaction (38.18), P is a cluster of monomers or an intrinsic point defect species x. A cluster of m intrinsic point defects of type x is represented by P_{m_x}. At a given temperature and intrinsic point defect concentration, the total free energy change associated with the formation of a cluster containing m intrinsic point defects from a supersaturated solution is given by

$$\Delta F_x\,(m_x) = -m_x k_\mathrm{B} T \ln \frac{C_x}{C_{x,\mathrm{e}}} + \lambda_x m_x^{2/3} , \tag{38.19}$$

where F is the free energy and λ is the surface energy coefficient for the cluster. Note that the subscript x (for i and v) denotes the intrinsic point defect x (as in C_x) as well as the clusters containing the intrinsic point defect x (as in F_x and λ_x), depending on the variable. The first term on the right-hand side of (38.19) is the bulk (volume) free energy change (per m intrinsic point defects) associated with the intrinsic point defect supersaturation, and the second term is the cluster surface energy associated with the formation of the new cluster surface. The number of intrinsic point defects in the so-called *critical cluster* is obtained by maximizing the free energy change with respect to m

$$m_x^* = \left(\frac{2\lambda_x}{3k_\mathrm{B} T \ln \frac{C_x}{C_{x,\mathrm{e}}}} \right)^3 . \tag{38.20}$$

The maximum free-energy change associated with the formation of the critical clusters containing m^* intrinsic point defects is interpreted as the nucleation barrier and is given by the substitution of m^* into (38.19)

$$\Delta F_x(m_x^*) = \Delta F_x^* = \frac{4}{27} \frac{\lambda_x^3}{\left(k_\mathrm{B} T \ln \frac{C_x}{C_{x,\mathrm{e}}}\right)^2} . \tag{38.21}$$

The classical nucleation theory gives the rate of formation of stable supercritical nuclei or clusters per unit volume, defined as the *nucleation rate*, as a function of various properties of the critical clusters and the intrinsic point defects as [38.57]

$$J_x = \eta_{x\leftrightarrow x}^* \times Z_x \times \phi_{x,\mathrm{e}}^* , \tag{38.22}$$

where J_x is the nucleation rate of clusters of type x, or x-clusters, $\eta_{x\leftrightarrow x}^*$ is the attachment frequency of monomer x to the critical clusters of x, $\phi_{x,\mathrm{e}}^*$ is the equilibrium density of the critical clusters, and Z_x is the Zheldovich factor. Clusters are assumed to be spherical in shape. The relevance and accuracy of this assumption are discussed in the following section. Using the classical expressions for the Zheldovich factor and the equilibrium concentration of the critical clusters, the nucleation rate is given as

$$\begin{aligned} J_x = {} & \left[4\pi D_x C_x \times R_x \left(m_x^*\right)\right] \\ & \times \left[\left\{12\pi k_\mathrm{B} T \times \Delta F_x \left(m_x^*\right)\right\}^{-\frac{1}{2}} \left(k_\mathrm{B} T \ln \frac{C_x}{C_{x,\mathrm{e}}}\right) \right] \\ & \times \left[\rho_x \exp\left(-\frac{\Delta F_x \left(m_x^*\right)}{k_\mathrm{B} T} \right) \right] . \end{aligned} \tag{38.23}$$

Here R is the radius (size) of a cluster, D is the diffusivity of any species, and ρ is the site density for nucleation. The first term in the square brackets on the right-hand side of (38.23) is the diffusion-limited intrinsic point defect attachment frequency to a critical cluster, the second term is the Zheldovich factor, and the third term is the equilibrium concentration of the critical clusters.

The Model. In this study, all microdefects are treated as spherical aggregates of either vacancies or self-interstitials and are termed *clusters*. Vacancy aggregates are termed v-clusters and represent D defects; self-interstitial aggregates are termed i-clusters and represent A and B defects. Since D defects are known to be octahedral voids, their approximation as spherical clusters is reasonably accurate [38.17, 18]. A defects are dislocation loops [38.10–13]. They presumably form, however, from globular self-interstitial aggregates

such as B defects [38.12, 19]. Therefore, the approximation of A defects as spherical clusters provides their density with a representative accuracy; globular B defects are quite accurately approximated as spherical clusters. All clusters are assumed to be immobile. The effect of impurities is ignored in this treatment.

The microdefect distribution in a CZ crystal is symmetric about the axis of growth. Thus, it suffices to treat the CZ defect dynamics by an axisymmetric model using the cylindrical coordinates r and z, where r is the radial coordinate and z is the axial coordinate. At any location (r, z) in the crystal, at any given time t, a population of clusters formed at various locations (r, ξ) at various moments of the elapsed time τ exists. The size (radius) R of these clusters is then a function of r, z, τ, and t. The clusters, once formed, are assumed to grow by a diffusion-limited kinetics. The growth equations for i-clusters and v-clusters are thus given by

$$\frac{\partial R_\mathrm{i}^2(r,z,\tau,t)}{\partial t} + V\frac{\partial R_\mathrm{i}^2(r,z,\tau,t)}{\partial z} = \frac{2D_\mathrm{i}}{\psi_\mathrm{i}^\mathrm{i}}(C_\mathrm{i} - C_\mathrm{i,e})\,, \tag{38.24}$$

$$\frac{\partial R_\mathrm{v}^2(r,z,\tau,t)}{\partial t} + V\frac{\partial R_\mathrm{v}^2(r,z,\tau,t)}{\partial z} = \frac{2D_\mathrm{v}}{\psi_\mathrm{v}^\mathrm{v}}(C_\mathrm{v} - C_\mathrm{v,e})\,, \tag{38.25}$$

where ψ_x^x is the concentration of any intrinsic point defect species, denoted by the superscript, in a cluster of any type, denoted by the subscript. The relationship among z, ξ, and τ is determined by the pull rate profile

$$\xi = z - \int_\tau^t V\,\mathrm{d}\tau'\,, \tag{38.26}$$

where τ' is the time between τ and t. The *cross-interaction* between i-clusters and vacancies and between v-clusters and self-interstitials is ignored, because the cluster growth is affected primarily by the dominant intrinsic point defect species. The dynamics at moderately low temperatures is strongly influenced by oxygen, which is beyond the scope of this model. The rate of the consumption of the intrinsic point defects x by x-clusters at any location at any time, q_x^x, is obtained by integrating the contributions from all the clusters present at the location

$$q_\mathrm{i}^\mathrm{i} = 4\pi D_\mathrm{i}(C_\mathrm{i} - C_\mathrm{i,e})\int_0^t R_\mathrm{i}(r,z,\tau,t)J_\mathrm{i}(r,\xi,\tau)\,\mathrm{d}\tau\,, \tag{38.27}$$

$$q_\mathrm{v}^\mathrm{v} = 4\pi D_\mathrm{v}(C_\mathrm{v} - C_\mathrm{v,e})\int_0^t R_\mathrm{v}(r,z,\tau,t)J_\mathrm{v}(r,\xi,\tau)\,\mathrm{d}\tau\,. \tag{38.28}$$

The balance of the intrinsic point defects must account for the change in the intrinsic point defect concentration by convection, diffusion, the Frenkel reaction, their consumption by the diffusion-limited growth of the formed clusters, and their consumption by their nucleation.

$$\begin{aligned}\frac{\partial C_\mathrm{i}}{\partial t} + V\frac{\partial C_\mathrm{i}}{\partial z} &= \nabla\,(D_\mathrm{i}\nabla C_\mathrm{i}) \\ &\quad - k_{\mathrm{i}\leftrightarrow\mathrm{v}}(C_\mathrm{i}C_\mathrm{v} - C_\mathrm{i,e}C_\mathrm{v,e}) \\ &\quad - 4\pi D_\mathrm{i}(C_\mathrm{i} - C_\mathrm{i,e}) \\ &\quad \times \int_0^t R_\mathrm{i}(r,z,\tau,t)J_\mathrm{i}(r,\xi,\tau)\,\mathrm{d}\tau \\ &\quad - J_\mathrm{i}(r,z,t)m_\mathrm{i}^*\,,\end{aligned} \tag{38.29}$$

$$\begin{aligned}\frac{\partial C_\mathrm{v}}{\partial t} + V\frac{\partial C_\mathrm{v}}{\partial z} &= \nabla\,(D_\mathrm{v}\nabla C_\mathrm{v}) \\ &\quad - k_{\mathrm{i}\leftrightarrow\mathrm{v}}(C_\mathrm{i}C_\mathrm{v} - C_\mathrm{i,e}C_\mathrm{v,e}) \\ &\quad - 4\pi D_\mathrm{v}(C_\mathrm{v} - C_\mathrm{v,e}) \\ &\quad \times \int_0^t R_\mathrm{v}(r,z,\tau,t)J_\mathrm{v}(r,\xi,\tau)\,\mathrm{d}\tau \\ &\quad - J_\mathrm{v}(r,z,t)m_\mathrm{v}^*\,.\end{aligned} \tag{38.30}$$

The consumption of intrinsic point defects by their nucleation events only is negligible, and is ignored. The nucleation rates of both vacancies and self-interstitials are given by the classical nucleation theory

$$\begin{aligned}J_\mathrm{i} &= \left[4\pi R\left(m_\mathrm{i}^*\right)D_\mathrm{i}C_\mathrm{i}\right] \\ &\quad \times\left[\left\{12\pi k_\mathrm{B}T\Delta F_\mathrm{i}\left(m_\mathrm{i}^*\right)\right\}^{-\frac{1}{2}}\left(k_\mathrm{B}T\ln\frac{C_\mathrm{i}}{C_\mathrm{i,e}}\right)\right] \\ &\quad \times\left[\rho_\mathrm{i}\exp\left\{-\frac{\Delta F_\mathrm{i}\left(m_\mathrm{i}^*\right)}{k_\mathrm{B}T}\right\}\right]\,,\end{aligned} \tag{38.31}$$

$$J_\mathrm{v} = \left[4\pi R_x \left(m_x^*\right) D_\mathrm{v} C_\mathrm{v}\right] \times \left[\left\{12\pi k_\mathrm{B} T \Delta F_\mathrm{v}\left(m_x^*\right)\right\}^{-\frac{1}{2}} \left(k_\mathrm{B} T \ln \frac{C_\mathrm{v}}{C_\mathrm{v,e}}\right)\right] \times \left[\rho_\mathrm{v} \exp\left\{-\frac{\Delta F_\mathrm{v}\left(m_x^*\right)}{k_\mathrm{B} T}\right\}\right] . \quad (38.32)$$

The discussed set of equations must be solved for a moving crystal. The transient domain of the computation is described by the shape of the crystal as a function of time

$$\Omega(r, z, t) = 0 . \quad (38.33)$$

The initial height of the crystal is assumed to be either zero or negligible. Equilibrium conditions are assumed on all crystal surfaces, on the basis of a fast surface kinetics. This assumption is valid in most CZ growth conditions [38.26, 34]. The initial size of the clusters formed at any location (r, ξ) is approximately described by the size of the critical clusters.

The cumulative density of all x-clusters in a population present at any location, n_x, can be explicitly defined using the solution of the described equations

$$n_\mathrm{i} = \int_0^t J_\mathrm{i}(r, \xi, \tau)\,\mathrm{d}\tau , \quad (38.34)$$

$$n_\mathrm{v} = \int_0^t J_\mathrm{v}(r, \xi, \tau)\,\mathrm{d}\tau . \quad (38.35)$$

The volumetric average radius of the cluster population at any location is given by

$$R_\mathrm{i,avg,vol} = \left(\frac{\int_0^t R_\mathrm{i}^3\,(r, z, \tau, t)\, J_\mathrm{i}(r, \xi, \tau)\,\mathrm{d}\tau}{\int_0^t J_\mathrm{i}(r, \xi, \tau)\,\mathrm{d}\tau}\right)^{1/3} , \quad (38.36)$$

$$R_\mathrm{v,avg,vol} = \left(\frac{\int_0^t R_\mathrm{v}^3\,(r, z, \tau, t)\, J_\mathrm{v}(r, \xi, \tau)\,\mathrm{d}\tau}{\int_0^t J_\mathrm{v}(r, \xi, \tau)\,\mathrm{d}\tau}\right)^{1/3} . \quad (38.37)$$

The subscript "avg,vol" denotes the volumetric average value.

Typically, the described set of equations can be solved in a quasistationary temperature field at any given height of the crystal. The CZ heat transport dynamics is much faster than the CZ defect dynam-

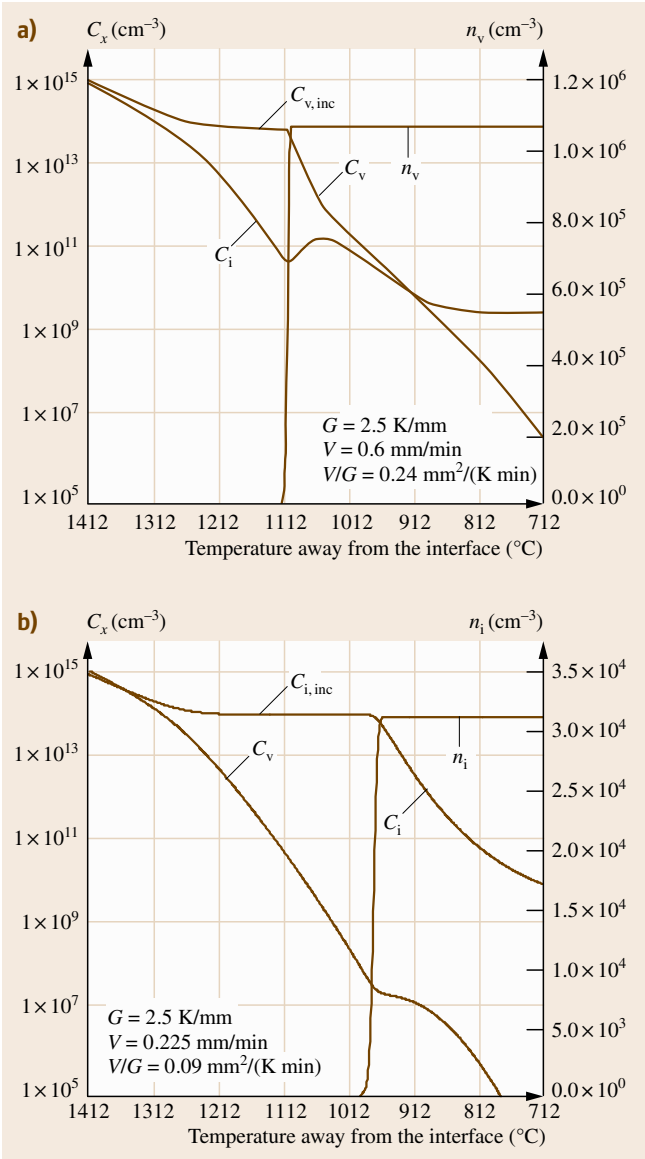


Fig. 38.7a,b The profiles of the concentrations of the intrinsic point defect species and the cluster densities as functions of the crystal temperature in a simulated **(a)** vacancy-rich crystal growing under steady state, **(b)** self-interstitial-rich crystal growing under steady state (after [38.1]). Note that the Frenkel reaction equilibrium remains valid until the cluster growth is complete and that the cross-interactions (i with v-clusters and v with i-clusters) are not included in the model

ics [38.26, 27, 34, 48, 49, 53, 54]. The temperature field in a growing crystal can be quantified accurately by the quasi-steady-state approximation [38.35–37]. In most cases, the assumption of one representative temperature field independent of the crystal height suffices [38.53, 54].

The accuracy of the model predictions strongly depends on the accuracy of the parameters describing the key properties of the intrinsic point defects. There is significant uncertainty in the reported values of the formation and migration energies of the intrinsic point defects. The parameters describing the Frenkel reaction kinetics, in particular, are highly approximate. Various studies have reported the acceptable parameters describing the CZ defect dynamics [38.1, 26, 34, 58–64]. In this section, the results obtained by *Kulkarni* et al. by solving the described equations using the properties of the intrinsic point defects listed in Table 38.1 are discussed [38.1]. The Frenkel reaction rate constant reported by *Sinno* et al. [38.26, 27] was used by *Kulkarni* and coworkers in their study [38.1]; the enthalpic barrier, however, was set to zero [38.1, 27]. An accurate estimation of this rate constant is not necessary, because the Frenkel reaction dynamics is very fast, leading to reaction equilibrium in the relevant temperature range.

It must be noted that the discussed model quantifies continuum-scale CZ defect dynamics. The effects of oxygen on the defect dynamics are also not included in the model.

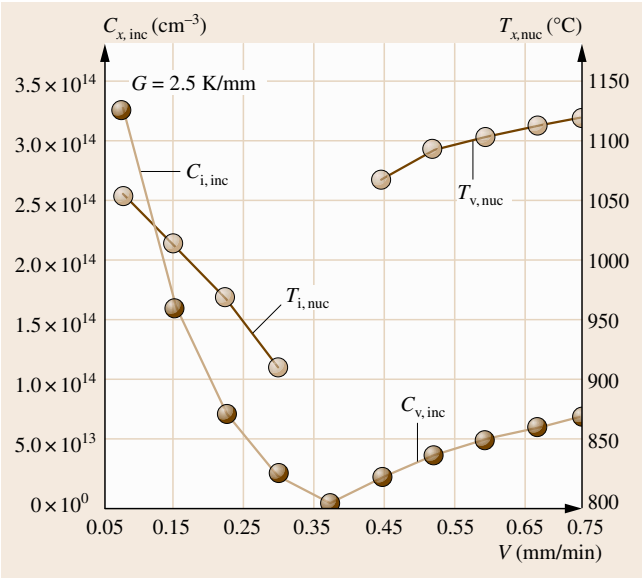


Fig. 38.8 Dependence of the concentration of the dominant incorporated point defect species and the nucleation temperature on the pull rate at a fixed G (subscript "inc" denotes the incorporated concentration and "nuc" denotes nucleation) (after [38.1])

Results and Discussion. *Voronkov* and *Falster* solved the described model by assuming that the effects of the axial and radial diffusion of the intrinsic point defects after the initial incorporation are negligible [38.46]. *Kulkarni* et al. solved the model describing both the steady-state and unsteady-state defect dynamics including axial diffusion effects [38.1]. The basic aspects of their study are discussed in this section.

Kulkarni et al. used a representative temperature profile described by (38.6) for their steady-state simulations [38.1]. A representative value for G of 2.5 K/mm was used. For the sake of simplicity, the units popularly applied in the crystal growth industry are used hereafter to describe key variables. Vacancies are incorporated as the dominant species at high V/G, as shown in Fig. 38.7a, and self-interstitials are the dominant incorporated species at low V/G, as shown in Fig. 38.7b. Figure 38.7 also shows that v-clusters and i-clusters are formed within a narrow range of temperature known as the *nucleation temperature range*. The nucleation temperature is defined as the temperature at which the nucleation rate is at its maximum. The predicted nucleation temperature of vacancies for the conditions studied is around 1100 °C and that for self-interstitials is around 950 °C. As reported by *Kulkarni*, the Frenkel reaction equilibrium prevails in the crystal until the cluster growth is complete [38.1]. At lower temperatures, when the intrinsic point defect concentrations are too low to affect the cluster distribution, the model predicts a deviation from the Frenkel reaction equilibrium. The model does not focus on an accurate quantification of the very low residual intrinsic point defect concentrations. Hence, it is not necessary to account for the Frenkel reaction kinetics in CZ growth. A set of simulations like this at a fixed G and varying V can capture the effect of V/G on the initial incorporation (Fig. 38.8). As shown in Fig. 38.8, the critical V/G is around 0.15 mm^2/(K min). It must be noted that various other groups report slightly different values of the critical V/G [38.26, 30–33]. Figure 38.8 also shows that the nucleation temperature increases with

increasing concentration of the incorporated dominant intrinsic point defect species, because the temperature at which the intrinsic point defect supersaturation is high enough to drive the nucleation events increases with increasing intrinsic point defect concentration. The cluster size distribution is influenced by an interplay between the formation of new clusters, which is driven by the dominant intrinsic point defect supersaturation, and the consumption of the intrinsic point defects by the existing clusters, which decreases the intrinsic point defect supersaturation. The conditions that allow the nucleation of the intrinsic point defects at a higher rate, before the intrinsic point defect concentration decreases by the cluster growth, lead to the formation of a large number of clusters, which remain very small in size; the conditions allowing the rapid growth of the formed clusters quickly reduce the intrinsic point defect concentration and the nucleation rate, leading to the formation of a small number of clusters that grow very large in size. More specifically, a higher cooling rate through the nucleation range leads to the evolution of smaller clusters at higher densities, whereas a higher incorporated intrinsic point defect concentration leads to the evolution of larger clusters at lower densities. This interplay is captured in the size distributions of the mature cluster populations in the simulated CZ crystals grown under varying conditions, as shown in Fig. 38.9. Using their simplified model, *Voronkov* and *Falster* quantified this interplay in terms of the cooling rate through the nucleation range and the concentration of the incorporated dominant intrinsic point defect species [38.46]

$$n_{x,\text{approx}} \propto \left(\frac{1}{D_x T_{\text{nuc},x}^2}\right)^{\frac{3}{2}} Q_{\text{nuc},x}^{\frac{3}{2}} C_{x,\text{nuc}}^{-\frac{1}{2}} \quad (38.38)$$

$$R_{x,\text{avg,approx}} \propto \left(D_x T_{\text{nuc},x}^2\right)^{\frac{1}{2}} \left(\frac{C_{x,\text{nuc}}}{Q_{\text{nuc},x}}\right)^{\frac{1}{2}} , \quad (38.39)$$

where Q is the cooling rate, given by the product of the local pull rate and the magnitude of the axial temperature gradient. The subscript "nuc" denotes the conditions at nucleation, and the subscript "approx" denotes an approximate value. *Kulkarni* et al. showed that the predictions of (38.38) and (38.39) agree quite well with the predictions of their rigorous model [38.1].

Many modern CZ processes enforce unsteady-state conditions in crystal growth by varying the crystal pull rate. *Kulkarni* et al. [38.1] captured the salient features of unsteady-state CZ growth by simulating the growth of a crystal pulled at the varying rate shown in Fig. 38.10a. The body of the simulated crystal was grown first by continuously decreasing the pull rate and then continuously increasing the pull rate. The predicted cluster type and the cluster density variation are shown in Fig. 38.10b. Each element of the crystal undergoes

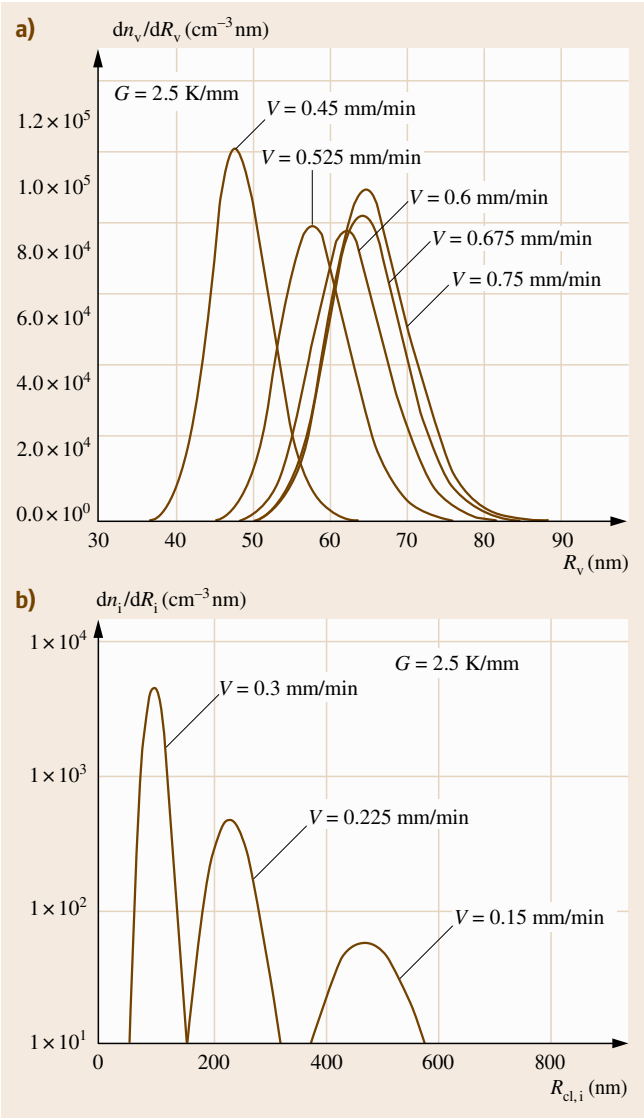


Fig. 38.9a,b The mature size distributions of **(a)** v-clusters and **(b)** i-clusters in various simulated crystals grown under steady states at various pull rates (after [38.1])

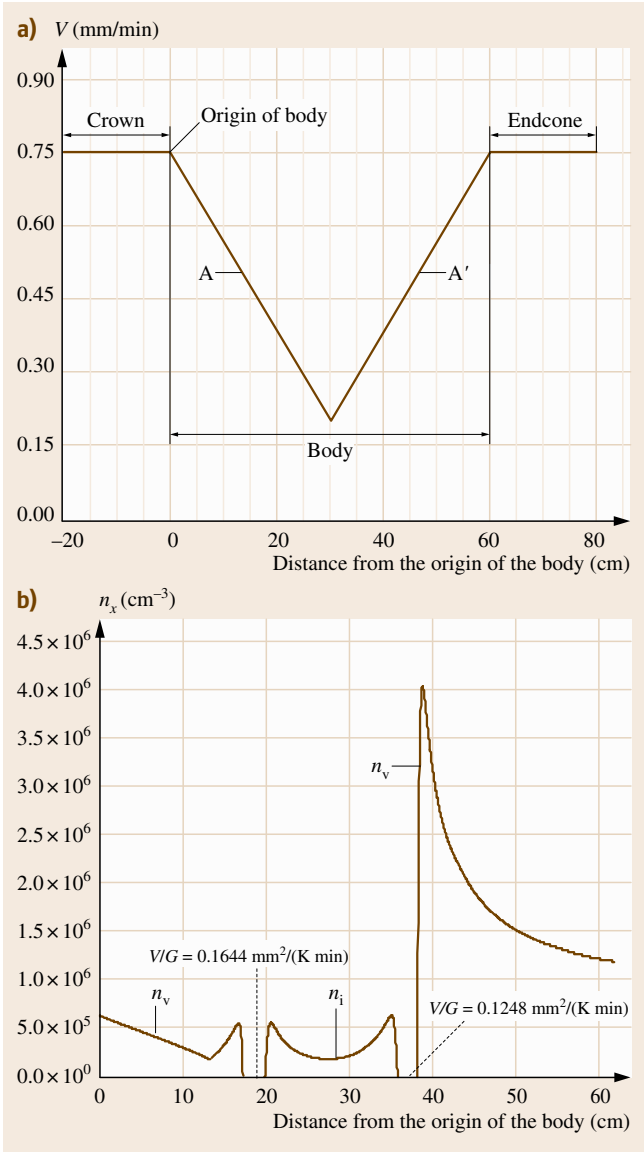


Fig. 38.10 **(a)** The crystal pull-rate profile used to understand the CZ defect dynamics under unsteady-state conditions. **(b)** The cluster density profile in the crystal grown using the pull-rate profile shown in **(a)** (after [38.1])

initial incorporation and nucleation under varying conditions, resulting in the interesting cluster distribution shown in Fig. 38.10. The most striking features of this study are the predicted shifts of the critical V/G, induced by the unsteady-state crystal growth. These shifts are caused by the variation of the excess intrinsic point defect flux associated with the variation of the crystal pull rate. When the pull rate continuously decreases, an element of silicon moves away from the melt–crystal interface at a decreasing rate, allowing a more efficient diffusion of self-interstitials from the interface. In addition, at a given time and for a given pull rate, the driving force for self-interstitial diffusion at the interface is slightly higher than that for steady-state growth, and the driving force for the vacancy diffusion is slightly lower than that for steady-state growth; this dynamics is caused by the prior higher pull rate. Hence, the critical V/G increases when the pull rate continuously decreases. Conversely, the exact opposite effects explain the decrease of the critical V/G when the pull rate continuously increases. These shifts of the critical V/G were quantified by *Kulkarni* et al. for engineering applications as functions of the rate of the change of the pull rate with respect to crystal length [38.1]

$$\left(\frac{V}{G}\right)_{\mathrm{c},-\mathrm{slope}} = \left(\frac{V}{G}\right)_{\mathrm{c}} - 7.85\,(\mathrm{mm}^2/\mathrm{K}) \times \left(\frac{\mathrm{d}V}{\mathrm{d}L}\right)(\mathrm{min}^{-1})\,, \tag{38.40}$$

$$\left(\frac{V}{G}\right)_{\mathrm{c},+\mathrm{slope}} = \left(\frac{V}{G}\right)_{\mathrm{c}} - 13.745\,(\mathrm{mm}^2/\mathrm{K}) \times \left(\frac{\mathrm{d}V}{\mathrm{d}L}\right)(\mathrm{min}^{-1})\,. \tag{38.41}$$

The subscripts "−slope" and "+slope" indicate the decreasing pull rate and the increasing pull rate, respectively.

Finally, *Kulkarni* et al. [38.1] also showed that the predictions of the applied model agree reasonably well with experimental observations, as shown in Fig. 38.11. The D defect density was experimentally determined in a crystal grown under unsteady-state conditions by the method of copper decoration and Secco etching [38.38–40]. These studies clearly establish the validity of the applied model.

The model applied for the quantification of CZ defect dynamics discussed so far predicts the size distributions of all populations of clusters at all locations in a CZ crystal. The model requires the solution of the integro-differential equations for the intrinsic point defect concentration fields and of the cluster growth equations describing the evolution of the cluster populations at all locations in a CZ crystal. As this model provides rigorous quantification of the CZ defect dynamics, it is termed the *rigorous model*.

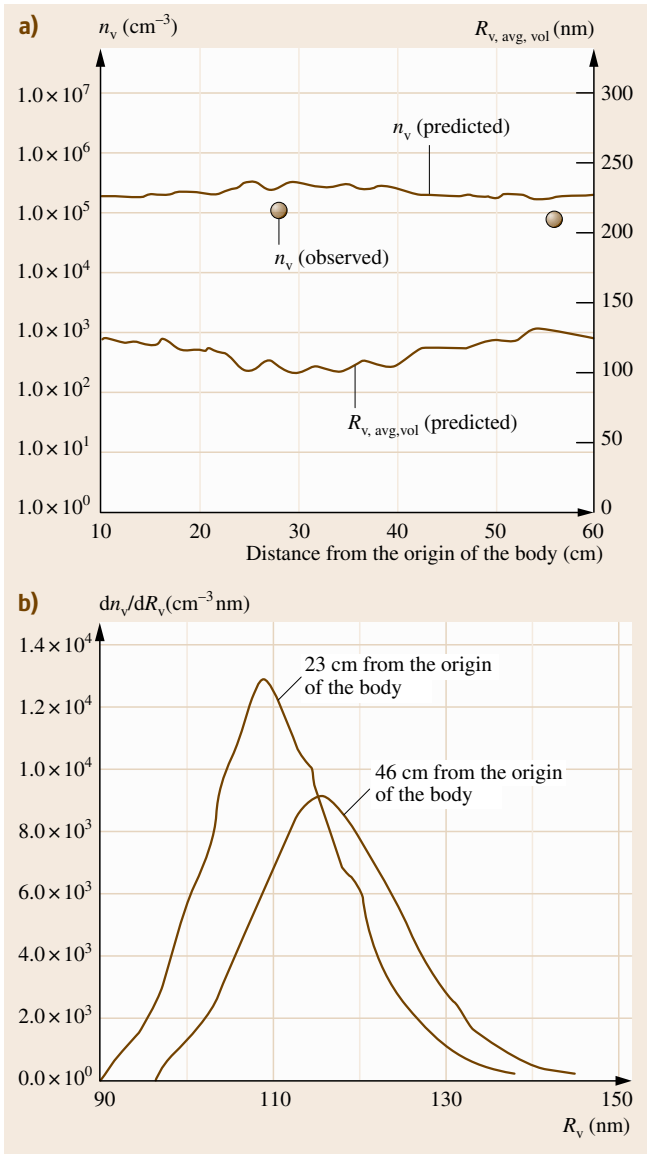


Fig. 38.11 **(a)** Comparison between the model predicted and experimentally determined approximate v-cluster density profiles in an experimental crystal. *Circles* indicate experimental data points. The profile of the volume-averaged v-cluster size in the crystal is also shown. **(b)** The v-cluster size distribution at two chosen axial locations in the crystal (after [38.1]) ▶

Second Approach: The Quantification of the CZ Defect Dynamics by the Lumped Model

The rigorous model can be simplified by representing a population of clusters of varying sizes at any given location in a CZ crystal by an equivalent population of identical clusters, as first shown by *Kulkarni* and *Voronkov* [38.53]. Thus, the complex rigorous model is reformulated by explicitly introducing the density and the average size of clusters in a population

$$n_i(r,z,t) = \int_0^t J_i(r,\xi,\tau)\,d\tau\,, \tag{38.42}$$

$$n_v(r,z,t) = \int_0^t J_v(r,\xi,\tau)\,d\tau\,, \tag{38.43}$$

$$\langle R_i\rangle = \frac{\int_0^t R_i(r,z,\tau,t)J_i(r,\xi,\tau)\,d\tau}{n_i}\,, \tag{38.44}$$

$$\langle R_v\rangle = \frac{\int_0^t R_v(r,z,\tau,t)J_v(r,\xi,\tau)\,d\tau}{n_v}\,, \tag{38.45}$$

where the brackets $\langle\cdots\rangle$ indicate the average value. Note that these average radii are different from the volumetric average radii defined in (38.36) and (38.37). The total consumption rate (per unit volume) of the intrinsic point defects by this population is now given as

$$q_i^i = 4\pi D_i(C_i - C_{i,e})\langle R_i\rangle n_i\,, \tag{38.46}$$

$$q_v^v = 4\pi D_v(C_v - C_{v,e})\langle R_v\rangle n_v\,. \tag{38.47}$$

The essential aspect of the simplified model is to replace the average radius of the cluster population by the square root of the average of the squares of the radii of all clusters in the population

$$\langle R_i\rangle \approx \left\langle R_i^2\right\rangle^{\frac{1}{2}} = \left(\frac{\int_0^t R_i^2(r,z,\tau,t)J_i(r,\xi,\tau)\,d\tau}{n_i}\right)^{1/2} = \left(\frac{U_i}{n_i}\right)^{1/2}\,, \tag{38.48}$$

$$\langle R_\mathrm{v}\rangle \approx \left\langle R_\mathrm{v}^2\right\rangle^{\frac{1}{2}} = \left(\frac{\int_0^t R_\mathrm{v}^2(r,z,\tau,t)J_\mathrm{v}(r,\xi,\tau)\,\mathrm{d}\tau}{n_\mathrm{v}}\right)^{1/2} = \left(\frac{U_\mathrm{v}}{n_\mathrm{v}}\right)^{1/2}, \tag{38.49}$$

where

$$U_\mathrm{i}(r,z,t) = \int_0^t R_\mathrm{i}^2(r,z,\tau,t)J_\mathrm{i}(r,\xi,\tau)\,\mathrm{d}\tau\,, \tag{38.50}$$

$$U_\mathrm{v}(r,z,t) = \int_0^t R_\mathrm{v}^2(r,z,\tau,t)J_\mathrm{v}(r,\xi,\tau)\,\mathrm{d}\tau\,. \tag{38.51}$$

The new auxiliary variable U is proportional to the total surface area of the cluster population. The intrinsic point defect consumption rate per unit volume is now rewritten as

$$q_\mathrm{i}^\mathrm{i} = 4\pi D_\mathrm{i}(C_\mathrm{i}-C_\mathrm{i,e})\left\langle R_\mathrm{i}^2\right\rangle^{1/2} n_\mathrm{i} = 4\pi D_\mathrm{i}(C_\mathrm{i}-C_\mathrm{i,e})\,(U_\mathrm{i}n_\mathrm{i})^{1/2}\,, \tag{38.52}$$

$$q_\mathrm{v}^\mathrm{v} = 4\pi D_\mathrm{v}(C_\mathrm{v}-C_\mathrm{v,e})\left\langle R_\mathrm{v}^2\right\rangle^{1/2} n_\mathrm{v} = 4\pi D_\mathrm{v}(C_\mathrm{v}-C_\mathrm{v,e})\,(U_\mathrm{v}n_\mathrm{v})^{1/2}\,. \tag{38.53}$$

The intrinsic point defect balances (38.29) and (38.30) are written using (38.52) and (38.53).

$$\frac{\partial C_\mathrm{i}}{\partial t} + V\frac{\partial C_\mathrm{i}}{\partial z} = \nabla\left(D_\mathrm{i}\nabla C_\mathrm{i}\right) - k_{\mathrm{i}\leftrightarrow\mathrm{v}}(C_\mathrm{i}C_\mathrm{v} - C_\mathrm{i,e}C_\mathrm{v,e}) - 4\pi D_\mathrm{i}(C_\mathrm{i}-C_\mathrm{i,e})\,(U_\mathrm{i}n_\mathrm{i})^{1/2}\,, \tag{38.54}$$

$$\frac{\partial C_\mathrm{v}}{\partial t} + V\frac{\partial C_\mathrm{v}}{\partial z} = \nabla\left(D_\mathrm{v}\nabla C_\mathrm{v}\right) - k_{\mathrm{i}\leftrightarrow\mathrm{v}}(C_\mathrm{i}C_\mathrm{v} - C_\mathrm{i,e}C_\mathrm{v,e}) - 4\pi D_\mathrm{v}(C_\mathrm{v}-C_\mathrm{v,e})\,(U_\mathrm{v}n_\mathrm{v})^{1/2}\,. \tag{38.55}$$

Thus the intrinsic point defect balances are described by (38.54) and (38.55) without the knowledge of the formation and growth histories of the cluster populations at any location. If the Frenkel reaction equilibrium is assumed, (38.54) and (38.55) are replaced by (38.3) and (38.56), obtained by subtracting (38.54) from (38.55)

$$\frac{\partial\left(C_\mathrm{v}-C_\mathrm{i}\right)}{\partial t} + V\frac{\partial\left(C_\mathrm{v}-C_\mathrm{i}\right)}{\partial z} = \nabla\left(D_\mathrm{v}\nabla C_\mathrm{v}\right) - \nabla\left(D_\mathrm{i}\nabla C_\mathrm{i}\right) - 4\pi D_\mathrm{v}(C_\mathrm{v}-C_\mathrm{v,e})\,(U_\mathrm{v}n_\mathrm{v})^{1/2} + 4\pi D_\mathrm{i}(C_\mathrm{i}-C_\mathrm{i,e})\,(U_\mathrm{i}n_\mathrm{i})^{1/2}\,. \tag{38.56}$$

The evolution of the auxiliary variable U is derived using the cluster growth (38.24) and (38.25) with the definitions (38.50) and (38.51)

$$\frac{\partial U_\mathrm{i}}{\partial t} + V\frac{\partial U_\mathrm{i}}{\partial z} = \frac{2D_\mathrm{i}n_\mathrm{i}}{\psi_\mathrm{i}^\mathrm{i}}(C_\mathrm{i}-C_\mathrm{i,e})\,, \tag{38.57}$$

$$\frac{\partial U_\mathrm{v}}{\partial t} + V\frac{\partial U_\mathrm{v}}{\partial z} = \frac{2D_\mathrm{v}n_\mathrm{v}}{\psi_\mathrm{v}^\mathrm{v}}(C_\mathrm{v}-C_\mathrm{v,e})\,. \tag{38.58}$$

The initial size of the formed clusters is assumed to be zero in the derivation of (38.57) and (38.58). This assumption is accurate and does not the affect the predictions of the model. The evolution of the density of clusters is directly obtained using the classical nucleation theory

$$\frac{\partial n_\mathrm{i}}{\partial t} + V\frac{\partial n_\mathrm{i}}{\partial z} = J_\mathrm{i}\,, \tag{38.59}$$

$$\frac{\partial n_\mathrm{v}}{\partial t} + V\frac{\partial n_\mathrm{v}}{\partial z} = J_\mathrm{v}\,. \tag{38.60}$$

The representative size of clusters at any location $\mathcal{R}$ is given as

$$\mathcal{R}_\mathrm{i} = \left\langle R_\mathrm{i}^2\right\rangle^{1/2} = \left(\frac{U_\mathrm{i}}{n_\mathrm{i}}\right)^{1/2}, \tag{38.61}$$

$$\mathcal{R}_\mathrm{v} = \left\langle R_\mathrm{v}^2\right\rangle^{1/2} = \left(\frac{U_\mathrm{v}}{n_\mathrm{v}}\right)^{1/2}. \tag{38.62}$$

The defect dynamics in a CZ crystal is now quantified by the intrinsic point defect concentration C, the auxiliary variable U, and the cluster density n. These variables are described by a set of partial differential equations without the necessity to quantify the formation and growth histories of clusters. The simplified model eliminates the elapsed time, introduced in the rigorous model to describe the size distribution of a cluster population at any given location in a CZ crystal, as an independent variable. Hence, this simplified model is computationally attractive. As the simplified model represents the population of clus-

ters of varying sizes at a given location in the CZ crystal by an equivalent population of identical clusters, it is termed the *lumped model* by *Kulkarni* and *Voronkov* [38.53].

The accuracy of the lumped model is verified by a comparison of its predictions with the predictions of the rigorous model, for many different crystal growth conditions, as shown in Fig. 38.12a. In addition, the lumped model is validated by experimental observations (Fig. 38.12b). Finally, as shown in Fig. 38.13, the two-dimensional microdefect distribution in a crystal pulled at a varying rate is predicted reasonably well by the lumped model. It must be noted that the mesh discretization used for the computation is relatively coarse. The inaccuracies associated with the mesh discretization are reduced in the simulations discussed in Sect. 38.3. In the discussed simulations, the lumped model was solved assuming the Frenkel reaction equilibrium. The method of copper decoration and Secco etching was used for the experimental determination of the microdefect distribution [38.38–40]. These studies establish the lumped model as a valuable engineering tool for the development of new CZ crystal growth processes.

Third Approach: The Quantification of the CZ Defect Dynamics by the Discrete Rate Equations and the Fokker–Planck Equation

The models described in the previous sections apply the classical nucleation theory to predict the formation of stable clusters and a diffusion-limited growth kinetics to quantify the growth of these clusters. This approach works very well for Czochralski crystal growth. A more rigorous treatment of CZ defect dynamics, however, must account for all reactions involved in the intrinsic point defect aggregation, as described by reaction (38.18). Considering the large size of microdefects in CZ crystals, this approach requires the solution of an impractically large number of equations. *Sinno* and *Brown* [38.47], *Mori* [38.27], and *Brown* et al. [38.49] quantified the defect dynamics in CZ crystals by applying a mixed approach involving the solution of a set of equations derived treating smaller clusters as discrete particles and the solution of a set of Fokker–Planck equations derived from the discrete equations for larger clusters. In this section, the contributions of this work to the field of CZ defect dynamics are discussed.

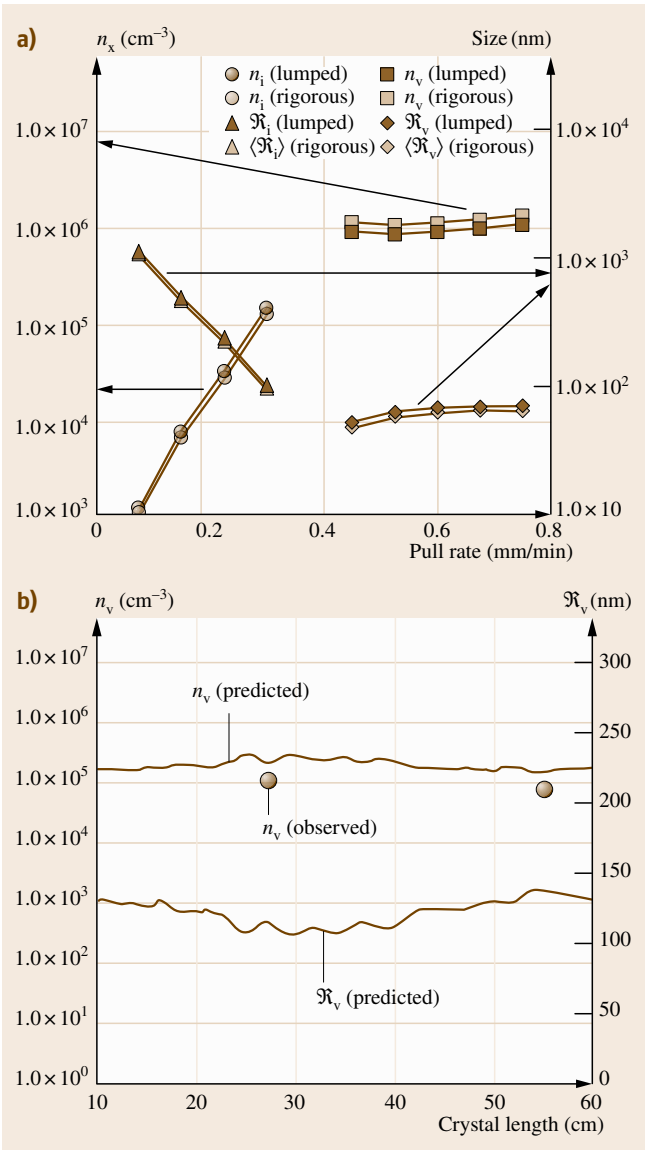


Fig. 38.12 **(a)** Comparison between the predictions of the rigorous model and the predictions of the lumped model. **(b)** Comparison between the predictions of the lumped model with the experimental observations (after [38.53])

The Discrete Rate Equations. Reaction set (38.18) defines the series of reactions driving the aggregation of vacancies and self-interstitials. In this reaction set, clusters of the same size and type are treated as a separate species. The cross-interactions between v-clusters and

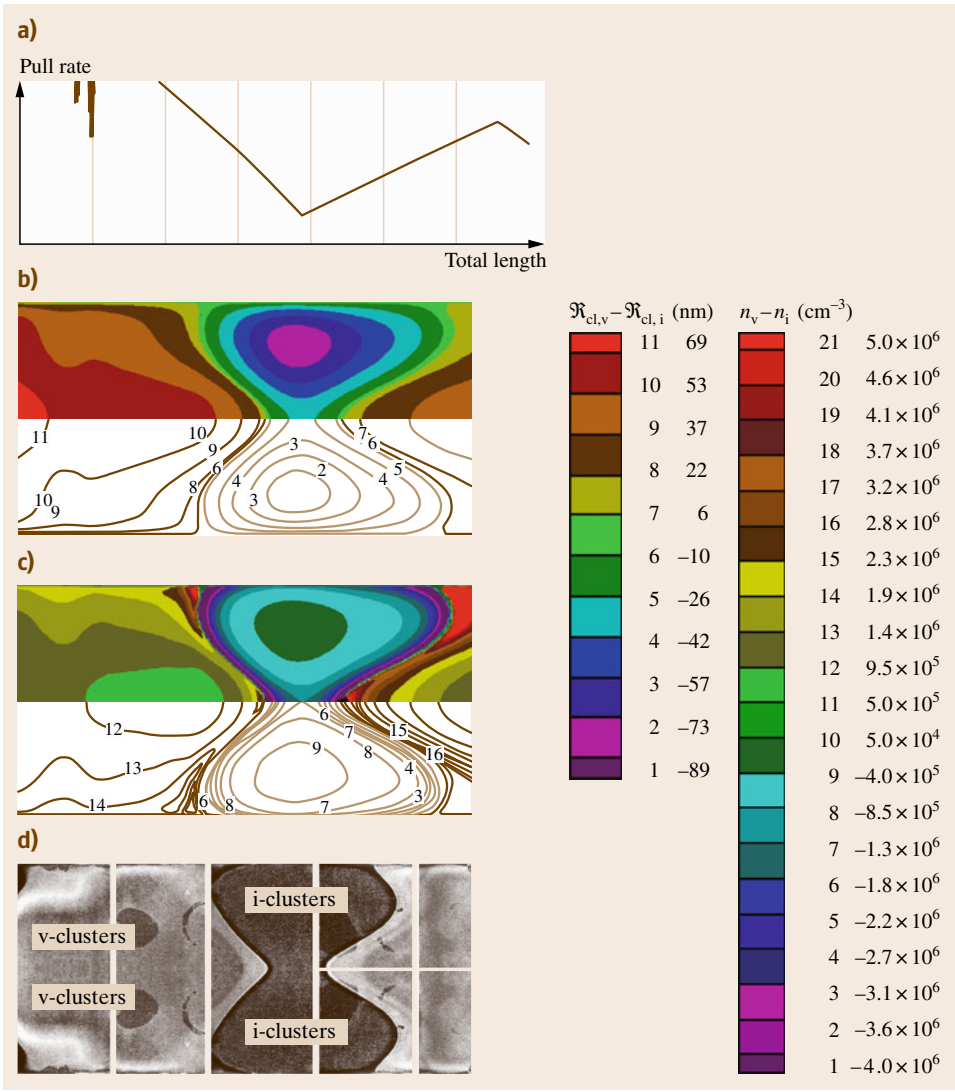


Fig. 38.13 **(a)** The pull-rate profile used to pull the experimental CZ crystal. **(b)** The predicted $\Re_v - \Re_i$. **(c)** The predicted $n_v - n_i$. **(d)** The observed defect distribution in the crystal. Positive values indicate the v-cluster size and density and the magnitudes of the negative values indicate the i-cluster size and density. v-clusters and i-clusters do not coexist (after [38.53])

self-interstitials and i-clusters and vacancies are ignored in this chapter.

In the previous sections, the size of a cluster was defined as its radius. In this section, the size of a cluster is defined by the number of intrinsic point defects in it rather than by its radius, following the work of *Sinno* and *Brown* [38.47], *Mori* [38.27], *Wang* and *Brown* [38.48], and *Brown* et al. [38.49]. In an element of silicon, the overall rate of formation of clusters of size m, Φ_m, is given by the difference between the net volumetric flux coming from the clusters of size $m-1$ to the clusters of size m, I_m, and the net volumetric flux going from clusters of size m to clusters of size $m+1$, $I_{(m+1)}$

$$\Phi_{m_x} = I_{m_x} - I_{(m+1)_x} \,. \tag{38.63}$$

The subscript "m_x" indicates an x-cluster containing m intrinsic point defects of type x. Note that J_x defines the nucleation rate of stable x-clusters, as defined in the previous sections, and Φ_{m_x} defines the net formation rate of x-clusters of size m. The net volumetric flux coming from the clusters containing $m-1$ intrinsic point defects to the clusters containing m intrinsic point defects is defined by the growth rate of the former and the dissolution rate of the latter

$$I_{m_x} = g_{(m-1)_x}\phi_{(m-1)_x} - d_{m_x}\phi_{m_x} \,, \tag{38.64}$$

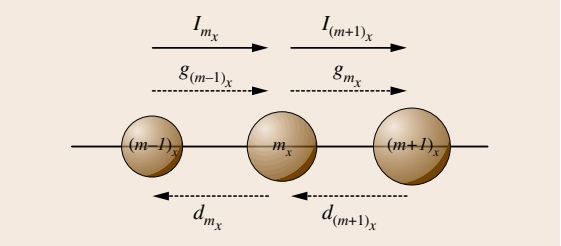


Fig. 38.14 The relationship among the nucleation flux I, the growth rate g, and the dissolution rate d (after [38.27])

where ϕ is the density of clusters, g is the growth rate of a cluster, and d is the dissolution rate. The relationship among I, g, and d is shown schematically in Fig. 38.14. *Sinno* and *Brown* [38.47], *Mori* [38.27], *Wang* and *Brown* [38.48], and *Brown* et al. [38.49] define the cluster growth rate and the dissolution rate by incorporating the kinetic interactions between monomers and clusters. In this section, these rates are defined assuming a diffusion-limited interaction between clusters and monomers, for the sake of consistency with the discussion in other sections of this chapter. The growth rate of a cluster containing m monomers of type x is given by the diffusion-limited attachment frequency of monomers to the cluster

$$g_{m_x} = 4\pi R_{m_x} D_x C_x \,, \tag{38.65}$$

where R_m is the radius of a cluster of size m. The cluster dissolution rate is given by its thermodynamic relationship with the cluster growth rate

$$d_{m_x} = \frac{g_{(m-1)_x}\phi_{(m-1)_x,\mathrm{e}}}{\phi_{m_x,\mathrm{e}}} \,. \tag{38.66}$$

The equilibrium concentration of clusters of size m, in a solution of a given composition, at a given temperature, is given by

$$\phi_{m_x,\mathrm{e}} = \rho_x \mathrm{e}^{-\left(\frac{\Delta F_{m_x}}{k_\mathrm{B}T}\right)} \,. \tag{38.67}$$

The total free energy change associated with the formation of a cluster of size m, from a solution of a given composition, at a given temperature, is

$$\Delta F_{m_x} = -m_x k_\mathrm{B} T \ln \frac{C_x}{C_{x,\mathrm{e}}} + \Gamma_{m_x} \,, \tag{38.68}$$

where Γ_m is the formation energy of a cluster of size m. The formation energy of a large unstrained cluster of size m is simply approximated as $\lambda m^{2/3}$. Using (38.67) and (38.68), (38.66) is written as

$$d_{m_x} = \frac{g_{(m-1)_x}}{\left(\frac{C_x}{C_{x,\mathrm{e}}}\right)\mathrm{e}^{\frac{\left(\Gamma_{(m-1)_x}-\Gamma_{m_x}\right)}{k_\mathrm{B}T}}} \,. \tag{38.69}$$

These equations define the diffusion limited growth rates and dissolution rates of all clusters in a Czochralski crystal.

The Cluster Balance Using Discrete Rate Equations. With the discrete formulation, the cluster conservation equations in an axisymmetric CZ crystal are written as follows

$$\frac{\partial \phi_{m_x}(r,z,t)}{\partial t} + V\frac{\partial \phi_{m_x}(r,z,t)}{\partial z} = I_{m_x}(r,z,t) - I_{(m+1)_x}(r,z,t) \,, \tag{38.70}$$

for $m_x \geq 2$.

In (38.70), the density of clusters of the same size and type is treated as a dependent variable. This equation must be solved with the intrinsic point defect balances. The quantification of the defect dynamics, in CZ crystals containing microdefects as large as 100–200 nm, using the discrete rate equations is computationally impractical. Therefore, reasonable approximations based on the discrete rate equations are necessary for a practical solution of the problem.

The Fokker–Planck Equation. The Fokker–Planck equation (FPE) is derived from the discrete rate equations by a Kramers–Moyal expansion treating m as a continuous independent variable [38.27,47]. The FPE reduces the number of equations defining the CZ defect dynamics. By this formulation, the size distribution of clusters is written as a continuous function of m

$$\frac{\partial f_x(r,z,t,m_x)}{\partial t} + V\frac{\partial f_x(r,z,t,m_x)}{\partial z} = -\frac{\partial}{\partial m_x}\left[A_x(r,z,t,m_x)\, f_x(r,z,t,m_x) - B_x(r,z,t,m_x)\frac{\partial f_x(r,z,t,m_x)}{\partial m_x}\right] \,. \tag{38.71}$$

The cluster density determined by the FPE is written as f and the subscript x defines the type of the cluster. A is termed the *drift coefficient* and B is termed the *diffusion coefficient* following the generalized transport equation written in m-space, and they are related to the discrete rate equations as

$$A_x(r,z,t,m_x) = g_x(r,z,t,m_x) - d_x(r,z,t,m_x) - \frac{\partial B_x(r,z,t,m_x)}{\partial m_x} \,, \tag{38.72}$$

$$B_x(r,z,t,m_x) = \frac{g_x(r,z,t,m_x) + d_x(r,z,t,m_x)}{2} \,. \tag{38.73}$$

The Fokker-Plank equation thus reduces the number of equations describing the cluster growth. Using this formulation, a simplified model can now be developed.

The Model. The Fokker–Planck equation is accurate when clusters are large. Hence, an accurate model involves describing the CZ defect dynamics by the discrete rate equations for smaller clusters and by the Fokker–Planck equation for larger clusters. Thus, the cluster balances are written as follows

$$\frac{\partial \phi_{m_\mathrm{i}}(r,z,t)}{\partial t} + V\frac{\partial \phi_{m_\mathrm{i}}(r,z,t)}{\partial z} = I_{m_\mathrm{i}}(r,z,t) - I_{(m+1)_\mathrm{i}}(r,z,t)\ , \quad \text{for } (m_\mathrm{dis} \geq m_\mathrm{i} \geq 2)\ , \tag{38.74}$$

$$\frac{\partial \phi_{m_\mathrm{v}}(r,z,t)}{\partial t} + V\frac{\partial \phi_{m_\mathrm{v}}(r,z,t)}{\partial z} = I_{m_\mathrm{v}}(r,z,t) - I_{(m+1)_\mathrm{v}}(r,z,t)\ , \quad \text{for } (m_\mathrm{dis} \geq m_\mathrm{v} \geq 2)\ , \tag{38.75}$$

$$\frac{\partial f_\mathrm{i}(r,z,t,m_\mathrm{i})}{\partial t} + V\frac{\partial f_\mathrm{i}(r,z,t,m_\mathrm{i})}{\partial z} = -\frac{\partial}{\partial m_\mathrm{i}}\left[A_\mathrm{i}(r,z,t,m_\mathrm{i})\, f_\mathrm{i}(r,z,t,m_\mathrm{i}) - B_\mathrm{i}(r,z,t,m_\mathrm{i})\frac{\partial f_\mathrm{i}(r,z,t,m_\mathrm{i})}{\partial m_\mathrm{i}}\right], \quad \text{for } (m_\mathrm{dis} < m_\mathrm{i} \leq m_\mathrm{max})\ , \tag{38.76}$$

$$\frac{\partial f_\mathrm{v}(r,z,t,m_\mathrm{v})}{\partial t} + V\frac{\partial f_\mathrm{v}(r,z,t,m_\mathrm{v})}{\partial z} = -\frac{\partial}{\partial m_\mathrm{v}}\left[A_\mathrm{v}(r,z,t,m_\mathrm{v})\, f_\mathrm{v}(r,z,t,m_\mathrm{v}) - B_\mathrm{v}(r,z,t,m_\mathrm{v})\frac{\partial f_\mathrm{v}(r,z,t,m_\mathrm{v})}{\partial m_\mathrm{v}}\right], \quad \text{for } (m_\mathrm{dis} < m_\mathrm{v} \leq m_\mathrm{max})\ . \tag{38.77}$$

The subscript "dis" denotes the maximum cluster size treated by the discrete rate equations, and the subscript "max" denotes the maximum cluster size quantified by the Fokker–Planck equations. The intrinsic point defect balances following this approach are written as

$$\frac{\partial C_\mathrm{i}}{\partial t} + V\frac{\partial C_\mathrm{i}}{\partial z} = \nabla(D_\mathrm{i}\nabla C_\mathrm{i}) - k_{\mathrm{i}\leftrightarrow\mathrm{v}}(C_\mathrm{i}C_\mathrm{v} - C_{\mathrm{i,e}}C_{\mathrm{v,e}}) - q_\mathrm{i}^\mathrm{i}\ , \tag{38.78}$$

$$\frac{\partial C_\mathrm{v}}{\partial t} + V\frac{\partial C_\mathrm{v}}{\partial z} = \nabla(D_\mathrm{v}\nabla C_\mathrm{v}) - k_{\mathrm{i}\leftrightarrow\mathrm{v}}(C_\mathrm{i}C_\mathrm{v} - C_{\mathrm{i,e}}C_{\mathrm{v,e}}) - q_\mathrm{v}^\mathrm{v}\ , \tag{38.79}$$

where the intrinsic point defect consumption rates by clusters are given by

$$q_\mathrm{i}^\mathrm{i} = \left(\frac{\partial}{\partial t} + V\frac{\partial}{\partial z}\right) \times \left[\sum_{m_\mathrm{i}=2}^{m_\mathrm{i}=m_\mathrm{dis}} m_\mathrm{i}\phi_{m_\mathrm{i}} + \int_{m_\mathrm{i}=m_\mathrm{dis}+1}^{m_\mathrm{i}=m_\mathrm{max}} m_\mathrm{i} f_\mathrm{i}\,\mathrm{d}m_\mathrm{i}\right], \tag{38.80}$$

$$q_\mathrm{v}^\mathrm{v} = \left(\frac{\partial}{\partial t} + V\frac{\partial}{\partial z}\right) \times \left[\sum_{m_\mathrm{v}=2}^{m_\mathrm{v}=m_\mathrm{dis}} m_\mathrm{v}\phi_{m_\mathrm{v}} + \int_{m_\mathrm{v}=m_\mathrm{dis}+1}^{m_\mathrm{v}=m_\mathrm{max}} m_\mathrm{v} f_\mathrm{v}\,\mathrm{d}m_\mathrm{v}\right]. \tag{38.81}$$

The first term in the square brackets on the right-hand side of (38.80) accounts for the consumption of self-interstitials by the discrete i-clusters, and the second term accounts for the consumption of self-interstitials by the self-interstitial *FP-cluster*; (38.81) describes the vacancy consumption. The coupled model using the discrete rate equations for smaller clusters and the Fokker–Planck equation for larger clusters is still computationally expensive. As the model uses the described coupled approach, it is termed the *discrete–continuous* model in this chapter. For further details, the reader is referred to [38.27, 47–49]. As noted earlier, these researchers account for the kinetic interactions between monomers and clusters.

Results. Although the unsteady-state discrete–continuous model is described in this chapter, only the quantification of the steady-state CZ defect dynamics, at a fixed pull rate, involving the cluster growth has been reported in the literature thus far. The results obtained by the solution of the discrete–continuous model agree well with the results described by the rigorous model and the lumped model. The initial incorporation and the effects of the cooling rate and the incorporated dominant intrinsic point defect concentration on the cluster size distribution are captured very well. The evolution of the intrinsic point defect and the microdefect concentration profiles in a CZ crystal, as predicted by *Wang* and *Brown*, are shown in Fig. 38.15 [38.48]. The two-dimensional intrinsic point defect concentration fields and the microdefect distributions captured by the discrete–continuous model reported by *Brown* et al. are shown in Fig. 38.16 [38.49]. The physics of the CZ defect dynamics is quantified accurately by the applied model.

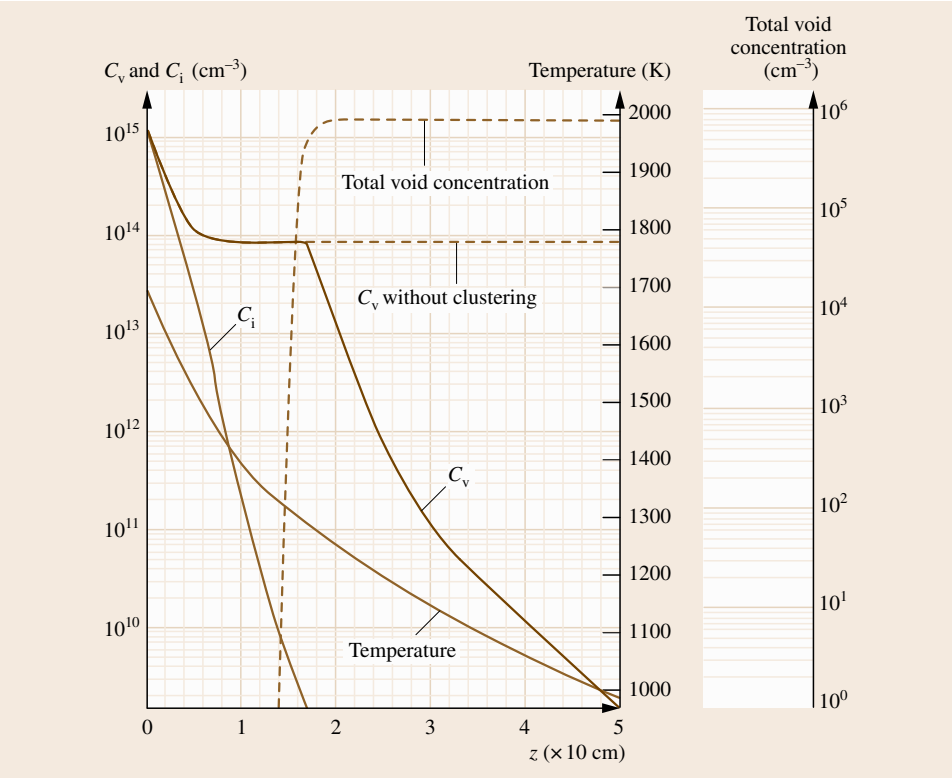


Fig. 38.15 The evolution of the intrinsic point defects and the microdefects as predicted by the discrete–continuous model (after [38.48], © Elsevier 2001)

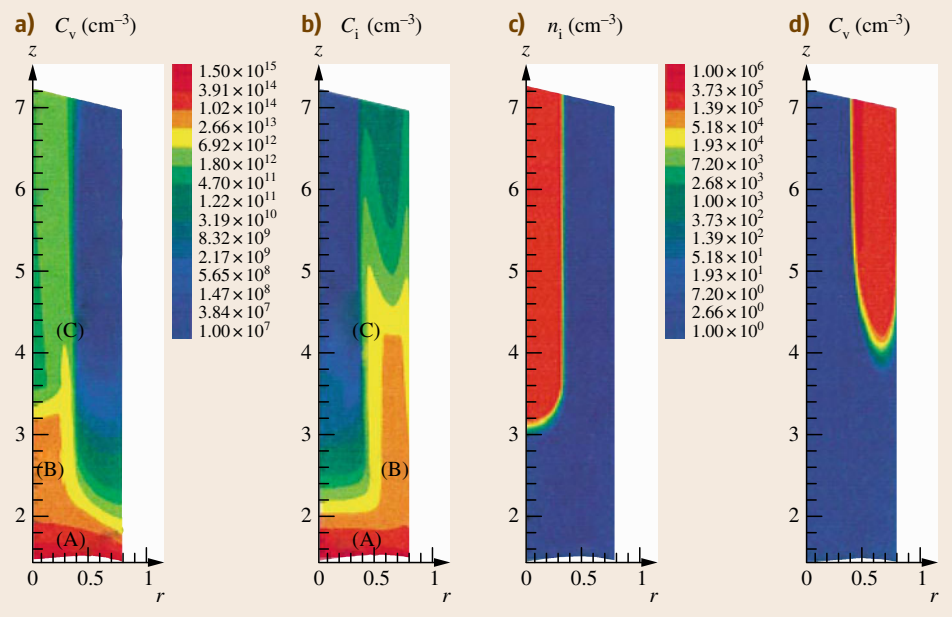


Fig. 38.16a–d Simulation results for steady-state crystal growth at $V = 0.6$ mm/min, showing **(a)** vacancy concentration, **(b)** self-interstitial concentration, **(c)** the total v-cluster density (> 50 nm diameter), and **(d)** the total i-cluster density (> 50 nm diameter). The *upper-case letters* in the figures are used by *Brown* et al. to describe the physics of the CZ defect dynamics and should not be mistaken for the figure labels, which are in lower case (after [38.49], © Elsevier 2001)

38.3 Czochralski Defect Dynamics in the Presence of Oxygen

As discussed so far, the aggregates of the intrinsic point defects commonly exist in silicon crystals grown by both the Czochralski process and the float zone (FZ) process. In addition, CZ crystals contain oxide particles, primarily silicon dioxide, termed oxygen clusters in this chapter. Oxygen clusters form only in CZ crystals, because CZ crystals, during their growth, incorporate oxygen in appreciable concentration from the crucible containing the silicon melt. Oxygen clusters in a growing CZ crystal are typically very small. These clusters facilitate the formation of stacking faults in the crystal subjected to selective heat treatments that generate self-interstitials (Fig. 38.3). Hence, oxygen clusters are popularly identified by these stacking faults known as oxidation-induced stacking faults (OSFs) [38.16].

The dynamics of the formation of various microdefects in CZ crystals is affected by many reactions involving the intrinsic point defects and oxygen, and their transport. This defect dynamics in the absence of oxygen has been discussed in detail in Sect. 38.2. The reported research on the direct quantification of the CZ defect dynamics in the presence of oxygen, in particular, and oxygen cluster formation in monocrystalline silicon, in general, are discussed in this section [38.48, 59, 65–70].

38.3.1 Reactions in Growing CZ Crystals

The essential aspect of understanding CZ defect dynamics is the quantification of the kinetics of all relevant reactions in a growing crystal. The Frenkel reaction involving the intrinsic point defects and silicon, the reactions involving vacancies and oxygen, and the aggregation reactions forming all microdefects influence the CZ defect dynamics.

Reactions Involving No Aggregation

The Frenkel reaction discussed in Sect. 38.1 and the reactions involving vacancies and interstitial oxygen (simply, oxygen) do not directly produce microdefects. The Frenkel reaction involves the mutual annihilation of a vacancy v and a self-interstitial i by their recombination to produce a silicon lattice atom Si and the backward production of a pair of a vacancy and a self-interstitial from a silicon lattice atom (reaction 38.1). Oxygen participates in a series of reversible reactions with vacancies and complexes of vacancies and oxygen in a growing CZ crystal. The following reactions involving vacancies and oxygen are of primary importance [38.59]:

$$\mathrm{v} + \mathrm{O} \rightleftharpoons \mathrm{vO}\,, \tag{38.82}$$

$$\mathrm{vO} + \mathrm{O} \rightleftharpoons \mathrm{vO_2}\,, \tag{38.83}$$

where vO and $\mathrm{vO_2}$ are vacancy–oxygen complexes. Each forward or reverse reaction listed above is considered to be an elementary reaction. The net rate of formation of a reacting species is given by the summation of the rates of formation of the species by each elementary reaction. At equilibrium, the net rate of production of any species is zero.

The Nucleation of the Intrinsic Point Defects and Oxygen

Octahedral voids or D defects are formed by the aggregation of vacancies. Globular B defects are formed by the aggregation of self-interstitials. A defects presumably form by the transformation of B defects. These microdefects are modeled as clusters of intrinsic point defects. The thermodynamics and kinetics of the formation of these microdefects is discussed in Sect. 38.2.

Oxygen clusters (O-clusters) are modeled as spherical aggregates of oxygen (silicon dioxide) [38.70]. The specific volume of an oxygen cluster is greater than that of silicon. Thus, the formation of an oxygen cluster is associated with the generation of stress. In the presence of vacancies, however, the clusters relieve stress by the consumption of vacancies. Oxygen cluster formation proceeds through a series of reactions involving oxygen and vacancies. Hence, this series of reactions is written as

$$\mathrm{O} + P_{m_\mathrm{O}} + \gamma\mathrm{v} + \tfrac{1}{2}\mathrm{Si} \rightleftharpoons P_{(m+1)_\mathrm{O}}\,, \tag{38.84}$$

where γ is the number of vacancies absorbed per oxygen atom participating in the reaction. It must be noted that an oxygen cluster containing m oxygen atoms also contains $m/2$ silicon atoms.

The volume (bulk) free-energy change associated with the formation of an oxygen cluster containing m oxygen atoms in an isolated element of silicon at a fixed temperature and composition is given by the contributions from the oxygen supersaturation and the vacancy supersaturation. Thus, the total free-energy change associated with the formation of an oxygen clus-

ter containing m oxygen atoms is

$$\Delta F_{\mathrm{O}}(m_{\mathrm{O}}) = \left[-m_{\mathrm{O}} k_{\mathrm{B}} T \ln \frac{C_{\mathrm{O}}}{C_{\mathrm{O,e}}} - \gamma m_{\mathrm{O}} k_{\mathrm{B}} T \ln \frac{C_{\mathrm{v}}}{C_{\mathrm{v,e}}} \right] + \left[\lambda_{\mathrm{O}} m_{\mathrm{O}}^{2/3} \right] . \tag{38.85}$$

The first term in the square brackets on the right-hand side of (38.85) is the volume (bulk) free energy change, and the second term is the energy required to form the surface of an oxygen cluster containing m oxygen atoms. The subscript "O" denotes both oxygen and oxygen clusters depending on the variable.

The formation kinetics of oxygen clusters is quite complex. An oxygen cluster undergoes morphological changes as it grows. This chapter does not address the details of these morphological changes. A broad macroscopic understanding of the oxygen cluster distribution is obtained by assuming these clusters to be spherical. The number of oxygen atoms in the critical cluster is obtained by maximizing the free-energy change ΔF with respect to m. The net rate of formation of stable oxygen clusters is obtained using the classical nucleation theory. Typically, it is accurate to assume that the formation rate of stable oxygen clusters is described by the diffusion-limited attachment of oxygen atoms to the critical oxygen clusters. For the sake of completeness, however, the attachment frequency is described both by the oxygen diffusion-limited mechanism and the vacancy diffusion-limited mechanism, depending on the ratio of $D_{\mathrm{v}} C_{\mathrm{v}}$ to $D_{\mathrm{O}} C_{\mathrm{O}}$. Hence, the formation rate of stable oxygen clusters per unit volume of silicon, or the oxygen nucleation rate, is given as

$$J_{\mathrm{O}} = \left[\eta^*_{\mathrm{O}\leftrightarrow\mathrm{O}} \right] \left[\left\{ 12\pi k_{\mathrm{B}} T \times \Delta F_{\mathrm{O}} \left(m^*_{\mathrm{O}} \right) \right\}^{-1/2} \times \left(k_{\mathrm{B}} T \ln \frac{C_{\mathrm{O}}}{C_{\mathrm{O,e}}} + \gamma k_{\mathrm{B}} T \ln \frac{C_{\mathrm{v}}}{C_{\mathrm{v,e}}} \right) \right] \times \left[\rho_{\mathrm{O}} \exp \left(- \frac{\Delta F_{\mathrm{O}} \left(m^*_{\mathrm{O}} \right)}{k_{\mathrm{B}} T} \right) \right] , \tag{38.86}$$

$$\eta^*_{\mathrm{O}\leftrightarrow\mathrm{O}} = \begin{cases} 4\pi R \left(m^*_{\mathrm{O}} \right) D_{\mathrm{O}} C_{\mathrm{O}} & \left| D_{\mathrm{v}} C_{\mathrm{v}} \geq \gamma D_{\mathrm{O}} C_{\mathrm{O}} \right. \\ \dfrac{4\pi R \left(m^*_{\mathrm{O}} \right) D_{\mathrm{v}} C_{\mathrm{v}}}{\gamma} & \left| D_{\mathrm{v}} C_{\mathrm{v}} < \gamma D_{\mathrm{O}} C_{\mathrm{O}} \right. \end{cases} , \tag{38.87}$$

where $\Delta F(m^*)$ is the free-energy change associated with the formation of a critical cluster containing m^* monomers; the subscript "O" denotes oxygen or oxygen clusters and the superscript asterisk denotes the critical clusters. $\eta^*_{\mathrm{O}\leftrightarrow\mathrm{O}}$ is the attachment frequency of oxygen atoms to a critical oxygen cluster. The second term in the square brackets on the right-hand side of (38.86) is the Zheldovich factor and the third term is the equilibrium concentration of the critical oxygen clusters. The discussed kinetics can now be applied in the development of the equations governing the CZ defect dynamics.

38.3.2 The Model

The model quantifying the CZ defect dynamics must account for the balances of all species, cluster formation, and cluster growth. All microdefects are approximated as spherical clusters. As discussed before, D defects are termed v-clusters, A and B defects are termed i-clusters, and the aggregates of oxygen (silicon dioxide) are termed O-clusters. At any given location of a growing CZ crystal, at a given time, one or more than one population of clusters formed at various other locations during the elapsed time period can exist. The clusters are assumed to be immobile; thus, they are only carried convectively from one location to the next by the physical movement of the growing crystal. In addition, there is a spatial distribution of these populations. A rigorous treatment of the spatial distribution of these cluster populations is computationally expensive. *Kulkarni* and *Voronkov* developed a lumped model that represents a population of clusters at any given location by an equivalent population of identical clusters [38.53]. Later, *Kulkarni* applied this model to quantify the CZ defect dynamics in the presence of oxygen [38.70]. In this chapter, this research reported by Kulkarni is discussed in detail.

The Governing Equations

The balance of self-interstitials includes their transport and their consumption by the Frenkel reaction and i-clusters

$$\frac{\partial C_{\mathrm{i}}}{\partial t} + V \frac{\partial C_{\mathrm{i}}}{\partial z} = \nabla \left(D_{\mathrm{i}} \nabla C_{\mathrm{i}} \right) + \left[k_{\mathrm{i}\leftrightarrow\mathrm{v}} (C_{\mathrm{i,e}} C_{\mathrm{v,e}} - C_{\mathrm{i}} C_{\mathrm{v}}) \right] - q^{\mathrm{i}}_{\mathrm{i}} . \tag{38.88}$$

The term in the square brackets in (38.88) is the net rate of formation (negative consumption rate) of self-interstitials per unit volume by the Frenkel reaction.

Vacancies are consumed by both v-clusters and O-clusters. In addition, vacancies participate in reactions

with self-interstitials, oxygen, and vO. Hence, the vacancy balance is written as

$$\frac{\partial C_\mathrm{v}}{\partial t}+V\frac{\partial C_\mathrm{v}}{\partial z}=\nabla\left(D_\mathrm{v}\nabla C_\mathrm{v}\right)+\left[k_{\mathrm{i}\leftrightarrow\mathrm{v}}(C_{\mathrm{i,e}}C_{\mathrm{v,e}}-C_\mathrm{i}C_\mathrm{v})-k_{\mathrm{v}\leftrightarrow\mathrm{O}}C_\mathrm{v}C_\mathrm{O}+k_{\mathrm{vO}}C_{\mathrm{vO}}\right]-q_\mathrm{v}^\mathrm{v}-q_\mathrm{O}^\mathrm{v}\,. \tag{38.89}$$

The rate constant for an elementary forward or an elementary reverse reaction is denoted by k. The subscripts of k indicate the reactants involved in a forward or a reverse reaction: "k_x" denotes the rate constant for the elementary reaction involving only x and "$k_{x\leftrightarrow y}$" indicates the elementary reaction involving x and y, where x and y represent the reacting species. q_x^y is the volumetric consumption rate of species y, denoted by the superscript, by the clusters containing species x, denoted by the subscript. The term in the square brackets in (38.89) is the net rate of vacancy production by reactions (38.1) and (38.82). The species vO is considered to be immobile. It is not directly consumed by clusters. Thus, the vO species balance must account only for the convection and reactions (38.82) and (38.83)

$$\frac{\partial C_{\mathrm{vO}}}{\partial t}+V\frac{\partial C_{\mathrm{vO}}}{\partial z}=\left(k_{\mathrm{v}\leftrightarrow\mathrm{O}}C_\mathrm{v}C_\mathrm{O}-k_{\mathrm{vO}}C_{\mathrm{vO}}-k_{\mathrm{vO}\leftrightarrow\mathrm{O}}C_{\mathrm{vO}}C_\mathrm{O}+k_{\mathrm{vO}_2}C_{\mathrm{vO}_2}\right)\,. \tag{38.90}$$

The species vO_2 is also considered to be immobile and it is also not directly consumed by clusters. It participates only in reaction (38.83)

$$\frac{\partial C_{\mathrm{vO}_2}}{\partial t}+V\frac{\partial C_{\mathrm{vO}_2}}{\partial z}=\left(k_{\mathrm{vO}\leftrightarrow\mathrm{O}}C_{\mathrm{vO}}C_\mathrm{O}-k_{\mathrm{vO}_2}C_{\mathrm{vO}_2}\right)\,. \tag{38.91}$$

Oxygen is in abundance

$$\frac{\partial C_\mathrm{O}}{\partial t}+V\frac{\partial C_\mathrm{O}}{\partial z}=0\,. \tag{38.92}$$

It is evident from (38.88–38.91) that the balance of the *excess total vacancy concentration*, defined as the difference between the sum of the concentrations of all species containing vacancies (v, vO, and vO_2) and the concentration of self-interstitials, $C_\mathrm{v}+C_{\mathrm{vO}}+C_{\mathrm{vO}_2}-C_\mathrm{i}$, is not explicitly affected by nonaggregation reactions (38.1), (38.82), and (38.83); this balance is written as

$$\frac{\partial(C_\mathrm{v}+C_{\mathrm{vO}}+C_{\mathrm{vO}_2}-C_\mathrm{i})}{\partial t}+V\frac{\partial(C_\mathrm{v}+C_{\mathrm{vO}}+C_{\mathrm{vO}_2}-C_\mathrm{i})}{\partial z}=\nabla\left(D_\mathrm{v}\nabla C_\mathrm{v}\right)-\nabla\left(D_\mathrm{i}\nabla C_\mathrm{i}\right)-q_\mathrm{v}^\mathrm{v}-q_\mathrm{O}^\mathrm{v}+q_\mathrm{i}^\mathrm{i}\,. \tag{38.93}$$

Assuming the reaction equilibrium for reactions (38.1), (38.82), and (38.83), the species balances (38.88–38.91) are defined by (38.93), (38.3), and the reaction equilibria [38.59]

$$\frac{C_{\mathrm{vO}}}{C_\mathrm{v}}=\sqrt{\frac{C_{\mathrm{vO}_2,\mathrm{e}}}{C_{\mathrm{v,e}}}}\,, \tag{38.94}$$

$$\frac{C_{\mathrm{vO}_2}}{C_\mathrm{v}}=\frac{C_{\mathrm{vO}_2,\mathrm{e}}}{C_{\mathrm{v,e}}}\,. \tag{38.95}$$

In CZ crystals, only oxygen nucleation facilitated by vacancies is of primary interest. The formation of O-clusters by ejection of self-interstitials is negligible. Once formed in the presence of vacancies, O-clusters initially grow by consuming vacancies without ejecting self-interstitials; later, when the vacancy concentration decreases, O-clusters can grow by ejection of self-interstitials. For the sake of simplicity, the growth of O-clusters by ejection of self-interstitials is ignored. When the vacancy concentration is sufficiently high, the O-cluster growth is assumed to be limited by the consumption of oxygen by the clusters; when the vacancy concentration is relatively low, the O-cluster growth is assumed to be limited by the consumption of vacancies by the clusters. Thus, O-clusters do not grow when vacancies are at equilibrium concentration. These approximations accurately quantify the density of O-clusters but underpredict their size by ignoring their growth by ejection of self-interstitials under vacancy-lean conditions. The assumptions used in the model are self-consistent, however, and provide meaningful insights into the CZ defect dynamics. If desired, the upper limit of O-cluster size can be quantified by simply assuming oxygen diffusion-limited cluster growth under all conditions; this assumption is not used in the formulation of the discussed model, although it can be implemented without much effort.

The diffusion-limited volumetric consumption rates of vacancies, self-interstitials, and oxygen by various clusters are defined following the methodology devel-

oped by *Kulkarni* and *Voronkov* [38.53]

$$q_{\mathrm{i}}^{\mathrm{i}} = 4\pi D_{\mathrm{i}}(C_{\mathrm{i}} - C_{\mathrm{i,e}})\,(U_{\mathrm{i}} n_{\mathrm{i}})^{1/2}\,, \tag{38.96}$$

$$q_{\mathrm{v}}^{\mathrm{v}} = 4\pi D_{\mathrm{v}}(C_{\mathrm{v}} - C_{\mathrm{v,e}})\,(U_{\mathrm{v}} n_{\mathrm{v}})^{1/2}\,, \tag{38.97}$$

$$q_{\mathrm{O}}^{\mathrm{v}} = \begin{cases} \gamma 4\pi D_{\mathrm{O}}(C_{\mathrm{O}} - C_{\mathrm{O,e}})\,(U_{\mathrm{O}} n_{\mathrm{O}})^{1/2} \\ \quad \left| D_{\mathrm{v}}(C_{\mathrm{v}} - C_{\mathrm{v,e}}) \geq \gamma D_{\mathrm{O}}(C_{\mathrm{O}} - C_{\mathrm{O,e}}) \right. \\ 4\pi D_{\mathrm{v}}(C_{\mathrm{v}} - C_{\mathrm{v,e}})\,(U_{\mathrm{O}} n_{\mathrm{O}})^{1/2} \\ \quad \left| D_{\mathrm{v}}(C_{\mathrm{v}} - C_{\mathrm{v,e}}) < \gamma D_{\mathrm{O}}(C_{\mathrm{O}} - C_{\mathrm{O,e}}) \right. \end{cases}. \tag{38.98}$$

The evolution of the auxiliary variable U, which is proportional to the surface area of the cluster population, is described by the cluster growth equation

$$\frac{\partial U_{\mathrm{i}}}{\partial t} + V\frac{\partial U_{\mathrm{i}}}{\partial z} = \frac{2D_{\mathrm{i}} n_{\mathrm{i}}}{\psi_{\mathrm{i}}^{\mathrm{i}}}(C_{\mathrm{i}} - C_{\mathrm{i,e}})\,, \tag{38.99}$$

$$\frac{\partial U_{\mathrm{v}}}{\partial t} + V\frac{\partial U_{\mathrm{v}}}{\partial z} = \frac{2D_{\mathrm{v}} n_{\mathrm{v}}}{\psi_{\mathrm{v}}^{\mathrm{v}}}(C_{\mathrm{v}} - C_{\mathrm{v,e}})\,, \tag{38.100}$$

$$\frac{\partial U_{\mathrm{O}}}{\partial t} + V\frac{\partial U_{\mathrm{O}}}{\partial z} = \begin{cases} \dfrac{2D_{\mathrm{O}} n_{\mathrm{O}}}{\psi_{\mathrm{O}}^{\mathrm{O}}}(C_{\mathrm{O}} - C_{\mathrm{O,e}}) \\ \quad \left| D_{\mathrm{v}}(C_{\mathrm{v}} - C_{\mathrm{v,e}}) \geq \gamma D_{\mathrm{O}}(C_{\mathrm{O}} - C_{\mathrm{O,e}}) \right. \\ \dfrac{2D_{\mathrm{v}} n_{\mathrm{O}}}{\gamma\psi_{\mathrm{O}}^{\mathrm{O}}}\left(C_{\mathrm{v}} - C_{\mathrm{v,e}}\right) \\ \quad \left| D_{\mathrm{v}}\left(C_{\mathrm{v}} - C_{\mathrm{v,e}}\right) < \gamma D_{\mathrm{O}}\left(C_{\mathrm{O}} - C_{\mathrm{O,e}}\right) \right. \end{cases}. \tag{38.101}$$

The total cluster density is directly obtained by the classical nucleation theory

$$\frac{\partial n_{\mathrm{i}}}{\partial t} + V\frac{\partial n_{\mathrm{i}}}{\partial z} = J_{\mathrm{i}}\,, \tag{38.102}$$

$$\frac{\partial n_{\mathrm{v}}}{\partial t} + V\frac{\partial n_{\mathrm{v}}}{\partial z} = J_{\mathrm{v}}\,, \tag{38.103}$$

$$\frac{\partial n_{\mathrm{O}}}{\partial t} + V\frac{\partial n_{\mathrm{O}}}{\partial z} = J_{\mathrm{O}}\,. \tag{38.104}$$

The representative radius of a cluster population at any location is given as

$$\mathcal{R}_{\mathrm{i}} = \left(\frac{U_{\mathrm{i}}}{n_{\mathrm{i}}}\right)^{1/2}, \tag{38.105}$$

$$\mathcal{R}_{\mathrm{v}} = \left(\frac{U_{\mathrm{v}}}{n_{\mathrm{v}}}\right)^{1/2}, \tag{38.106}$$

$$\mathcal{R}_{\mathrm{O}} = \left(\frac{U_{\mathrm{O}}}{n_{\mathrm{O}}}\right)^{1/2}. \tag{38.107}$$

The domain of computation is transient, because a CZ crystal is continuously pulled. The equation describing this domain transience must be solved with the discussed equations. Vacancies and self-interstitials are assumed to exist at equilibrium on all crystal surfaces including the melt–crystal interface; the concentrations of the vO and vO_2. species are determined by the re-

Table 38.2 Key properties of various species, including oxygen and vacancy–oxygen complexes, participating in reactions in growing CZ crystals

Property set II	Property set III
$D_{\mathrm{i}}\,(\mathrm{cm}^2/\mathrm{s}) = 0.19497\exp\left(\frac{-0.9\,(\mathrm{eV})}{k_{\mathrm{B}}T}\right)$	$D_{\mathrm{i}}(\mathrm{cm}^2/\mathrm{s}) = 4\times10^{-3}\exp\left(\frac{-0.3\,(\mathrm{eV})}{k_{\mathrm{B}}T}\right)$
$D_{\mathrm{v}}\,(\mathrm{cm}^2/\mathrm{s}) = 6.2617\times10^{-4}\exp\left(\frac{-0.4\,(\mathrm{eV})}{k_{\mathrm{B}}T}\right)$	$D_{\mathrm{v}}(\mathrm{cm}^2/\mathrm{s}) = 2\times10^{-3}\exp\left(\frac{-0.38\,(\mathrm{eV})}{k_{\mathrm{B}}T}\right)$
$D_{\mathrm{O}}\,(\mathrm{cm}^2/\mathrm{s}) = 1.3\times10^{-1}\exp\left(\frac{-2.53\,(\mathrm{eV})}{k_{\mathrm{B}}T}\right)$	$D_{\mathrm{O}}(\mathrm{cm}^2/\mathrm{s}) = 1.3\times10^{-1}\exp\left(\frac{-2.53\,(\mathrm{eV})}{k_{\mathrm{B}}T}\right)$
$C_{\mathrm{i,e}}\,(\mathrm{cm}^{-1}) = 6.1759\times10^{26}\exp\left(\frac{-4.0\,(\mathrm{eV})}{k_{\mathrm{B}}T}\right)$	$C_{\mathrm{i,e}}(\mathrm{cm}^{-1}) = 4.725\times10^{27}\exp\left(\frac{-4.3492\,(\mathrm{eV})}{k_{\mathrm{B}}T}\right)$
$C_{\mathrm{v,e}}\,(\mathrm{cm}^{-1}) = 7.52\times10^{26}\exp\left(\frac{-4.0\,(\mathrm{eV})}{k_{\mathrm{B}}T}\right)$	$C_{\mathrm{v,e}}(\mathrm{cm}^{-1}) = 1.2\times10^{27}\exp\left(\frac{-4.12\,(\mathrm{eV})}{k_{\mathrm{B}}T}\right)$
$C_{\mathrm{O}}\,(\mathrm{cm}^{-1}) = 9\times10^{22}\exp\left(\frac{-1.52\,(\mathrm{eV})}{k_{\mathrm{B}}T}\right)$	$C_{\mathrm{O}}(\mathrm{cm}^{-1}) = 9\times10^{22}\exp\left(\frac{-1.52\,(\mathrm{eV})}{k_{\mathrm{B}}T}\right)$
$C_{\mathrm{vO_2,e}}\,(\mathrm{cm}^{-1}) = \frac{C_{\mathrm{O}}^2}{5\times10^{22}}\exp\left(\frac{-0.5\,(\mathrm{eV})}{k_{\mathrm{B}}T}\right)$	$C_{\mathrm{vO_2,e}}(\mathrm{cm}^{-1}) = \frac{C_{\mathrm{O}}^2}{5\times10^{22}}\exp\left(\frac{-0.5\,(\mathrm{eV})}{k_{\mathrm{B}}T}\right)$
$\lambda_{\mathrm{i}}\,(\mathrm{eV}) = 2.75–2.85^*$	$\lambda_{\mathrm{i}}\,(\mathrm{eV}) = 2.75–2.85^*$
$\lambda_{\mathrm{v}}\,(\mathrm{eV}) = 1.75$	$\lambda_{\mathrm{v}}\,(\mathrm{eV}) = 1.75$
$\lambda_{\mathrm{O}}\,(\mathrm{eV}) = 1.7$	$\lambda_{\mathrm{O}}\,(\mathrm{eV}) = 1.7$
$\gamma = 0.42$	$\gamma = 0.42$

* Values between 2.75–2.85 eV give acceptable results. The simulations presented in this chapter were performed using $\lambda_{\mathrm{i}} = 2.75$ eV

action equilibria. As the final size of a cluster is far greater than its critical size, the initial size of clusters upon their formation is assumed to be zero. The initial length of a growing crystal is assumed to be finite but negligible. The discussed model describes the defect dynamics in CZ crystals growing under both steady as well as unsteady states.

There is a considerable uncertainty in the parameters describing the properties of many species participating in the CZ defect dynamics. Particularly, properties of self-interstitials are not very well known. Reported self-interstitial migration energies vary from 0.95 to 0.3 eV [38.1, 26, 34, 58–64]. As there are many parameters describing the properties of various species, a reasonably accurate prediction of the general characteristics of the observed microdefect distributions is possible for many different sets of values of these parameters. Two sets of properties, listed in Table 38.2, were used for simulations discussed in this section. The intrinsic point defect properties listed under property set II were derived from the property set proposed by *Kulkarni* et al., on the basis of further fine-tuning to fit experimental data [38.1, 70]. The intrinsic point defect properties proposed by *Voronkov* and *Falster* were fine-tuned to derive property set III [38.64, 70]. The formation energy of vO_2. species and the surface energies of all clusters were tuned to predict experimental data and well-accepted nucleation ranges of self-interstitials, vacancies, and oxygen. Both property sets yield qualitatively similar results.

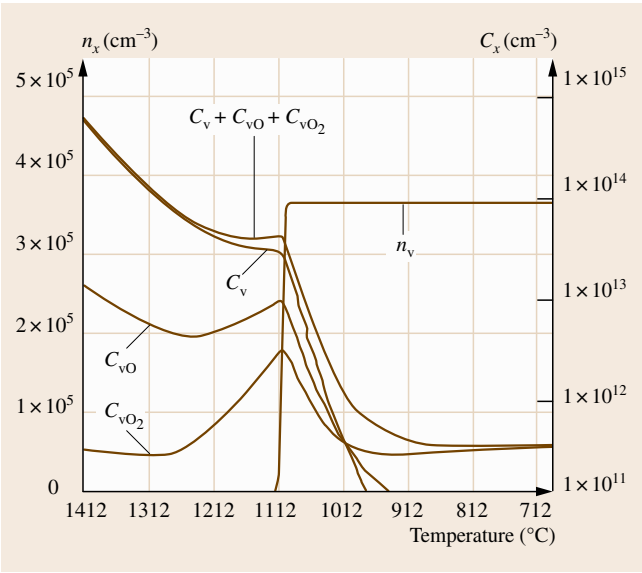


Fig. 38.17 Evolution of the concentrations of various reacting species and the density of v-clusters in a CZ crystal growing at a high rate ($G = 3.5$ K/mm, $V = 0.7$ mm/min, $C_O = 6.25 \times 10^{17}$ cm^{-3} or 12.5 ppma) (after [38.70])

38.3.3 Defect Dynamics in One-Dimensional Crystal Growth

A CZ crystal growing at a fixed rate through a fixed temperature field remains in a steady state, with respect to a fixed coordinate system, far from the regions formed in the beginning of the growth. A solution of the one-dimensional version of the developed model assuming only axial variation of the microdefect distribution provides insights into the basics of the CZ defect dynamics in the presence of oxygen. For these simulations, the crystal is assumed to grow through a temperature profile described by the linear dependence of $1/T$ on z, according to (38.6).

As discussed in Sect. 38.2, *Voronkov* described the conditions leading to the formation of various microdefects in growing FZ and CZ crystals in the early 1980s, in the absence of oxygen [38.19]. In the interest of continuity and in the context of understanding the influence of oxygen on the CZ defect dynamics, salient features of this theory are discussed again in this section. According to Voronkov's theory, an interplay between the Frenkel reaction and the transport of the intrinsic point defects of silicon determines the concentration fields of the intrinsic point defects in the vicinity of the melt–crystal interface. Vacancies and self-interstitials are assumed to exist at equilibrium at the interface. The temperature drop in the crystal in the vicinity of the interface drives the recombination of vacancies and self-interstitials, decreasing their concentrations. The developed concentration gradients drive the diffusion of vacancies and self-interstitials from the interface into the crystal. The vacancy concentration at the interface is higher than the concentration of self-interstitials, whereas self-interstitials diffuse faster. Thus, when the convective transport is relatively appreciable, vacancies remain the dominant species in the crystal; when the diffusion is relatively appreciable, self-interstitials are replenished at a higher rate from the interface and become the dominant species. Voronkov approximately quantified the relative effect of the convection over the diffusion by the ratio of V to G (V/G). At higher V/G, convection is appreciable; at lower V/G, diffusion is appreciable; at the criti-

cal V/G, the flux of vacancies is equal to the flux of self-interstitials. This analysis does not take into account the effect of oxygen. Oxygen introduces reactions involving vacancies and oxygen into this dynamics. In the presence of oxygen, free vacancies for the recombination with self-interstitials are supplied from the dissociation of vO and vO_2 species as well as from the interface, which is an infinite source. Hence, the presence of oxygen shifts the balance of this dynamics in favor of vacancies.

The evolution of the concentrations of v, vO, and vO_2 species as functions of the temperature in a CZ crystal growing at very high V/G, or under highly vacancy-rich conditions, is shown in Fig. 38.17. Near the interface, where the recombination rate is significant, concentrations of all three species decrease, as both free vacancies (v) and bound vacancies (bound as vO and vO_2) participate in the recombination reaction; the participation of free vacancies in the recombination is direct, whereas the participation of bound vacancies results through the coupling of reactions (38.1), (38.82), and (38.83). Once the recombination rate decreases, the *total vacancy concentration*, $C_v + C_{vO} + C_{vO_2}$, remains essentially constant. The *bound vacancy concentration*, $C_{vO} + C_{vO_2}$, increases with decreasing temperature because of a shift in the reaction equilibrium. Free vacancies, however, remain dominant and nucleate at around 1100 °C. The growth of voids predominantly consumes all vacancy species, as shown in Fig. 38.17. It must be noted, however, that the residual total vacancy concentration left at lower temperatures remains appreciable because of the binding between vacancies and oxygen.

At close to the critical yet moderately vacancy-rich condition, the free vacancy concentration does not remain high enough to form voids at higher temperatures, in a growing CZ crystal. As the temperature drops further, the concentration of bound vacancies increases. Under these conditions, free vacancies facilitate O-cluster formation at lower temperatures. The formation and growth of O-clusters predominantly consumes both free and bound vacancies, as shown in Fig. 38.18. The predicted total vacancy concentration at lower temperatures is approximate, because of the assumptions discussed in the previous section.

Conditions leading to the growth of crystals free of large v-clusters and i-clusters are desired in many microelectronic applications. Hence, the range of the pull rate within which a CZ crystal free of large clusters can be grown at different oxygen concentrations is of primary interest in industrial crystal growth. This range can be determined by a series of simulations at

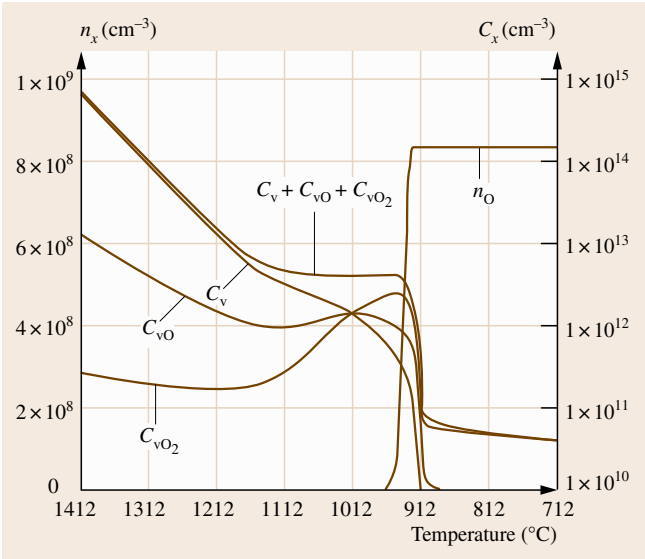


Fig. 38.18 Evolution of the concentrations of various reacting species and the density of O-clusters in a CZ crystal growing close to but moderately above the critical condition ($G = 3.5\,\mathrm{K/mm}$, $V = 0.48\,\mathrm{mm/min}$, $C_O = 6.25\times10^{17}\,\mathrm{cm}^{-3}$ or $12.5\,\mathrm{ppma}$) (after [38.70])

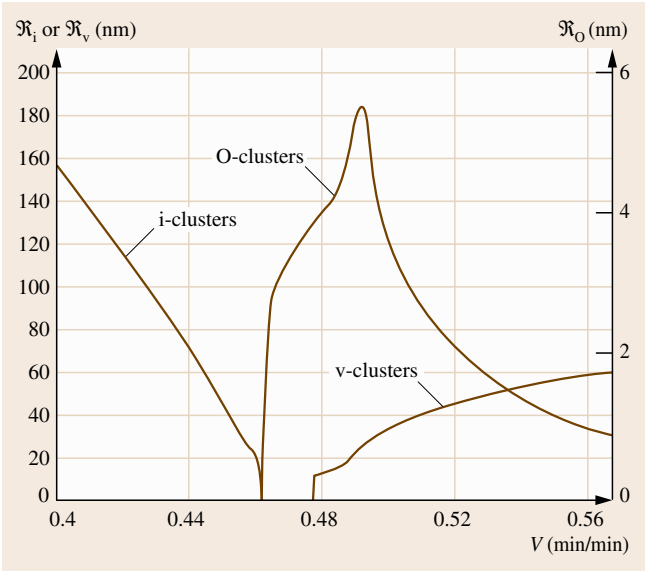


Fig. 38.19 The simulated sizes of various clusters as functions of the CZ crystal pull rate (after [38.70])

different oxygen concentrations at different fixed pull rates. These simulations, however, are computationally expensive. An approximate pull rate range that allows the growth of a crystal without large v-clusters and i-clusters, at a given oxygen concentration, can be determined by simulating the growth at a continuously decreasing rate such that the microdefect distribution in the crystal continuously shifts as a function of the pull rate. A direct but approximate correlation between the microdefect distribution and the pull rate can thus be obtained. Figure 38.19 shows one such simulation, defining the microdefect distribution as a function of the pull rate for a given oxygen concentration and temperature profile. It must be noted that the region free of large v-clusters and i-clusters contains O-clusters. O-clusters are quite large in the vicinity of the boundary between v-clusters and O-clusters, as they are formed at higher temperatures in the presence of a relatively higher vacancy concentration. The size of O-clusters decreases as the vacancy concentration during their formation decreases. A series of such simulations at different oxygen concentrations shows how oxygen affects the range of the pull rate within which the growth of a crystal free of large v-clusters and i-clusters is possible. Oxygen clearly expands this range because of the binding between vacancies and oxygen, as shown in Fig. 38.20. In this figure, v-clusters greater than 20 nm in radius are defined as large. It must be noted that this definition is arbitrary. Figure 38.20 also shows how V/G defining the boundary between O-clusters and i-clusters, known as the v/i boundary, shifts with the oxygen concentration in the discussed one-dimensional crystals growing through the temperature profile defined by (38.6). As discussed before, the presence of vO and vO_2 species near the interface increases the total vacancy concentration available for the recombination with self-interstitials, thus decreasing the V/G marking the v/i boundary; in effect, the crystal becomes marginally more vacancy rich in the presence of oxygen. The surface energy of voids is assumed to be a constant and independent of the oxygen concentration in all these simulations.

The series of one-dimensional simulations discussed in this section establishes the salient effects of oxygen on the CZ defect dynamics. All simulations discussed in this section are performed using property set III.

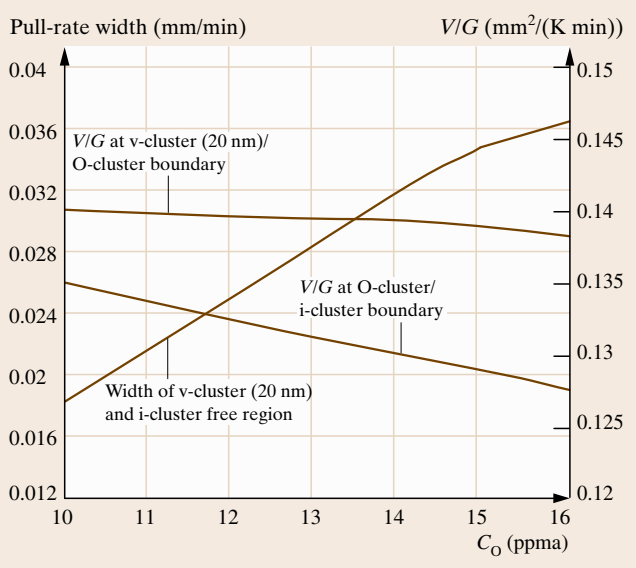


Fig. 38.20 The sensitivity of the microdefect distribution near the critical condition to the oxygen concentration. Note: The i-cluster and large v-cluster free region includes v-clusters smaller than 20 nm (radius) and O-clusters (after [38.70])

38.3.4 Defect Dynamics in Two-Dimensional Crystal Growth

The radial variation of the temperature field in a growing CZ crystal and the radial diffusion of the intrinsic point defects, induced by the lateral surface of the crystal and the radial variation of the intrinsic point defect concentration, introduce a two-dimensional variation of the microdefect distribution in the crystal. In addition, variation of the crystal pull rate, commonly observed in modern CZ processes, introduces an axial variation of the microdefect distribution. Hence, it is necessary to validate the discussed model by a comparison of its predictions with the microdefect distribution observed in a crystal grown under an unsteady state representing a variety of possible conditions in modern CZ growth. An experimental crystal was pulled by the varying rate shown in Fig. 38.21. The crystal was cut longitudinally and the microdefect distribution was characterized by the method of copper decoration followed by etching [38.38–40]. The crystal was assumed to grow through a fixed temperature field predicted by the commercial software MARC, using the algorithm developed by *Virzi* [38.37]. As shown in Fig. 38.21, v-clusters are observed in the regions of crystal grown at higher rates, and i-clusters are observed in the regions grown at lower rates. The dense bands at the

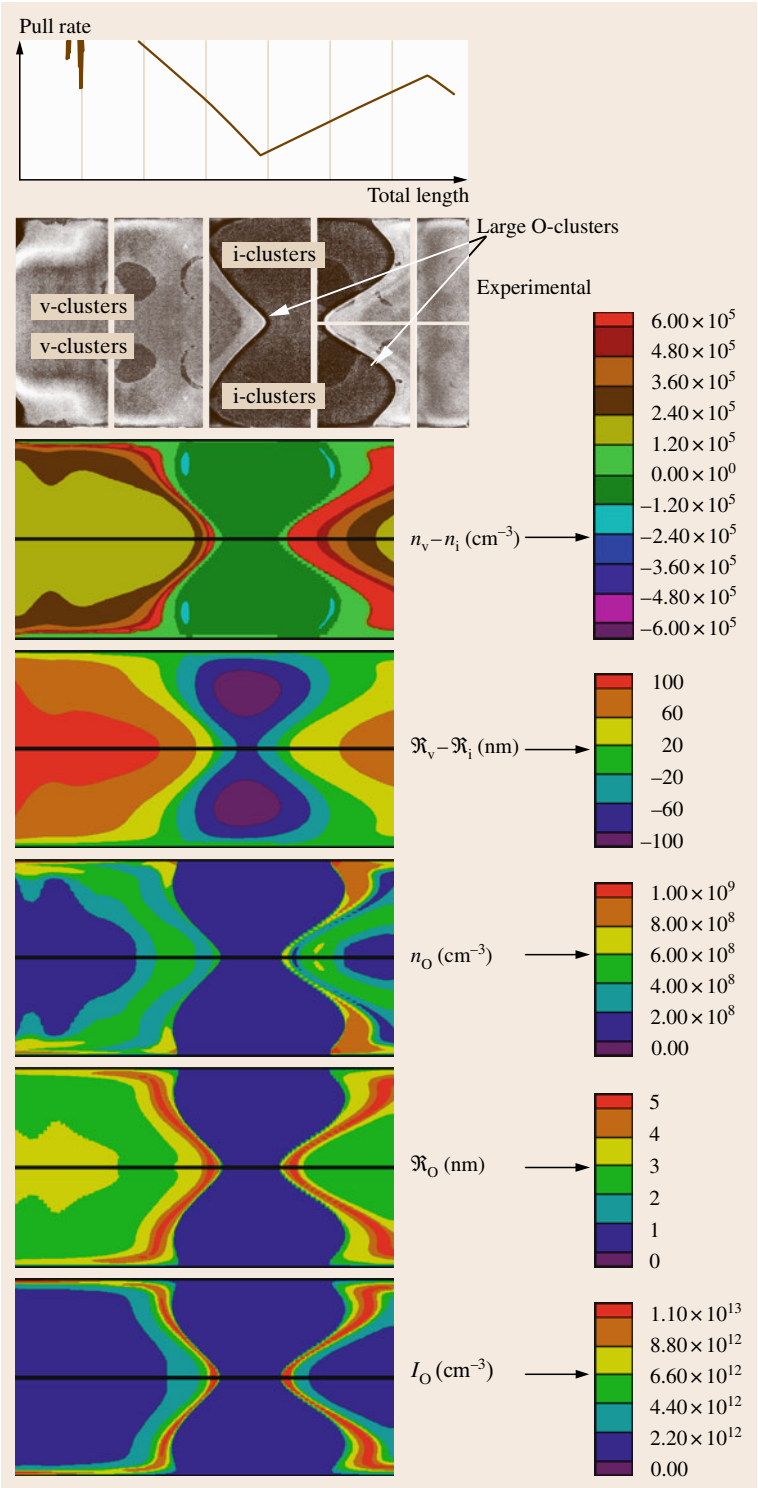


Fig. 38.21 Comparison of the predictions of the model, using property set III, with the experimental observations. The dense bands along the v/i boundary, near the edge of the region containing v-clusters, indicate intense oxygen precipitation (note: scales truncated for clarity) (after [38.70])

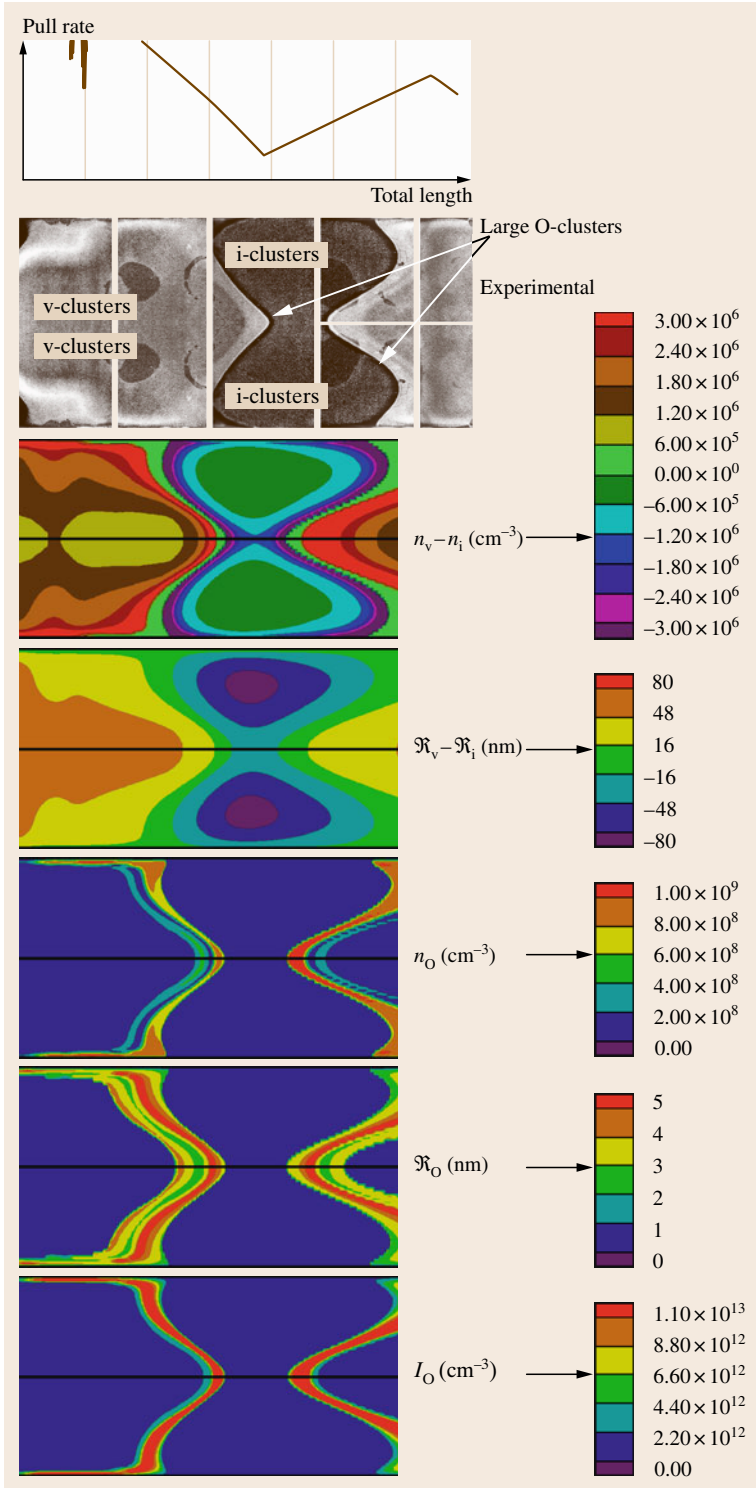


Fig. 38.22 Comparison of the predictions of the developed model, using property set II, with the experimental observations. The dense bands along the v/i boundary, near the edge of the region containing v-clusters, indicate intense oxygen precipitation (note: scales truncated for clarity) (after [38.70])

edge of the region containing large v-clusters, close to the featureless region in the vicinity of i-clusters, indicate the presence of large O-clusters at higher densities. This observation is only qualitative, as all O-clusters are not revealed by the applied characterization technique. The featureless band or region very close to the region of i-clusters can contain unobservable, small microdefects.

The microdefect distributions in the crystal, predicted using property set III, are also shown in Fig. 38.21. As i-clusters and v-clusters do not coexist, a distribution of $\mathcal{R}_v - \mathcal{R}_i$ simultaneously shows the distribution of both types of clusters; a positive value of $\mathcal{R}_v - \mathcal{R}_i$ indicates the size of v-clusters and a negative value of $\mathcal{R}_v - \mathcal{R}_i$ indicates the size of i-clusters. Similarly, the distribution of the densities of i-clusters and v-clusters are quantified by the distribution of $n_v - n_i$. The size and density distribution of O-clusters are shown separately in Fig. 38.21. It must be noted that the lumped model predicts the representative size of an entire population of microdefects present at any given location. In crystal regions formed under the appreciable influence of the convection of the intrinsic point defects, the dominant clusters formed are v-clusters; in regions formed under the appreciable influence of the diffusion of the intrinsic point defects, the dominant clusters formed are i-clusters; near the critical but moderately vacancy-rich condition, O-clusters are dominant. Since the lumped model applies the continuum-scale kinetics and thermodynamics in the prediction of the representative size and density of any given cluster population, it typically predicts negligible but finite oxygen precipitation even under highly vacancy-rich conditions. *Kulkarni* implemented various algorithms to ignore this negligible oxygen precipitation in the developed computer program that solves the model equations; but depending on the operating conditions and the species property set used for the simulation, a negligible but finite oxygen precipitation is predicted by the computer program even under highly vacancy-rich conditions [38.70]. This is evident in Fig. 38.21, which shows a negligible but definite presence of O-clusters in the regions grown under appreciable influence of convection of the intrinsic point defects. Both the size (radius) and density of O-clusters must be studied simultaneously to quantify the intensity of oxygen precipitation. The intensity of oxygen precipitation (I_O) can be measured by the amount of oxygen precipitated per unit volume of the crystal, which is given as

$$I_O = \tfrac{4}{3}\pi \mathcal{R}_O^3 \psi_O^O n_O \,. \tag{38.108}$$

As shown in Fig. 38.21, the intensity of oxygen precipitation is very high in the regions formed marginally above the critical condition and is negligible in the regions dominated by v-clusters and i-clusters.

The microdefect distributions, in the same crystal shown in Fig. 38.21, predicted using the property set II are shown in Fig. 38.22. The model captures the spatial distributions of various microdefects quite well. Although the predicted densities and sizes of various microdefects are not directly compared with the experimental data, the predictions are consistent with the reported experimental data [38.38, 39, 48].

38.4 Czochralski Defect Dynamics in the Presence of Nitrogen

As microdefects in silicon substrates produced from CZ crystals can adversely affect the performance of devices built on them, the development of crystal growth processes to reduce the size of the microdefects in growing crystals is of industrial significance. Various studies have shown that the microdefect distribution in CZ crystals can be influenced by the presence of nitrogen in general, and that the size of voids can be reduced, in particular [38.20, 71–75]. Various mechanisms to describe the effect of nitrogen on the CZ defect dynamics have been proposed [38.20, 72, 74, 75]. *Voronkov* and *Falster* quantified the effect of nitrogen on the void distribution in an isolated element of a growing CZ crystal [38.74]. An accurate model for the prediction of the effect of nitrogen on CZ defect dynamics, however, must account for the diffusion of mobile reacting species in a growing crystal. It must also incorporate reactions involving vacancies, self-interstitials, oxygen, and complexes of vacancies with oxygen and nitrogen, along with the formation of various microdefects in CZ crystals growing under both steady as well as unsteady states. This chapter discusses such a model, proposed and solved by *Kulkarni*, which quantifies both steady-state and unsteady-state CZ defect dynamics in the presence of nitrogen and oxygen [38.75].

38.4.1 The Model

An accurate quantification of CZ defect dynamics requires a treatment of a series of elementary reactions

describing the evolution of the populations of various clusters, which is computationally impractical. Therefore, reasonable approximations are required for the quantification of the microdefect distribution in a CZ crystal. The lumped model proposed by *Kulkarni* and *Voronkov* that approximates a population of microdefects of various sizes present at a given location in the crystal as an equivalent population of identical microdefects is adopted for the treatment of the growth of microdefects [38.53]. All microdefects are approximated as spherical clusters. The stable clusters are assumed to form according to the classical nucleation theory and to grow by diffusion-limited kinetics. The cluster formation and growth dynamics is treated simultaneously with the reactions involving vacancies, self-interstitials, oxygen, nitrogen, and complexes of oxygen and nitrogen.

Reactions in Growing CZ Crystals

The primary reaction influencing the CZ defect dynamics is the Frenkel reaction, which involves the mutual annihilation of a vacancy and a self-interstitial to produce a silicon lattice atom, and the generation of a pair of a vacancy and a self-interstitial from a lattice silicon atom, as shown in reaction (38.1). Oxygen participates in reactions with vacancies and the complexes of vacancies and oxygen, according to (38.82) and (38.83). Nitrogen exists as both a monomer and a dimer, and interacts with vacancies [38.74]

$$\mathrm{N}+\mathrm{N} \rightleftharpoons \mathrm{N}_2\,, \tag{38.109}$$

$$\mathrm{v}+\mathrm{N} \rightleftharpoons \mathrm{vN}\,. \tag{38.110}$$

The microdefect distributions in CZ crystals are affected by defect dynamics at fairly higher temperatures, at which nitrogen predominantly exists as N, N_2, and vN; hence, the presence of any other species containing nitrogen can be ignored [38.74]. The species N and N_2 are assumed to exist in the interstitial form in silicon, and vN can be viewed to exist in the substitutional form. The equilibria for reactions (38.109) and (38.110) are given by

$$\frac{C_\mathrm{N}^2}{C_{\mathrm{N}_2}} = K_{\mathrm{N}_2}^{\mathrm{N}\leftrightarrow\mathrm{N}}\,, \tag{38.111}$$

$$\frac{C_\mathrm{v} C_\mathrm{N}}{C_\mathrm{vN}} = K_\mathrm{vN}^{\mathrm{v}\leftrightarrow\mathrm{N}}\,, \tag{38.112}$$

where K is the equilibrium constant; the subscripts specify the reaction type.

The aggregation of the intrinsic point defects, vacancies and self-interstitials, and the aggregation of oxygen proceed as discussed in Sects. 38.2 and 38.3, respectively. The thermodynamics and kinetics of formation of all clusters are also discussed in Sects. 38.2 and 38.3.

The Governing Equations

The governing equations describe the balances of all species participating in the CZ defect dynamics. The total vacancy concentration, defined as the sum of the concentration of free vacancies (v), and of the cumulative concentration of bound vacancies (vO, vO_2, and vN), is not affected by reactions (38.82), (38.83), (38.109), and (38.110); both free vacancies and self-interstitials are consumed by the Frenkel reaction at an equal rate and are generated at an equal rate. Hence, the rate of change of the *total excess vacancy concentration*, defined as the difference between the total vacancy concentration and the concentration of self-interstitials, is affected only by the transport of individual species, and their consumption by the clusters

$$\begin{aligned}&\frac{\partial\left(C_\mathrm{v}+C_\mathrm{vO}+C_{\mathrm{vO}_2}+C_\mathrm{vN}-C_\mathrm{i}\right)}{\partial t}\\&\quad+V\frac{\partial\left(C_\mathrm{v}+C_\mathrm{vO}+C_{\mathrm{vO}_2}+C_\mathrm{vN}-C_\mathrm{i}\right)}{\partial z}\\&=\nabla\left(D_\mathrm{v}\nabla C_\mathrm{v}\right)-\nabla\left(D_\mathrm{i}\nabla C_\mathrm{i}\right)-q_\mathrm{v}^\mathrm{v}-q_\mathrm{O}^\mathrm{v}+q_\mathrm{i}^\mathrm{i}\,.\end{aligned} \tag{38.113}$$

All species except for vacancies and self-interstitials are assumed to be immobile. It must be noted that the formation of N_2 from N (reaction 38.109) requires N to be a mobile species. The diffusivity of N is implicitly assumed to be high enough to validate the assumption of equilibrium of reaction (38.109), but the diffusion flux of N is assumed to be negligible compared with the convective flux of N, in the overall species balance. These are reasonable simplifying assumptions. Equation (38.113) can also be derived from individual balances for all species. The overall nitrogen balance is given by

$$\begin{aligned}&\frac{\partial\left(C_\mathrm{N}+2C_{\mathrm{N}_2}+C_\mathrm{vN}\right)}{\partial t}\\&\quad+V\frac{\partial\left(C_\mathrm{N}+2C_{\mathrm{N}_2}+C_\mathrm{vN}\right)}{\partial z}=0\,.\end{aligned} \tag{38.114}$$

The assumption of equilibrium of reactions (38.1), (38.82), (38.83), (38.109), and (38.110) does not necessitate the solution of the individual rate equations. Equations (38.3), (38.94), (38.95), (38.111), and (38.112), which describe the reaction equilibria, and equations (38.113), (38.114), and (38.92) define the balances of i, v, O, vO, vO_2, N, N_2, and vN.

As discussed before, O-clusters are formed in the presence of vacancies without the ejection of silicon atoms as self-interstitials. Once formed, the clusters grow by the consumption of vacancies. At higher vacancy concentrations, the O-cluster growth is assumed to be determined by the diffusion-limited consumption of oxygen by the clusters. At lower vacancy concentrations, the O-cluster growth is assumed to be determined by the diffusion-limited consumption of vacancies by the clusters. The growth of O-clusters by the ejection of silicon atoms as self-interstitials, at lower vacancy concentrations, is ignored in the model. These approximations accurately predict the density of O-clusters but underpredict their size. If desired, the upper limit on the size of O-clusters can be computed by assuming the oxygen diffusion-limited growth of the clusters under all conditions. This assumption is not made in the discussed model, although it can be incorporated without much effort. The volumetric consumption rates of different species by different clusters (i, v, and O), and the average sizes and densities of the cluster populations are given by (38.96–38.107) in Sect. 38.3, as per the methodology developed by *Kulkarni* and *Voronkov* [38.53].

The domain of the calculation is transient, because a CZ crystal is continuously pulled. The described equations must be solved with the equations defining the domain transience, the temperature field in the crystal, and the segregation of nitrogen between the melt and the crystal at the melt–crystal interface.

The heat transport dynamics is assumed to be faster than the defect dynamics. Hence, quasi-steady-state assumptions for the calculation of the temperature field suffice. The temperature field is first calculated by the commercial software MARC, using the algorithms developed by *Virzi*, and then corrected using the experimentally measured interface shape [38.37, 53, 54].

The concentration field of oxygen in a crystal is determined by the process conditions. Hence, the concentration of oxygen in a crystal at the melt–crystal interface is directly assigned on the basis of the process conditions. The concentrations of all vacancy–oxygen complexes are defined by the reaction equilibria, and the intrinsic point defects (vacancies and self-interstitials) are assumed to be at their respective equilibrium concentrations at the interface.

The concentration of nitrogen in a crystal at the melt–crystal interface can be computed by assuming the thermodynamic segregation of nitrogen between the crystal and a well-mixed melt. If it is assumed that nitrogen exists as a monomer in the melt, the thermodynamic segregation of monomeric nitrogen between the crystal and the melt defines its concentration in the crystal, and the concentrations of N_2 and vN species in the crystal are defined by the reaction equilibria (38.111) and (38.112), respectively. The evolution of the nitrogen concentration in the well-mixed melt as a function of time is calculated by following the loss of silicon by solidification and by recognizing that the melt is the only source of nitrogen in all forms (N, N_2, and vN) in the crystal. It is a popular industrial practice, however, to define the relationship between the total nitrogen content in a crystal at the melt–crystal interface and the total nitrogen content in the melt (from which the crystal grows) using an apparent or effective segregation coefficient, which is just an engineering parameter that describes experimental observations. It is reiterated that this apparent segregation coefficient is not a thermodynamic parameter. This approach does not require any explicit assumptions on the state of nitrogen in a melt or a crystal and provides a simple yet reasonably accurate relationship between the total nitrogen content in the crystal and that in the melt. For a given total nitrogen content in the melt, the concentrations of all nitrogen-containing species in the crystal at the interface are defined by the apparent segregation of total nitrogen between the melt and the crystal, the stoichiometric relationship between the total nitrogen concentration and the concentrations of individual nitrogen containing species, and the reaction equilibria (38.111) and (38.112). The evolution of the nitrogen concentration in the well-mixed melt is quantified by following the silicon and nitrogen losses to the growing crystal across the interface. The results discussed in this section are generated using the apparent segregation coefficient, in accordance with popular industrial practice.

The accuracy of the predictions of the model depends on the accuracy of the parameters describing the properties of all reacting species and the parameters defining the thermodynamics of the discussed reactions. The parameters proposed by *Kulkarni* et al. for the simulation of the CZ defect dynamics in the presence of oxygen are used in this study [38.70]. The equilibrium constants defining the equilibria of reactions (38.109) and (38.110), $K_{N_2}^{N\leftrightarrow N}$ and $K_{vN}^{v\leftrightarrow N}$, are adopted from *Voronkov* and *Falster* [38.74]. The apparent segregation coefficient of nitrogen determining its segregation between a crystal and a melt is adopted from *Yatsurugi* et al. and others [38.20, 76, 77]. There is significant uncertainty in the parameters defining

the thermodynamics of the reactions and the properties of all species involved in the CZ defect dynamics.

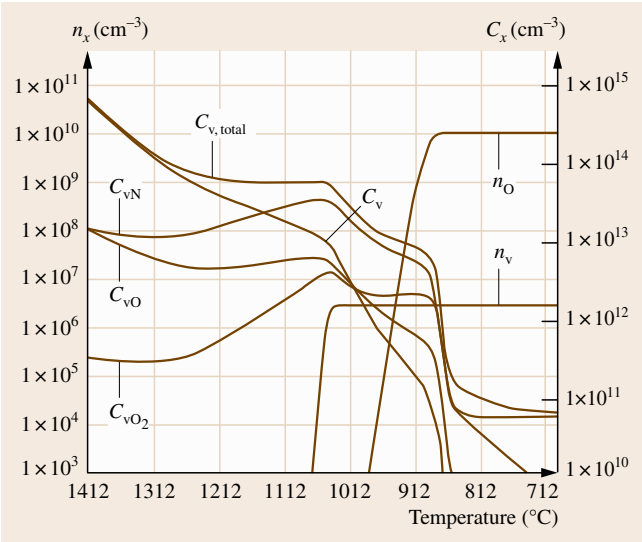


Fig. 38.23 Evolution of the concentrations of various reacting species in a nitrogen-doped CZ crystal growing at a high rate ($G = 3.9\,\mathrm{K/mm}$, $V = 0.8\,\mathrm{mm/min}$, $C_{N,\mathrm{total}} = 2\times10^{14}\,\mathrm{cm}^{-3}$, $C_O = 7\times10^{17}$ or 14 ppma)

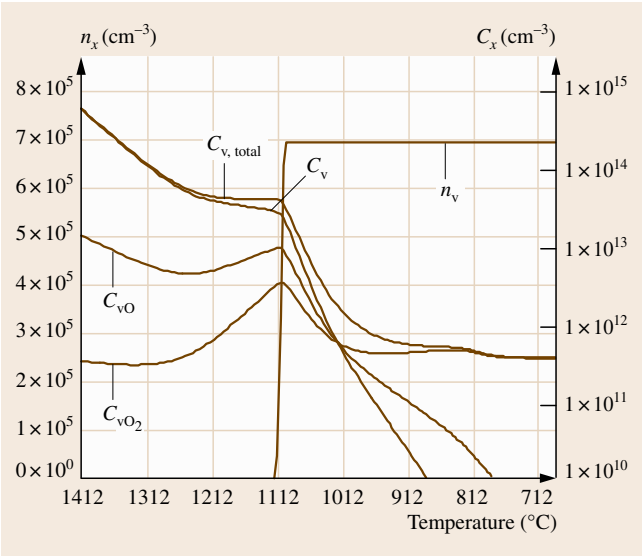


Fig. 38.24 Evolution of the concentrations of various reacting species in a CZ crystal growing at a high rate ($G = 3.9\,\mathrm{K/mm}$, $V = 0.8\,\mathrm{mm/min}$, $C_{N,\mathrm{total}} = 0\,\mathrm{cm}^{-3}$, $C_O = 7\times10^{17}$ or 14 ppma). Note the difference between the *scale* of the density axis in this figure and in Fig. 38.23

Hence, the parameter set used must be interpreted as just a suggestion.

The developed model can capture the microdefect distributions in the presence of oxygen and nitrogen in CZ crystals growing under steady as well as unsteady states.

38.4.2 CZ Defect Dynamics in One-Dimensional Crystal Growth

The important effects of nitrogen on the CZ defect dynamics are better demonstrated by the simulation of the growth of one-dimensional crystals showing no radial variation in the microdefect distributions. Further simplification can be achieved by simulating steady-state crystal growth. A crystal growing at a fixed rate through a fixed temperature field reaches a steady state far from the regions formed in the beginning of the growth. For simulation of crystal growth at steady state, the crystal is assumed to grow through a fixed representative temperature profile.

Nitrogen affects the reaction dynamics near the melt–crystal interface by interacting with vacancies according to reaction (38.110). The vN species produced by this interaction provides free vacancies (v) by dissociation, for the annihilation of self-interstitials. As free vacancies are also available from the interface, which is an infinite source, the interaction of nitrogen with vacancies shifts the balance of this dynamics in favor of vacancies. The crystal becomes richer in vacancies in the presence of nitrogen. Oxygen also shifts this balance in favor of vacancies by producing vO and vO_2 species, which provide free vacancies by their dissociation; but the effect of oxygen is weaker than that of nitrogen [38.59, 70, 74, 75].

The evolution of the concentrations of key reacting species, when the convective transport is appreciable, as functions of the temperature along the length of a nitrogen-doped CZ crystal growing from a very large melt (effectively infinite), is shown in Fig. 38.23. Since the melt is very large, the total nitrogen concentration, $C_N + 2C_{N_2} + C_{vN}$ or $C_{N,\mathrm{total}}$ in the crystal is effectively constant. Near the interface, all species containing vacancies participate in the recombination with self-interstitials; the participation of free vacancies is direct, and the participation of vN, vO, and vO_2 species is indirect through their dissociation. After the self-interstitial concentration drops significantly, the total vacancy concentration, $C_v + C_{vO} + C_{vO_2} + C_{vN}$ or $C_{v,\mathrm{total}}$ remains essentially constant until v-clusters are formed in appreciable numbers at lower temperatures.

Free vacancies directly participate in the formation and growth of v-clusters; the bound vacancies, vO, vO_2, and vN species, participate indirectly by their dissociation. The strong binding of vacancies with nitrogen reduces the free vacancy concentration before appreciable nucleation, compared with that in the absence of nitrogen, decreasing the vacancy nucleation temperature. The free vacancy consumption rate by v-clusters is low in the presence of lower vacancy concentration, facilitating their formation in appreciably higher densities. As the supply of total vacancies is fixed at the onset of the vacancy nucleation, the formation of v-clusters at higher densities reduces their size. The effect of nitrogen on the CZ defect dynamics is studied effectively by a comparison of the evolution of various reacting species in a growing CZ crystal in the presence of nitrogen (Fig. 38.23) with that in the absence of nitrogen (Fig. 38.24), under otherwise identical conditions.

A distinct effect of nitrogen is the facilitation of the formation of O-clusters in the presence of voids at appreciable densities. The strong binding between vacancies and nitrogen maintains appreciably high concentrations of bound and free vacancy species at lower temperatures at which vacancy-facilitated O-cluster formation and growth mark the CZ defect dynamics. After their formation, O-clusters become the dominant vacancy sinks compared with v-clusters, as shown in Fig. 38.23. In the absence of nitrogen, however, vN species is also absent, and both weakly bound (vO and vO_2) and free vacancies are consumed at higher temperatures primarily by v-clusters, suppressing the formation of O-clusters (Fig. 38.24). In a nitrogen-doped crystal exhibiting moderate dominance of vacancies near the interface, the free vacancy concentration does not remain high enough to form v-clusters at higher temperatures; at lower temperatures, however, O-clusters are formed in the absence of v-clusters (Fig. 38.25). It must be noted that the total vacancy concentration, even in the presence of clusters, remains negligible but finite in a fully grown CZ crystal because of the binding of vacancies with oxygen at lower temperatures.

The pull rate marking the boundary between O-clusters and i-clusters, known as the v/i boundary, represents the critical operating condition. The quantification of the variation of this critical pull rate with the nitrogen concentration, for otherwise fixed crystal growth conditions, is important. For a given temperature field in a crystal-puller, this pull rate can be quantified

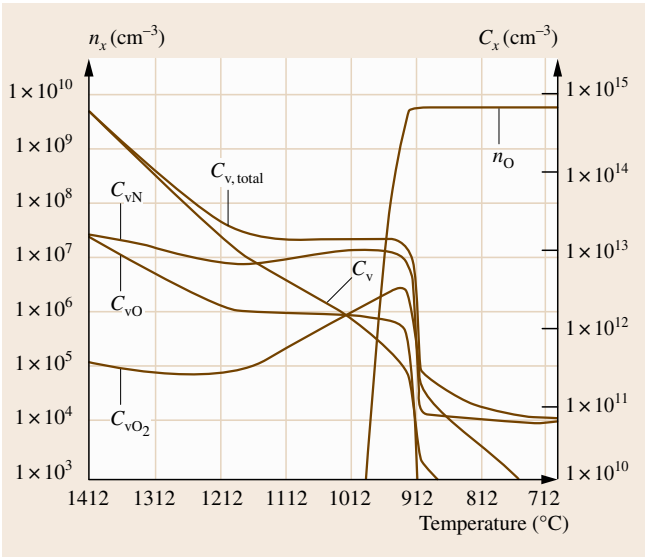


Fig. 38.25 Evolution of the concentrations of various reacting species in a nitrogen-doped CZ crystal growing close to the critical condition ($G = 3.9\,\mathrm{K/mm}$, $V = 0.50\,\mathrm{mm/min}$, $C_{N,total} = 2\times10^{14}\,\mathrm{cm^{-3}}$, $C_O = 7\times10^{17}$ or 14 ppma)

by a series of crystal growth simulations at various fixed or constant pull rates. These computations are numerous and can be expensive. An approximate variation of the cluster distribution in a crystal as a function of its pull rate can be captured by predicting this distribution in the crystal pulled by either continuously increasing or decreasing rate as shown in Fig. 38.26. One such simulation provides the approximate correlation between the cluster distribution in a crystal and its pull rate. The crystal growth conditions marking the v/i boundary can also be quantified in terms of the popular but approximate growth parameter V/G. The V/G marking this boundary, known as the critical V/G, continuously decreases with increasing nitrogen concentration (Fig. 38.27). The total vacancy concentration near the interface increases with increasing nitrogen concentration because of the strong binding between vacancies and nitrogen, resulting in a decrease in the critical V/G. This shift is particularly significant at higher nitrogen concentrations.

The conditions that allow the growth of CZ crystals in the absence of large clusters are also of interest in the crystal growth industry. Therefore, the quantification of the range of the crystal pull rate or V/G within which CZ crystals can grow free of large v-clusters and i-clusters is important. The large cluster-free range is

bracketed by the v/i boundary and the boundary between O-clusters and large v-clusters, arbitrarily defined

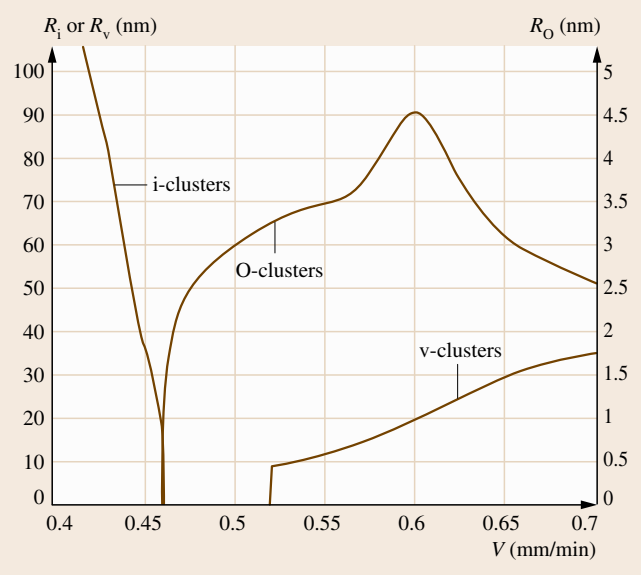


Fig. 38.26 The simulated sizes of various clusters as functions of the CZ crystal pull rate ($G = 3.9\,\mathrm{K/mm}$, $C_{N,total} = 2\times10^{14}\,\mathrm{cm^{-3}}$, $C_O = 7\times10^{17}$ or 14 ppma)

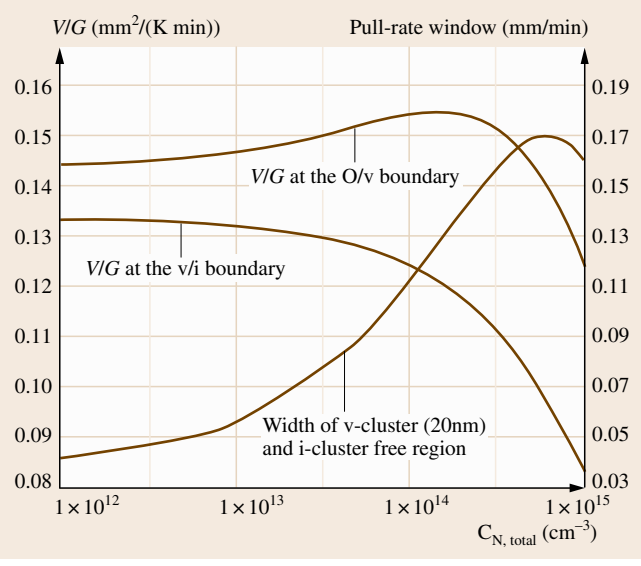


Fig. 38.27 The sensitivity of the cluster distributions near the critical condition to the nitrogen concentration. Note: The v-cluster and i-cluster free region includes v-clusters smaller than 20 nm ($G = 3.9\,\mathrm{K/mm}$, $C_O = 7\times10^{17}$ or 14 ppma)

to be 20 nm large in radius; the latter boundary is termed the O/v boundary. It must be noted that the crystals free of such large clusters contain O-clusters, as discussed before. The size of O-clusters increases with increasing concentration of the free vacancies available during their formation, as shown in Fig. 38.26. The V/G marking the O/v boundary increases with the nitrogen concentration first, reaches a maximum, and then decreases; the increase in this V/G at lower nitrogen concentration, in spite of the decrease in the V/G marking the v/i boundary, is caused by the strong binding between v and N that reduces the free vacancy concentration near the vacancy nucleation temperature range. At moderately higher nitrogen concentrations, however, the shift in the V/G marking the v/i boundary is too strong because of a significant increase in the total vacancy concentration near the interface, forcing a decrease in the V/G defining the O/v boundary. The pull rate range and the V/G range allowing the large cluster-free crystal growth still increase until the nitrogen concentration is significantly high. In the presence of nitrogen at very high concentrations, the total vacancy concentration in the crystal near the interface dramatically increases with the nitrogen concentration such that the free vacancy concentration around the vacancy nucleation temperature increases. Hence, the pull rate range (as well as V/G range) within which crystals free of large clusters can be grown also marginally decreases with increasing nitrogen concentration, at very high nitrogen concentration, as shown in Fig. 38.27. This predicted increase in the size of v-clusters with increasing nitrogen concentration does not agree with reported data [38.75, 78]. Therefore, the model predictions are far less accurate at very high nitrogen doping levels.

The salient features of the CZ defect dynamics and the strengths and weaknesses of the discussed model are thus adequately described by the discussed simulations of one-dimensional crystal growth.

38.4.3 CZ Defect Dynamics in Two-Dimensional Crystal Growth

The observed microdefect distributions in a CZ crystal vary along its axis and radius. This radial variation is caused by the radial variation of the temperature field and the radial diffusion of the intrinsic point defects induced by the temperature field and the lateral surface of a growing crystal. Axial variations are primarily induced by the change in the pull rate

and, in the case of nitrogen-doped crystals, by the axial variation in the nitrogen concentration. Quantification of these two-dimensional microdefect distributions requires a solution of the discussed model for two-dimensional crystal growth.

Kulkarni predicted the microdefect distribution in a crystal, at the end of its growth, pulled at a fixed rate through a fixed temperature field, shown in Fig. 38.28 [38.75]. The oxygen concentration in the crystal is assumed to be 16 ppma. The nitrogen concentration varies from 8.25×10^{13} to $6.7\times10^{14}\,\mathrm{cm}^{-3}$, from one end (left-hand side) of the crystal to the other (right-hand side), and from 8.4×10^{13} to $3.9\times10^{14}\,\mathrm{cm}^{-3}$, along the central axis within the cylindrical body of the crystal. The crystal growth at a fixed rate well above the pull rate at the critical condition is assumed. The crystal grows under an appreciable influence of convection. The central region of the crystal contains larger v-clusters and relatively smaller O-clusters; the peripheral region is dominated by smaller v-clusters, relative to those in the center, and larger O-clusters, relative to those in the center. This variation, away from the lateral surface of the crystal, is primarily caused by the radial variation of the temperature field; the relative effect of the axial convection with respect to the axial diffusion becomes weaker along the radial position of the crystal, away from the lateral surface, because of the radial variation of the temperature field. The defect distribution in the crystal near the lateral surface is strongly influenced by the surface-induced diffusion of the intrinsic point defects, details of which will be discussed in Sect. 38.5. The vacancy-rich region, defined as the region marked by the presence of v-clusters, O-clusters or any vacancy containing species, expands in the crystal as the nitrogen concentration increases. It must be noted that the microdefect distributions in the crystal when the crystal is still in contact with the melt at the end of its growth are shown in Fig. 38.28. Hence, the vacancy nucleation

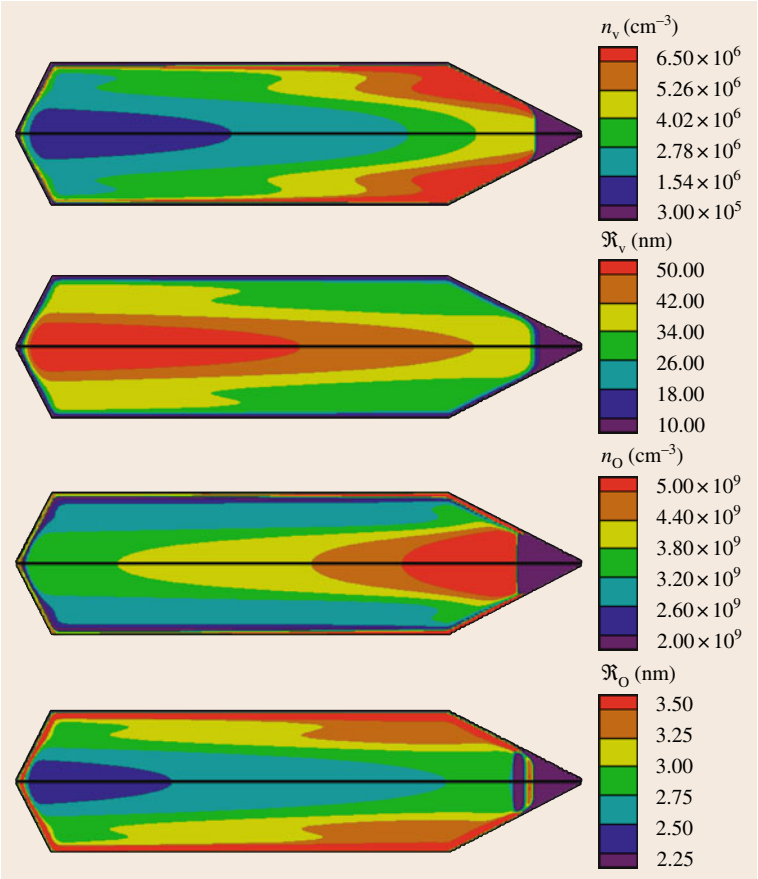


Fig. 38.28 The model predicted cluster distributions in a CZ crystal, pulled at a fixed rate well above the pull rate at the critical condition, at the end of its growth. Note: The melt–crystal interface is not flat. $C_{N,total}$ varies from 8.25×10^{13} to $6.7\times10^{14}\,\mathrm{cm}^{-3}$, from one end (*left-hand side*) of the crystal to the other (*right-hand side*), and from 8.4×10^{13} to $3.9\times10^{14}\,\mathrm{cm}^{-3}$, along the central axis within the cylindrical body of the crystal; $C_O = 16$ ppma. The crystal cooling rates are moderate. The *legend* for each plot is shown to the *right*

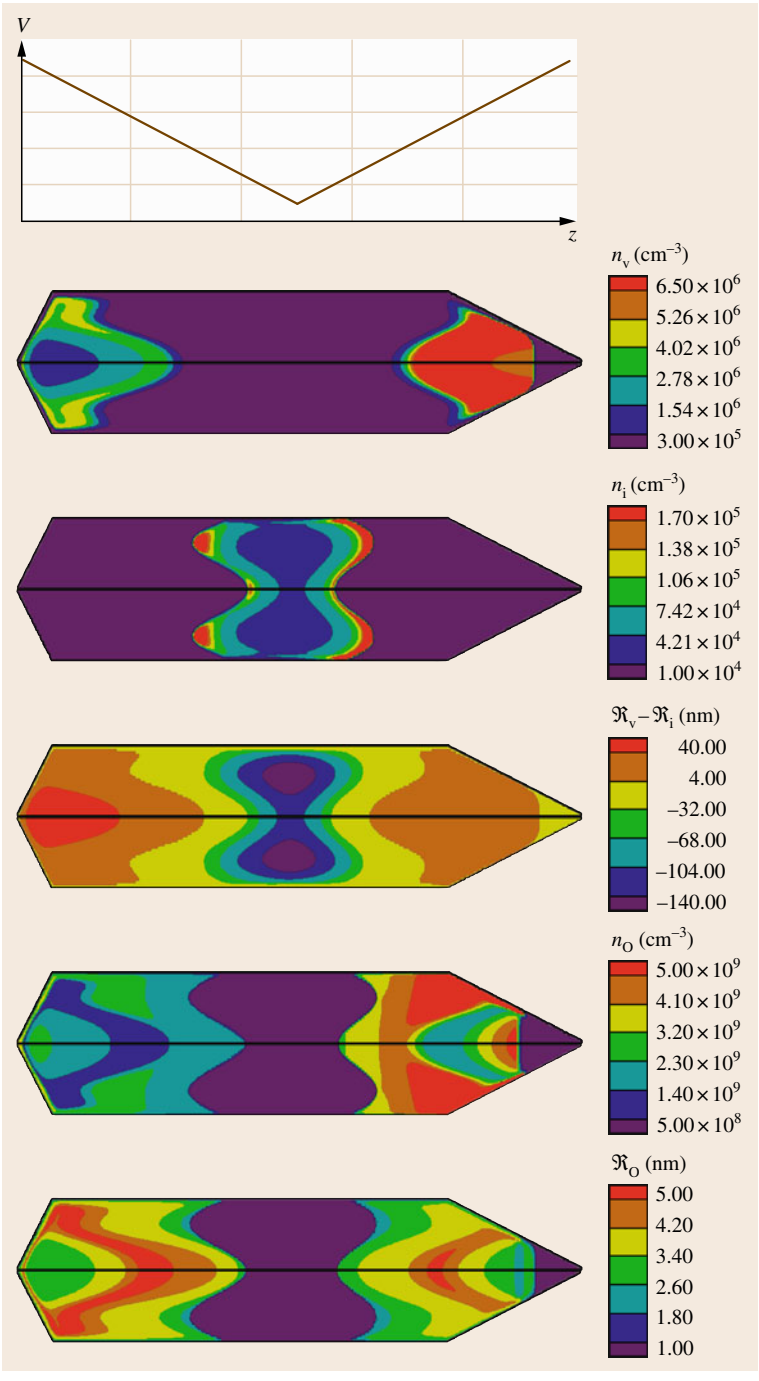


Fig. 38.29 The predicted cluster distributions in a simulated crystal at the end of its growth under transient conditions. Note: The melt–crystal interface is not flat. $C_{N,total}$ varies from 8.25×10^{13} to 6.7×10^{14} cm^{-3}, from one end (*left-hand side*) of the crystal to the other (*right-hand side*), and from 8.4×10^{13} to 3.9×10^{14} cm^{-3}, along the central axis within the cylindrical body of the crystal; $C_O = 16$ ppma. The *legend* for each plot is shown on the *right*

front and the O-cluster formation front are also visible near the right end of the crystal.

The capability of the discussed model can be demonstrated by predicting the microdefect distribu-

tion in the crystal pulled at the varying rate shown in Fig. 38.29. The oxygen and nitrogen concentrations in the crystal were set at the same level as those set for the simulation of the crystal discussed in Fig. 38.28 [38.75]. Hence, the effect of varying pull rate can be observed by a comparison between the cluster distributions shown in Fig. 38.28 and those shown in Fig. 38.29. The varying pull rate generates a distribution of all discussed clusters in this chapter. As i-clusters and v-clusters do not coexist, the simultaneous distributions of these clusters can be quantified by the distributions of $\mathcal{R}_v - \mathcal{R}_i$ and $n_v - n_i$. Positive values indicate the distribution of v-clusters and negative values indicate the distribution of i-clusters. Considering their differences, however, the density distributions of i-clusters and v-clusters are shown separately in Fig. 38.29. At a higher pull rate, v-clusters and O-clusters are formed; at a lower pull rate, i-clusters are formed; at a pull rate marking the moderate dominance of vacancies closer to the v/i boundary, only large O-clusters are formed. It must be noted that the microdefect distributions in the crystal at the end of its growth but still in contact with the melt are shown in Fig. 38.29. Hence, the vacancy nucleation front and the oxygen nucleation front are visible near the opposite end of the crystal.

The discussed model captures the salient features of the typical two-dimensional microdefect distributions observed in CZ crystals doped with nitrogen. In the presence of nitrogen at very high concentrations, however, the model predictions are less accurate.

38.5 The Lateral Incorporation of Vacancies in Czochralski Silicon Crystals

A crystal pulled at a very high rate, well above the critical condition, to meet industrial throughput goals, contains voids everywhere except very close to the lateral surface, where the surface-induced diffusion of both the intrinsic point defect species reduces the vacancy concentration at higher temperatures and allows the formation of oxygen clusters, during crystal growth (Fig. 38.30a). It must be noted that the Frenkel reaction couples the vacancy concentration with the self-interstitial concentration; hence, both vacancies that diffuse toward the lateral surface and self-interstitials that diffuse from the surface reduce the vacancy concentration near the surface. The temperature field of a typical growing CZ crystal shows significant radial variation; G increases significantly along the crys-



Fig. 38.30 **(a)** Typical microdefect distribution observed by copper precipitation followed by etching in a wafer produced from a CZ crystal grown at a very high rate well above the critical condition. **(b)** Typical microdefect distribution observed by copper precipitation followed by etching in a wafer produced from a CZ crystal grown through a radially highly nonuniform temperature field; note that a fraction of the crystal is grown above and the rest below the critical condition. **(c)** Typical microdefect distribution observed by copper precipitation followed by etching in a wafer produced from a CZ crystal grown close to the critical condition through a relatively radially uniform temperature field (Note: only $\frac{1}{4}$ of the wafer is shown) (after [38.79]) ▶

tal radius. Hence, many Czochralski crystals pulled at a moderate rate contain a central core of D defects and a peripheral region of A defects, as shown in Fig. 38.30b. The narrow detectable microdefect-free region close to the lateral surface of the crystal is also a result of the surface-induced diffusion of self-interstitials and vacancies, during its growth; while fast-diffusing self-interstitials dominantly contribute to their depletion near the surface, vacancies that diffuse from the surface also contribute to the depletion of self-interstitials through Frenkel reaction. Figure 38.30b also shows a very narrow detectable microdefect-free region termed the v/i boundary between the region of D defects and the region of A defects.

Recent advances in device industry demand either the reduction in the size of microdefects or their elimination from the CZ crystal. To address this requirement, many modern crystals are grown under a more radially uniform temperature field close to the critical condition. Hence, typical microdefects formed in these crystals are oxygen clusters and B defects. In practice, a perfect radial uniformity of the temperature field is not achieved; G marginally increases along the radius of a typical crystal. A crystal grown close to the critical condition, however, always exhibits a peripheral region, close to the surface, containing either oxygen clusters or both oxygen clusters and small voids, irrespective of the central microdefect distribution, as shown in Fig. 38.30c. *Kulkarni* described the physics of the evolution of microdefects in this region, known as the *edge ring* or the *peripheral ring* [38.79]. The peripheral ring loses its distinct appearance at higher crystal pull rates. This section addresses the evolution of the microdefect distribution in CZ crystals pulled at various rates, from the critical condition to highly vacancy-rich conditions.

38.5.1 General Defect Dynamics: A Brief Revisit

The theory of the initial point defect incorporation describes the defect dynamics in a growing crystal in the absence of oxygen [38.19]. According to this theory, the formation of microdefects in an oxygen-less crystal is driven by the supersaturation of crystal with either vacancies or self-interstitials. The concentrations of the intrinsic point defects in a growing crystal, away from its lateral surface, evolve primarily by an interplay between the species transport from the melt–crystal interface, both by convection and diffusion, and the Frenkel reaction. The intrinsic point defects of silicon exist at their respective equilibrium concentrations at the interface. The recombination of vacancies and self-interstitials in a growing crystal, driven by the decreasing temperature with increasing distance from the interface, generates their concentration gradients, inducing their diffusion from the interface into the crystal. The species convection is induced by the physical growth of the crystal. Self-interstitials diffuse faster than vacancies near the interface at higher temperatures but the vacancy concentration is higher than the self-interstitial concentration at the interface, where the equilibrium conditions prevail. Thus, when the species diffusion is relatively appreciable, or *below the critical condition*, the self-interstitial flux into the crystal from the interface is greater than the vacancy flux and the concentration of self-interstitials increases, away from the interface, with increasing effect of diffusion relative to convection. When the species convection is relatively appreciable, or *above the critical condition*, the vacancy flux is greater than the self-interstitial flux and the concentration of vacancies, away from the interface, increases with increasing effect of convection relative to diffusion. The dominant intrinsic point defect species annihilates the other species to very low concentrations and aggregates to form microdefects, at lower temperatures. At the critical condition, the vacancy flux is equal to the self-interstitial flux and the two intrinsic point defect species annihilate each other to very low concentrations, without forming large microdefects. This dynamics, defined ignoring the radial diffusion of the reacting species and the effects of oxygen, essentially establishes the incorporated concentration fields of the intrinsic point defects in a growing crystal, prior to the formation of microdefects.

Vacancies also interact with oxygen present in CZ silicon, producing various vacancy–oxygen complexes. The sum of the concentrations of all vacancy-containing species ($C_{v,total}$), defined as the total vacancy concentration, is relevant in considering the interactions between vacancies and self-interstitials. Free vacancies (v) directly participate in the annihilation of self-interstitials through the Frenkel reaction; bound vacancies (vO and vO_2) are also available for the annihilation of self-interstitials by dissociation. Because the total vacancy concentration available for the annihilation of self-interstitials increases in the presence of oxygen, the crystal becomes richer in vacancies with increasing oxygen concentration. Bound vacancies are assumed to be immobile; hence they affect the CZ defect dynamics only through reactions and physical movement of the crystal.

The model discussed in Sect. 38.3 describes the CZ defect dynamics in the presence of oxygen, and, hence, was applied by *Kulkarni* in his study on the peripheral ring formation. The results of this study are discussed below [38.79].

38.5.2 Defect Dynamics Under Highly Vacancy-Rich Conditions

The simulation of the defect dynamics in a CZ crystal growing close to the critical condition through a radially uniform temperature field can provide insights into the formation of the peripheral ring. To understand the difference between the evolution of microdefects in a CZ crystal growing at a very high rate well above the critical condition and that in a crystal growing close to the critical condition, the simulations of defect dynamics under both conditions must be performed.

Figure 38.31a shows the radial concentration profiles of key reacting species (C_x) in a crystal growing at a very high rate through a radially uniform temperature field. In the interior or bulk of the crystal away from the lateral surface, vacancy supersaturation and self-interstitial undersaturation develop. Since the equilibrium conditions prevail on the lateral surface of the crystal, the self-interstitial concentration gradually increases and the vacancy concentration gradually decreases along the radius of the crystal in the vicinity of the surface, at a given axial location. Both the diffusion of vacancies toward the surface and the diffusion of self-interstitials from the surface into the crystal contribute to the depletion of vacancies near the surface; self-interstitials, which diffuse faster than vacancies at higher temperatures, reduce the vacancy concentration by recombining with vacancies. Although the coupled transport of all species and reactions involving all species determine the species concentration fields in the crystal, the Frenkel reaction and the transport of self-interstitials and vacancies primarily influence their concentration fields, at higher temperatures.

As the vacancy supersaturation increases in the crystal interior (bulk) with decreasing temperature, voids are formed at around 1100 °C. Very close to the lateral surface, however, the radial diffusion of the intrinsic point defects maintains a lower concentration of vacancies. Vacancies in lower concentrations escape consumption by void formation at higher temperatures; they facilitate the formation of oxygen clusters at lower temperatures, around 950 °C. Figure 38.31b shows the radial profile of the amount of vacancies absorbed by voids, per unit volume of the crystal ($C_{\text{v-v-clusters}}$), and

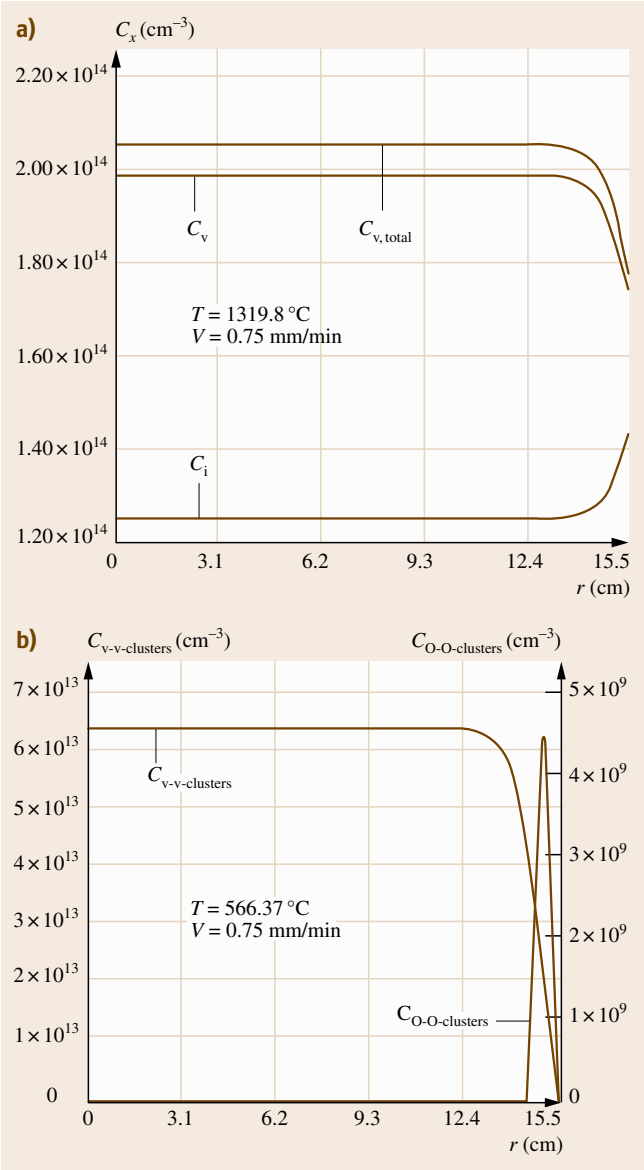


Fig. 38.31 (a) The radial profiles, close to the interface, of the concentrations of key reacting species in a CZ crystal growing well above the critical condition through a radially uniform temperature field. (b) The radial profiles of the concentration of vacancies in voids, per unit volume of crystal, and of oxygen in oxygen clusters, per unit volume of crystal, in a CZ crystal growing well above the critical condition through a radially uniform temperature field

of the amount of oxygen atoms absorbed by oxygen clusters, per unit volume of the crystal ($C_{\text{O-O-clusters}}$). There is negligible formation of oxygen clusters close

to the lateral crystal surface when the pull rate is very high, as shown in Fig. 38.31b.

38.5.3 Defect Dynamics Near the Critical Condition

A large microdefect-free crystal can be typically grown if the intrinsic point defect supersaturation in the crystal prior to the onset of nucleation is very low and radially uniform. This is achieved by growing the crystal through a relatively radially uniform temperature field, at close to the critical condition, at which an acceptable balance between the total vacancy flux (convection and diffusion of total vacancies) and the self-interstitial flux (convection and diffusion of self-interstitials) is achieved, in the vicinity of the interface. This discussion

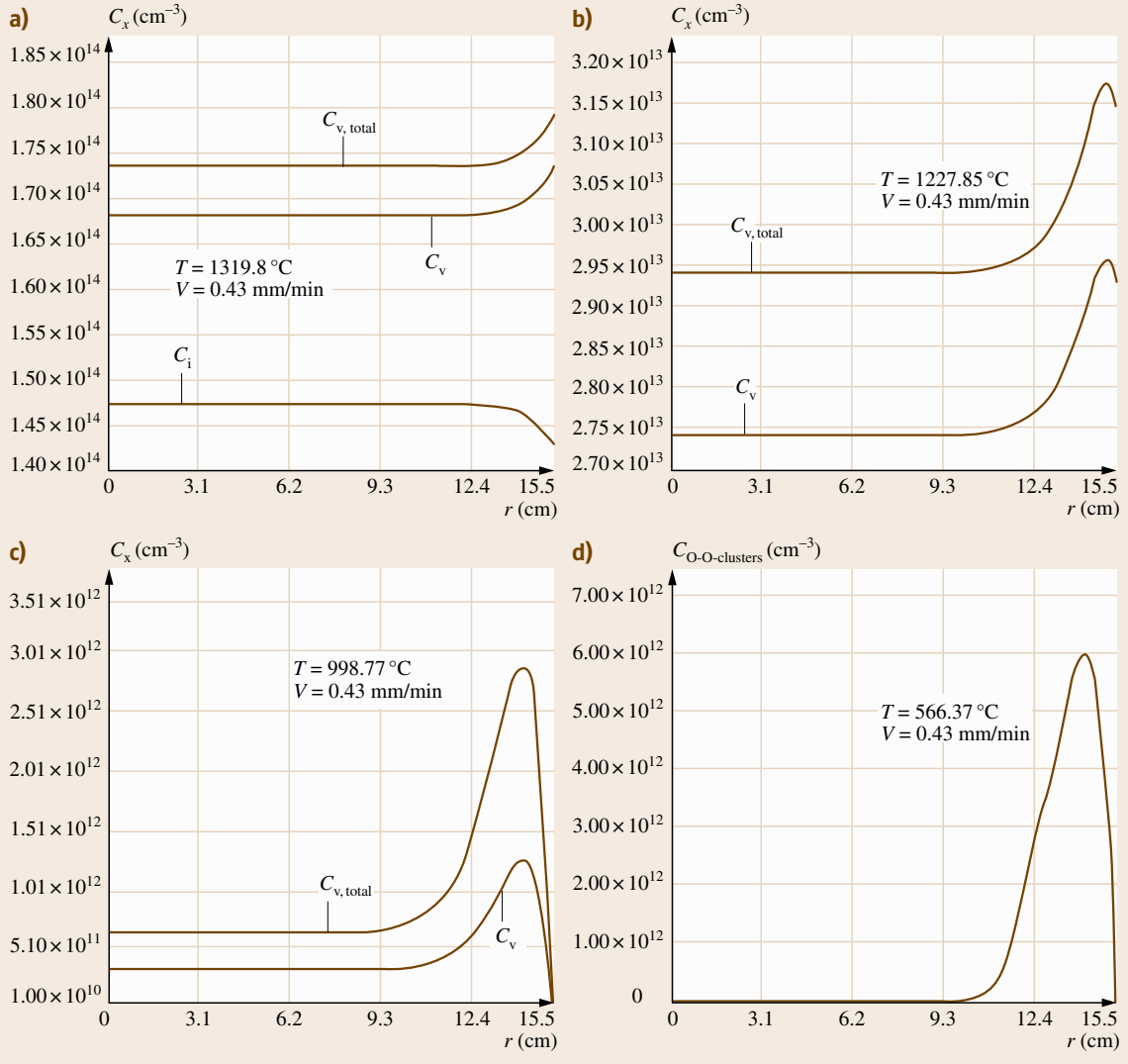


Fig. 38.32a–d The radial profiles of the concentrations of key reacting species in a CZ crystal growing close to the critical condition, through a radially uniform temperature field: **(a)** close to the interface, **(b)** at 1227.85 °C, **(c)** and at 998.77 °C. **(d)** The radial profiles of the concentration of oxygen in oxygen clusters, per unit volume of the crystal, in a CZ crystal growing close to the critical condition through a radially uniform temperature field

focuses on the growth of such marginally vacancy-rich or self-interstitial-rich crystals growing close to the critical condition. As total vacancies and self-interstitials remain in comparable concentrations under these growth conditions, the continuous interplay between the Frenkel reaction and the species transport remains relevant over the entire length of the crystal prior to the microdefect formation. Hence, the concentration fields of the intrinsic point defects and other species are sensitive to the relative shifts in their equilibrium concentrations and diffusivities with temperature.

Figure 38.32a shows the radial profiles of the concentrations of vacancies, self-interstitials, and total vacancies close to the interface of a crystal growing close to the critical condition through a radially uniform temperature field. The vacancy concentration in the bulk or the interior region of the crystal, away from the lateral surface, is higher than the self-interstitial concentration, but the crystal is marginally undersaturated with vacancies and marginally supersaturated with self-interstitials; the equilibrium conditions, however, prevail on the crystal surface. Therefore, self-interstitials diffuse towards the lateral surface from the crystal interior and vacancies diffuse from the lateral surface into the crystal. At higher temperatures, self-interstitials diffuse faster than vacancies and, hence, significantly affect the vacancy concentration field through the Frenkel reaction.

As the temperature decreases in the crystal, the diffusivities of both vacancies and self-interstitials decrease, along with their equilibrium concentrations, which prevail on the lateral surface. The Frenkel reaction remains at equilibrium, as shown by *Kulkarni* et al. [38.1]. All other homogeneous reactions are also assumed to be at equilibrium [38.70]. As the temperature decreases, the influence of the lateral surface-induced diffusion decreases and vacancies that were incorporated from the lateral surface at higher temperatures become the dominant species in the vicinity of the surface, annihilating self-interstitials to lower concentrations. Moreover, both the vacancy concentration at the lateral surface and the vacancy concentration in the crystal interior are lower than the vacancy concentration established in the vicinity of the surface. Hence, a relatively distinct vacancy supersaturation and a self-interstitial undersaturation in the vicinity of the lateral surface are established, as shown in Fig. 38.32b. In short, the interplay among decreasing equilibrium concentrations of all species, the decreasing effect of the species diffusion, and the fast reaction dynamics establishes a self-interstitial-lean and vacancy-rich ring near the lateral surface. At the stage shown in Fig. 38.32b, self-interstitials diffuse to this region from the lateral

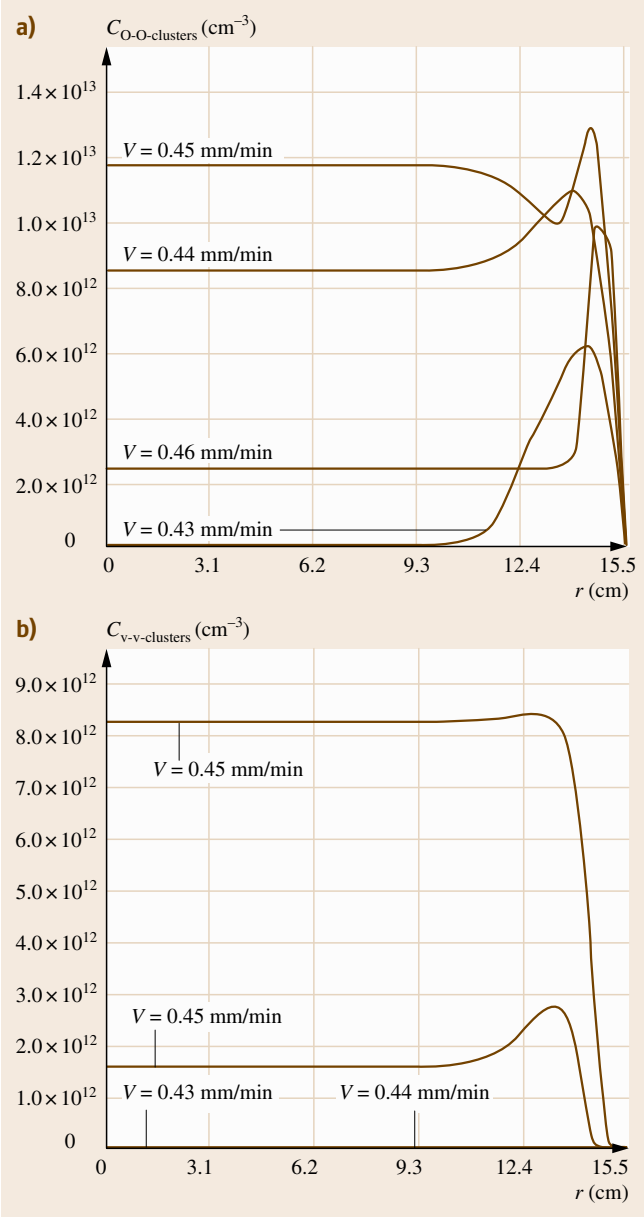


Fig. 38.33a,b The predicted radial profiles of the concentrations of **(a)** oxygen in oxygen clusters, per unit volume of crystal and **(b)** vacancies in voids, per unit volume of crystal, in CZ crystals growing close to the critical condition through a radially uniform temperature field, as functions of the pull rate. The predictions shown in this figure were generated using property set II

surface as well as from the crystal center and vacancies diffuse from this region, both radially inward and outward. Figure 38.32b shows that the interior (bulk) of the crystal is still undersaturated with vacancies when the region near the lateral surface is supersaturated with vacancies at around 1228 °C.

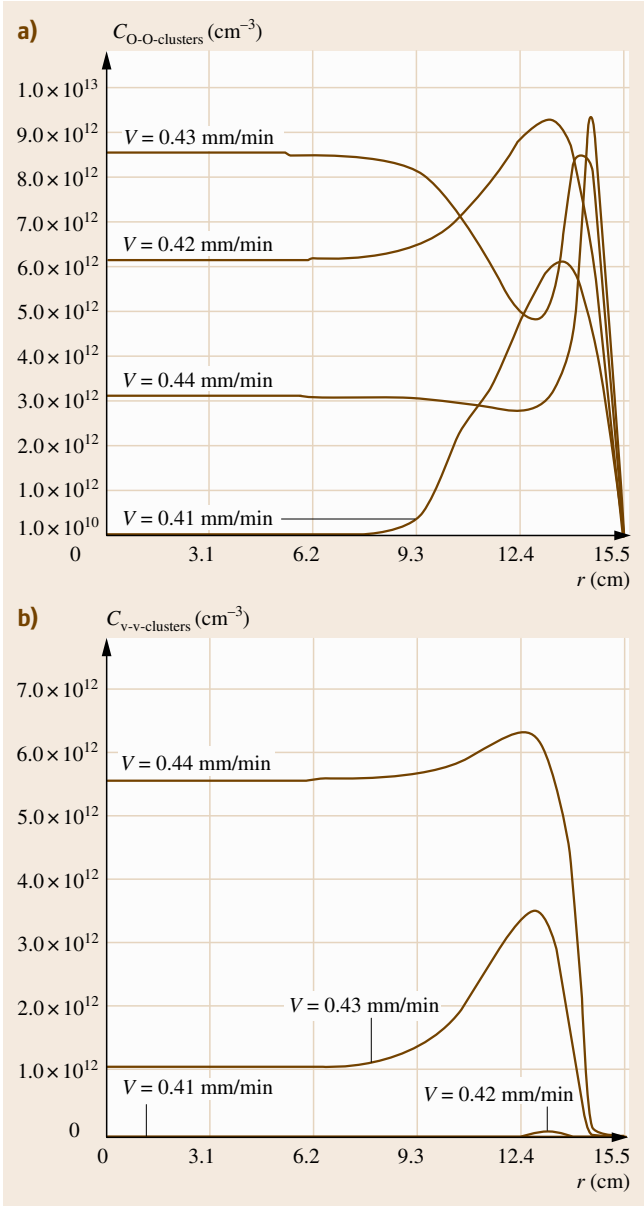


Fig. 38.34a,b The predicted radial profiles of the concentrations of **(a)** oxygen in oxygen clusters, per unit volume of crystal and **(b)** vacancies in voids, per unit volume of crystal, in CZ crystals growing close to the critical condition through a radially uniform temperature field, as functions of the pull rate. The predictions shown in this figure were generated using property set III

As the temperature drops further, the vacancy undersaturation in the crystal interior (bulk) decreases, because of a fast reduction in the equilibrium vacancy concentration. Eventually, the crystal can become marginally supersaturated with vacancies in the interior. The region near the lateral surface, however, becomes distinctly more supersaturated with vacancies, as shown in Fig. 38.32c.

In essence, the equilibrium conditions prevail on the surface of the growing crystal; near the critical condition, in the first stage of cooling at higher temperatures, the difference between the vacancy concentration and the self-interstitial concentration on the lateral crystal surface is much higher than that in the crystal interior (bulk), driving the vacancy enrichment of the near-surface region relative to the crystal interior; the vacancy supersaturation builds in the near-surface region with the decreasing temperature, because of the decreasing intrinsic point defect diffusivities and equilibrium concentrations.

It must be noted that, as the temperature decreases, the ratio of the bound vacancy concentration to the free vacancy concentration increases. In spite of the vacancy supersaturation established by the lateral surface, the vacancy concentration is not high enough to form large voids. Free and bound vacancies survive high temperatures to facilitate the formation of oxygen clusters (and small voids, in a few cases) at lower temperatures. The radial profile of the concentration of oxygen contained in oxygen clusters, per unit volume of the crystal, is shown in Fig. 38.32d. As can be observed, there is a distinct ring of oxygen clusters near the lateral surface of the crystal, which is the peripheral ring.

The peripheral ring becomes richer in vacancies but loses its distinct structure with increasing crystal pull rate. The microdefect distribution near the critical condition in a crystal as a function of the crystal pull rate is shown in Fig. 38.33. Figure 38.33a shows the radial profiles of $C_{\text{O-O-clusters}}$, and Fig. 38.33b shows the $C_{\text{v-v-clusters}}$ profiles. The competition between the void formation and the oxygen cluster formation for vacancies is evident as the pull rate increases. The peripheral ring cannot be formed when the relative effect of species convection is significant, because the vacancy supersaturation is maintained everywhere in the crystal under such condition, resulting in the diffusion of vacancies toward the lateral surface close to the interface, as already shown in Fig. 38.31. These results were generated

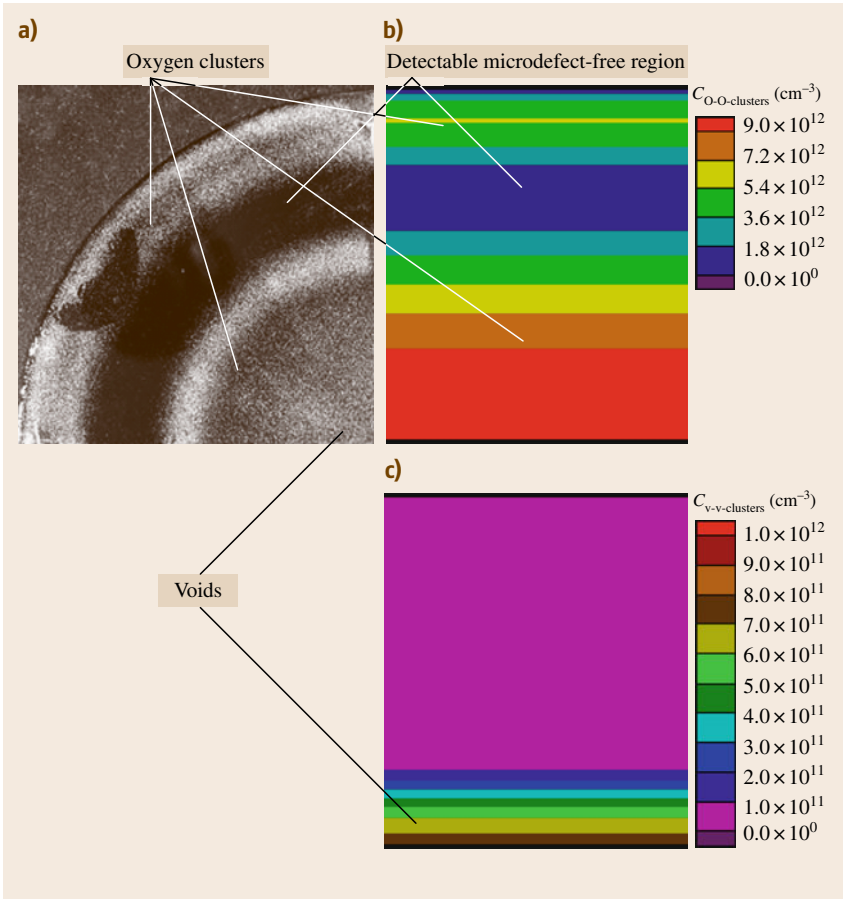


Fig. 38.35 **(a)** Observed microdefect distribution in a wafer showing a peripheral ring (Note: only 1/4 of the wafer is shown). **(b)** Predicted $C_{\text{O-O-clusters}}$ distribution in a longitudinal cross-section of the crystal showing a peripheral ring, **(c)** Predicted $C_{\text{v-v-clusters}}$ distribution in a longitudinal cross-section of the crystal showing a peripheral ring. (Note: the central core of oxygen clusters and voids surrounded by the featureless region is a result of a marginal radial variation in the temperature field)

using property set II; qualitatively similar results were generated using property set III (Fig. 38.34).

A comparison between the predicted microdefect distribution and the observed microdefect distribution in a crystal containing a peripheral ring is shown in Fig. 38.35. The simulation was performed using property set III. The central region of the crystal contains oxygen clusters and small voids. There is a detectable microdefect-free region between the central region containing oxygen clusters and the peripheral ring. This distribution shows that the crystal growth occurred through a radially nonuniform temperature field. However, the presence of the peripheral ring implies that the temperature field was sufficiently uniform.

Influence of the Intrinsic Point Defect Properties on the Edge Ring

There is uncertainty in the properties of many species influencing the CZ defect dynamics. Although the defect dynamics describing the evolution of the peripheral ring is driven by many parameters determining the properties of all participating species, the formation and migration energies of the intrinsic point defects have a particularly stronger influence.

The effect of the properties of the intrinsic point defects on the formation of the peripheral ring in a CZ crystal can be explained by considering the defect dynamics in the absence of oxygen, for simplicity. In the interior region of a CZ crystal growing through a uniform temperature field, the effects of radial species diffusion are negligible; thus, the evolution of the intrinsic point defect concentrations in this region near the melt–crystal interface is explained by the (axial) one-dimensional theory of initial incorporation [38.19]. According to this theory, the concentrations of the intrinsic point defects in an oxygen-less crystal growing at the critical condition through the temperature field approximated by a linear dependence of $1/T$ on z (38.6)

evolve according to (38.12) and (38.13), provided that the diffusivities of both species are temperature independent. In a narrow range of temperature near the interface, the assumption of constant species diffusivities, and a linear $1/T$ profile are acceptable, to some extent. Thus, (38.12) and (38.13) can be used as analytical guidelines for evaluating the evolution of the intrinsic point defect concentrations near the interface in a typical Czochralski crystal growing at the critical condition.

According to (38.4), (38.5), (38.12), and (38.13), both of the intrinsic point defect species are at their respective equilibrium concentrations in the interior of the crystal growing at the critical condition, if their formation energies are identical. The equilibrium conditions prevail on the crystal surface as well, because of the fast surface kinetics. Since there is no radial concentration gradient in the crystal under this condition, there is no surface-induced vacancy enrichment at the higher temperatures, which is required for the formation of the peripheral ring.

It is necessary to establish the vacancy undersaturation and self-interstitial supersaturation in the crystal interior (bulk) at higher temperatures to drive the near-surface vacancy enrichment. Moreover, the intensity of the peripheral ring increases with increasing difference between the equilibrium vacancy concentration and the vacancy concentration in the interior of the crystal, at higher temperatures, at the critical condition. This difference increases with decreasing vacancy formation energy, as given by (38.5) and (38.13). This difference also depends on the variation of the intrinsic point defect diffusivities with the temperature. Therefore, the assumption of constant diffusivities of the intrinsic point defects made by *Voronkov* must be relaxed [38.19]. At the critical condition, the self-interstitial flux is equal to the vacancy flux, i. e., the excess intrinsic point defect flux is zero.

When the self-interstitial migration energy is higher than the vacancy migration energy, the relative decrease in the self-interstitial diffusivity is higher than the relative decrease in the vacancy diffusivity, per unit length, in the vicinity of the interface, relatively reducing the replenishment of self-interstitials by diffusion from the interface. Hence, the relative self-interstitial diffusion flux decreases, at any given location in the vicinity of the interface, with increasing self-interstitial migration energy. As the relative self-interstitial diffusion flux decreases, proper changes in the other components of the excess intrinsic point defect flux (38.11) are required to satisfy the necessary equality of the vacancy flux and the self-interstitial flux, at the critical condition. These changes involve a reduction in the vacancy flux by reductions in the critical pull rate (for a given temperature field) and the vacancy concentration. As a result, the established vacancy concentration decreases in the crystal interior with increasing self-interstitial migration energy.

The effect of the formation and migration energies of the intrinsic point defects on the peripheral ring dynamics is multifold. The difference between the vacancy concentration and the self-interstitial concentration on the lateral surface of the crystal at higher temperatures increases with decreasing vacancy formation energy. If the migration energy of a self-interstitial is higher than that of a vacancy, the lateral diffusion of self-interstitials from the surface into the near-surface crystal region, at a later stage of cooling, becomes less significant, which maintains a higher vacancy concentration within the peripheral ring. The near-surface vacancy concentration field is influenced by the diffusion of both self-interstitials and vacancies, coupled with the Frenkel reaction, and not by the vacancy diffusion alone. It must be noted that the migration energy of a self-interstitial is higher than the migration energy of a vacancy, according to property set II and that the formation energy of a vacancy is lower than that of a self-interstitial, according to property set III. It is also important to note that the presence of oxygen and a deviation from the assumed crystal temperature field described by the linear variation of $1/T$ along the crystal axis also influence the dynamics of the formation of the peripheral ring, in addition to the formation and migration energies of the intrinsic point defects.

38.6 Conclusions

The quality of modern microelectronic devices built on silicon substrates manufactured by the Czochralski (CZ) and float-zone (FZ) crystal growth processes is determined by the distribution of the crystallographic imperfections known as microdefects present in the substrates. Microdefects are formed by the aggregation of the intrinsic point defects of silicon, vacancies (v) and self-interstitials (i), and by the vacancy-assisted aggre-

gation of oxygen (O) with silicon. Understanding the dynamics of the formation of these aggregates requires the quantification of this dynamics.

38.6.1 CZ Defect Dynamics in the Absence of Impurities

Early attempts at the quantification of defect dynamics were restricted to understanding the distribution of the intrinsic point defects in growing crystals, without accounting for the formation of microdefects, or the effects of oxygen. The theory of initial incorporation, proposed in 1982, explains and quantifies the conditions leading to the formation of microdefects [38.19]. According to this theory, in a growing CZ crystal, an interplay among the diffusion and convection of the intrinsic point defects and the Frenkel reaction driven by decreasing temperature near the vicinity of the melt–crystal interface establishes a dominant intrinsic point defect species, which nucleates at lower temperatures to form its aggregates. Vacancies, which exist at a higher concentration at the melt–crystal interface, remain dominant when the species convection is relatively appreciable; self-interstitials, which are fast diffusers in the higher temperature range, become the dominant species when the species diffusion is relatively appreciable. Vacancies form vacancy aggregates known as D defects; self-interstitials form self-interstitial aggregates known as A and B defects. This interplay is quantified in terms of the ratio of the crystal pull rate (V) to the magnitude of the axial temperature gradient at the interface (G), V/G, at any radial location of the crystal. Thus, D defects are formed at a radial location with higher V/G, A and B defects are formed at lower V/G, and no large microdefects are formed at V/G closer to its critical value. The radial distribution of microdefects in many crystals is reasonably explained by the radial variation of G.

The one-dimensional theory of initial incorporation does not effectively capture the effects of the radial diffusion of the intrinsic point defects induced by the radial variation in the temperature field and by the lateral surface of a crystal. Therefore, in the 1990s, many successful efforts were made to quantify the two-dimensional intrinsic point defect concentration fields in growing axisymmetric CZ crystals, to more accurately predict the microdefect distributions. These efforts, however, did not capture the effects of oxygen or the microdefect size distribution, as the actual events of the intrinsic point defect aggregation were not quantified.

Successful quantification of the microdefect distributions in CZ crystals, ignoring the effects of oxygen, was made only in and after the late 1990s. Capturing the CZ defect dynamics including the aggregation of the intrinsic point defects involves the quantification of the microdefect populations everywhere in a crystal. A microdefect population, at any given time and location in a crystal, consists of the microdefects formed at various locations during the elapsed time period. Therefore, a direct numerical treatment of the defect dynamics is impractical. Three popular approaches emerged to address the computational complexity of the CZ defect dynamics, in and after the late 1990s. All three approaches approximate the microdefects as spherical clusters; vacancy clusters or v-clusters represent D defects and self-interstitial clusters or i-clusters represent A and B defects.

The first approach, in its most rigorous form, requires the application of the classical nucleation theory to quantify the formation of the stable cluster nuclei and the prediction of the final microdefect distribution in a CZ crystal by quantifying the formation and growth histories of all clusters in the cluster populations everywhere in the crystal. Thus, the microdefect distribution is a function of location, time, and elapsed time. As the quantification of the CZ defect dynamics by the most rigorous form of this approach is computationally expensive, it is applied in its various reduced forms. An elegant treatment of the cluster formation and growth in an isolated element of a CZ crystal, cooled at a rate determined by its time–temperature history during the crystal growth, was first published in 1998. This treatment also quantifies the effect of the interesting interplay between the conditions facilitating the nucleation of the intrinsic point defects and the conditions facilitating the growth of the formed clusters on the final average size and the density of the clusters. This approximation, however, does not take into account the effect of the axial diffusion of the intrinsic point defects. The most elaborate model based on the first approach, reported in 2004, treats only the one-dimensional (axial) CZ defect dynamics. This model, however, is quite rigorous and addresses both the steady-state and unsteady-state defect dynamics. The shifts in the critical V/G associated with unsteady-state crystal growth were quantified for the first time by the unsteady-state model in 2004. The basic model used in the quantification of the defect dynamics by the first approach is termed the *rigorous model*.

The second approach followed in the quantification of the CZ defect dynamics approximates a population of

clusters present at any location in a crystal by another population of identical clusters. This is accomplished by approximating the average radius of the clusters in a population by the square root of the average of the squared radii of the clusters and by quantifying an auxiliary variable proportional to the total surface area of the clusters. The model based on this approach is termed the *lumped model*. The lumped model is computationally more attractive and effectively quantifies the two-dimensional microdefect distributions in CZ crystals.

The third approach followed in capturing the CZ defect dynamics couples the direct treatment of the reactions taking part in the aggregation events, for smaller clusters, and the Fokker–Plank formulation, for larger clusters. The formation of clusters and their growth is initially quantified by treating clusters of the same size and type as a species participating in the reactions. The discrete reaction rate equations derived by applying the classical nucleation theory are used for capturing the evolution of smaller clusters. The concentration (density) of all clusters of the same type exceeding a threshold size is treated as a continuous variable, which is a function of location, time, and the cluster size itself, by using the Fokker–Plank formulation. The Fokker–Plank equation (FPE) for larger clusters is derived from the discrete rate equations by a Kramers–Moyal expansion, treating the cluster size as a continuous independent variable. This approach is termed the *discrete–continuous approach* and is computationally very expensive.

38.6.2 CZ Defect Dynamics in the Presence of Oxygen

CZ crystals contain oxygen, which influences the defect dynamics. One-dimensional steady-state CZ defect dynamics in the presence of oxygen in a growing CZ crystal showing no radial variations has been modeled by both the Fokker–Planck formulation and the lumped model. The unsteady-state oxygen-influenced CZ defect dynamics accounting for the two-dimensional variations in microdefect distributions in growing CZ crystals, however, has been quantified only by the lumped model. The presence of oxygen primarily generates two bound vacancy species, vO and vO_2, and the aggregates of oxygen. The aggregates of vacancies are modeled as spherical v-clusters; the aggregates of self-interstitials are modeled as spherical i-clusters; and the aggregates of oxygen, primarily silicon dioxide, are modeled as spherical O-clusters.

The key element of the model describing the oxygen-influenced CZ defect dynamics is the vacancy-assisted formation of O-clusters. Effectively, all large O-clusters in the CZ growth are formed by absorbing vacancies, as the specific volume of O-clusters is greater than that of silicon. The growing O-clusters directly consume only free vacancies (v); as the free vacancy concentration decreases, however, more free vacancies are generated by the disassociation of vO and vO_2 species. Thus, both free vacancies and vacancies bound in vO and vO_2 species are consumed.

The type of microdefect formed in a given region in a crystal depends on the concentration of the intrinsic point defects and of vO and vO_2 species established a short distance away from the melt–crystal interface. In the regions marked by a high free vacancy concentration, voids or v-clusters are formed at higher temperatures by the nucleation of vacancies. The v-cluster growth consumes both free and bound vacancies. In the regions marked by a moderate free vacancy concentration, v-cluster formation is suppressed at higher temperatures; free and bound vacancies are consumed by the formation and growth of O-clusters. The binding between vacancies and oxygen allows survival of vacancies in the bound form at very low concentrations at lower temperatures even in the presence of voids and O-clusters. In the regions marked by the dominance of self-interstitials, i-clusters are formed. The concentration fields of the intrinsic point defects in the vicinity of the interface are established primarily by the interplay between the Frenkel reaction and the intrinsic point defect transport. Oxygen increases the effective vacancy concentration available for the recombination with self-interstitials by increasing the concentration of vO and vO_2 species and marginally aids the conditions leading to the survival of vacancies as the dominant intrinsic point defect species, for fixed crystal growth conditions. The increase in the pull rate range within which crystals free of large v-clusters and i-clusters can be grown, with increasing oxygen concentration, is also predicted and explained by the reported lumped model. This behavior is caused by a strong binding between vacancies and oxygen.

38.6.3 CZ Defect Dynamics in the Presence of Nitrogen

The microdefect distribution in a CZ crystal is strongly influenced by impurities such as nitrogen. A reasonably good approximation of the dynamics of microdefect

formation in growing CZ crystals in the presence of nitrogen and oxygen requires the quantification of the Frenkel reaction, the interactions between oxygen and vacancies and between nitrogen and vacancies, along with the aggregation events. Vacancies and self-interstitials annihilated by the recombination near the interface are partly supplied by their transport from the interface and by the dissociation of different reacting species. Nitrogen binding with vacancies is much stronger than that of oxygen with vacancies. The complexes of vacancies with nitrogen in particular, and oxygen to some extent, provide free vacancies by dissociation for the annihilation of self-interstitials near the melt–crystal interface. These free vacancies are available in addition to the vacancies from the interface, which is an infinite vacancy source. This dynamics increases the total concentration of vacancies in all forms, defined as the total vacancy concentration, and establishes relatively vacancy-rich conditions compared with those in the absence of nitrogen. In addition, nitrogen reduces the free vacancy concentration by strong binding at lower temperatures, reducing the vacancy nucleation temperature. The formation of vacancy aggregates in the presence of lower vacancy concentration results in an increase in their density at the expense of their size. The strong binding between vacancies and nitrogen provides an appreciable supply of vacancies for the facilitation of the formation and growth of the aggregates of oxygen with silicon at lower temperatures. Hence, vacancy aggregates and oxygen aggregates coexist in a wide range of crystal growth conditions in the presence of nitrogen.

The defect dynamics in nitrogen-doped CZ crystals, showing two-dimensional variations in the microdefect distribution and growing under both steady and unsteady states, has been quantified by the lumped model. The model approximates all aggregates as spherical clusters. Classical nucleation theory captures the formation of all clusters with a reasonable accuracy. The formed clusters grow by diffusion-limited kinetics. Vacancy clusters and self-interstitial clusters are formed by homogeneous nucleation of vacancies and self-interstitials, respectively. Oxygen clusters, because of their higher specific volume, are formed by the facilitation by vacancies. The growth of oxygen clusters is limited by the diffusion of oxygen when vacancies are in abundance and by the diffusion of vacancies when vacancies are scarce.

The conditions leading to the formation of different microdefects in CZ crystal growth have been quantified. The effects of varying pull rate and the nitrogen concentration are also captured. The model predictions agree well with the reported microdefect distributions in the presence of moderate and high nitrogen concentrations. In the presence of very high nitrogen concentrations, however, the model predictions are less accurate.

38.6.4 The Lateral Incorporation of Vacancies

A Czochralski crystal grown through a relatively radially uniform temperature field close to the critical condition is effectively free of large microdefects, but exhibits a prominent region of oxygen clusters (and small voids, in a few cases) near the lateral surface. Because of its ring-like appearance, the region is termed the peripheral ring. The peripheral ring is formed because of the vacancy supersaturation induced by the lateral surface of the crystal. The vacancies in moderate abundance facilitate the formation of oxygen clusters (and, possibly, small voids) at lower temperatures, giving the peripheral ring its characteristic appearance.

The equilibrium conditions prevail on the surface of a growing crystal; near the critical condition, in the first stage of cooling at higher temperatures, the difference between the vacancy concentration and the self-interstitial concentration on the lateral crystal surface is much higher than that in the crystal interior (bulk), driving the vacancy enrichment of the near-surface region relative to the crystal interior (bulk); the vacancy supersaturation builds in the near-surface region with the decreasing temperature, because of the decreasing intrinsic point defect equilibrium concentrations and diffusivities.

The moderate supersaturation of vacancies prior to the formation of oxygen clusters and small voids in the peripheral ring, when the bulk of the crystal is relatively vacancy lean, is driven by the Frenkel reaction, lateral surface-induced diffusion, and the crystal movement itself. This dynamics is termed the *lateral incorporation*. The lateral incorporation generates the *peripheral ring* only in crystals grown close to the critical condition.

References

38.1 M.S. Kulkarni, V.V. Voronkov, R. Falster: Quantification of unsteady-state and steady-state defect dynamics in the Czochralski growth of monocrystalline silicon, J. Electrochem. Soc. **151**, G663–G678 (2004)

38.2 W.C. Dash: Silicon crystals free of dislocations, J. Appl. Phys. **29**, 736–737 (1958)

38.3 W.C. Dash: Growth of silicon crystals free from dislocations, J. Appl. Phys. **30**, 459–474 (1959)

38.4 T. Abe, T. Samizo, S. Maruyama: Etch pits observed in dislocation free silicon crystals, Jpn. J. Appl. Phys. **5**, 458–459 (1966)

38.5 A.J.R. de Kock: Vacancy clusters in dislocation-free silicon, Appl. Phys. Lett. **16**, 100–102 (1970)

38.6 A.J.R. de Kock: The elimination of vacancy-cluster formation in dislocation-free silicon crystals, J. Electrochem. Soc. **118**, 1851–1856 (1971)

38.7 A.J.R. de Kock: Microdefects in dislocation-free silicon and germanium crystals, Acta Electron. **16**, 303 (1973)

38.8 A.J.R. de Kock, P.J. Roksnoer, P.G.T. Boonen: Effect of growth parameters on formation and elimination of vacancy clusters in dislocation-free silicon crystals, J. Cryst. Growth **22**, 311–320 (1974)

38.9 A.J.R. de Kock, P.J. Roksnoer, P.G.T. Boonen: Formation and elimination of growth striations in dislocation-free silicon crystals, J. Cryst. Growth **28**, 125–137 (1975)

38.10 P.M. Petroff, A.J.R. de Kock: Characterization of swirl defects in floating-zone silicon crystals, J. Cryst. Growth **30**, 117–124 (1975)

38.11 P.M. Petroff, A.J.R. de Kock: The formation of interstitial swirl defects in dislocation-free floating-zone silicon crystals, J. Cryst. Growth **36**, 4–12 (1976)

38.12 H. Föll, U. Gösele, B.O. Kolbesen: The formation of swirl defects in silicon by agglomeration of self-interstitials, J. Cryst. Growth **40**, 90–108 (1977)

38.13 A.J.R. de Kock, W.M. van de Wiljert: The effect of doping on the formation of swirl defects in dislocation-free Czochralski-grown silicon crystals, J. Cryst. Growth **49**, 718–734 (1980)

38.14 P.J. Roksnoer, M.M.B. van den Boom: Microdefects in a non-striated distribution in floating-zone silicon crystals, J. Cryst. Growth **53**, 563–573 (1981)

38.15 P.J. Roksnoer: The mechanism of formation of microdefects in silicon, J. Cryst. Growth **68**, 596–612 (1984)

38.16 M. Hasebe, Y. Takeoka, S. Shinoyama, S. Naito: Formation process of stacking faults with ringlike distribution in CZ-Si wafers, Jpn. J. Appl. Phys. **28**, L1999–L2002 (1989)

38.17 M. Kato, T. Yoshida, Y. Ikeda, Y. Kitagawara: Transmission electron microscope observation of "IR scattering defects" in As-grown Czochralski Si crystals, Jpn. J. Appl. Phys. **35**, 5597–5601 (1996)

38.18 T. Ueki, M. Itsumi, T. Takeda: Octahedral void defects observed in the bulk of Czochralski silicon, Appl. Phys. Lett. **70**, 1248–1250 (1997)

38.19 V.V. Voronkov: The mechanism of swirl defects formation in silicon, J. Cryst. Growth **59**, 625–643 (1982)

38.20 M. Iida, W. Kusaki, M. Tamatsuka, E. Iino, M. Kimura, S. Muraoka: Effects of light element impurities on the formation of grown-in defects free region of Czochralski silicon single crystal, Proc. Electrochem. Soc. **99-1**, 499–510 (2000)

38.21 E. Dornberger, D. Gräf, M. Suhren, U. Lambert, P. Wagner, F. Dupret, W. von Ammon: Influence of boron concentration on the oxidation-induced stacking fault ring in Czochralski silicon crystals, J. Cryst. Growth **180**, 343–352 (1997)

38.22 T. Sinno, H. Susanto, R.A. Brown, W. von Ammon, E. Dornberger: Boron-retarded self-interstitial diffusion in Czochralski growth of silicon crystals and its role in oxidation-induced stacking-fault ring dynamics, Appl. Phys. Lett. **75**, 1544–1546 (1999)

38.23 V.V. Voronkov, R. Falster: Dopant effect on point defect incorporation into growing silicon crystal, J. Appl. Phys. **87**, 4126–4129 (2000)

38.24 K. Terashima, H. Noguchi: The effects of boron impurity on the extended defects in CZ silicon crystals grown under interstitial rich conditions, J. Cryst. Growth **237-239**, 1663–1666 (2002)

38.25 M.S. Kulkarni: A selective review of the quantification of defect dynamics in growing Czochralski silicon crystals, Ind. Eng. Chem. Res. **44**, 6246–6263 (2003)

38.26 T. Sinno, R.A. Brown, W. von Ammon, E. Dornberger: Point defect dynamics and the oxidation-induced stacking-fault ring in Czochralski-grown silicon crystals, J. Electrochem. Soc. **145**, 302–318 (1998)

38.27 T. Mori: Modeling the Linkages Between Heat Transfer and Microdefect Formation in Crystal Growth: Examples of Czochralski Growth of Silicon and Vertical Bridgman Growth of Bismuth Germanate. Ph.D. Thesis (Massachusetts Institute of Technology, Massachusetts 2000)

38.28 T.Y. Tan, U. Gösele: Point defects, diffusion processes, and swirl defect formation in silicon, Appl. Phys. A **37**, 1–17 (1985)

38.29 S. Sadamitsu, S. Umeno, Y. Koike, M. Hourai, S. Sumita, T. Shigematsu: Dependence of the grown-in defect distribution on growth rates in Czochralski silicon, Jpn. J. Appl. Phys. **32**, 3675–3681 (1993)

38.30 W. von Ammon, E. Dornberger, H. Oelkrug, H. Weidner: The dependence of bulk defects on the

axial temperature gradient of silicon crystals during Czochralski growth, J. Cryst. Growth **151**, 273–277 (1995)
38.31 E. Dornberger, W. von Ammon: The dependence of ring-like distributed stacking faults on the axial temperature gradient of growing Czochralski silicon crystals, J. Electrochem. Soc. **143**, 1648–1653 (1996)
38.32 E. Dornberger, W. von Ammon, N. Van den Bogaert, F. Dupret: Transient computer simulation of a CZ crystal growth process, J. Cryst. Growth **166**, 452–457 (1996)
38.33 R. Falster, V.V. Voronkov, J.C. Holzer, S. Markgraf, S. McQuaid, L. Muléstagno: Intrinsic point defects and reactions in the growth of large silicon crystals, Proc. Electrochem. Soc. **98**(1), 468–489 (1998)
38.34 K. Nakamura, T. Saishoji, T. Kubota, T. Iida, Y. Shimanuki, T. Kotooka, J. Tomioka: Formation process of grown-in defects in Czochralski grown silicon crystals, J. Cryst. Growth **180**, 61–72 (1997)
38.35 P.A. Ramachandran, M.P. Dudukovic: Simulation of temperature distribution in crystals grown by Czochralski method, J. Cryst. Growth **71**, 399–408 (1985)
38.36 R.K. Srivastava, P.A. Ramachandran, M.P. Dudukovic: Interface shape in Czochralski grown crystals: Effect of conduction and radiation, J. Cryst. Growth **73**, 487–504 (1985)
38.37 A. Virzi: Computer modelling of heat transfer in Czochralski silicon crystal growth, J. Cryst. Growth. **112**, 699–722 (1991)
38.38 M.S. Kulkarni, J. Libbert, S. Keltner, L. Muléstagno: A theoretical and experimental analysis of macrodecoration of defects in monocrystalline silicon, J. Electrochem. Soc. **149**(2), G153–G165 (2002)
38.39 M.S. Kulkarni: A review and unifying analysis of defect decoration and surface polishing by chemical etching in silicon processing, Ind. Eng. Chem. Res. **42**, 2558–2588 (2003)
38.40 M.S. Kulkarni, H.F. Erk: Acid-based etching of silicon wafers: mass-transfer and kinetic effects, J. Electrochem. Soc. **147**, 176–188 (2000)
38.41 R.A. Brown, D. Maroudas, T. Sinno: Modelling point defect dynamics in the crystal growth of silicon, J. Cryst. Growth **137**, 12–25 (1994)
38.42 N.I. Puzanov, A.M. Eidenzon, D.N. Puzanov: Modelling microdefect distribution in dislocation-free Si crystals grown from the melt, J. Cryst. Growth **178**, 468–478 (1997)
38.43 E. Dornberger, W. von Ammon, J. Virbulis, B. Hanna, T. Sinno: Modelling of transient point defect dynamics in Czochralski silicon crystals, J. Cryst. Growth **230**, 291–299 (2001)
38.44 M. Okui, M. Nishimoto: Effect of the axial temperature gradient on the formation of grown-in defect regions in Czochralski silicon crystals; reversion of the defect regions between the inside and outside of the Ring-OSF, J. Cryst. Growth **237**, 1651–1656 (2002)
38.45 V.V. Kalaev, D.P. Lukanin, V.A. Zabelin, Y.N. Makarov, J. Virbulis, E. Dornberger, W. von Ammon: Calculation of bulk defects in CZ Si growth: impact of melt turbulent fluctuations, J. Cryst. Growth **250**, 203–208 (2003)
38.46 V.V. Voronkov, R. Falster: Vacancy-type microdefect formation in Czochralski silicon, J. Cryst. Growth **194**, 76–88 (1998)
38.47 T. Sinno, R.A. Brown: Modeling microdefect formation in Czochralski silicon, J. Electrochem. Soc. **146**, 2300–2312 (1999)
38.48 Z. Wang, R.A. Brown: Simulation of almost defect-free silicon crystal growth, J. Cryst. Growth **231**, 442–447 (2001)
38.49 R.A. Brown, Z. Wang, T. Mori: Engineering analysis of microdefect formation during silicon crystal growth, J. Cryst. Growth **225**, 97–109 (2001)
38.50 E. Dornberger, J. Virbulis, B. Hanna, R. Hölzl, E. Daub, W. von Ammon: Silicon crystals for future requirements of 300 mm wafers, J. Cryst. Growth **229**, 11–16 (2001)
38.51 E. Dornberger, D. Temmler, W. von Ammon: Defects in silicon crystals and their impact on DRAM device characteristics, J. Electrochem. Soc. **149**, G226–G231 (2002)
38.52 K. Kitamura, J. Furukawa, Y. Nakada, N. Ono, Y. Shimanuki, A.M. Eidenzon, N.I. Puzanov, D.N. Puzanov: Radial distribution of temperature gradients in growing CZ-Si crystals and its application to the prediction of microdefect distribution, J. Cryst. Growth **242**, 293–301 (2002)
38.53 M.S. Kulkarni, V.V. Voronkov: Simplified two-dimensional quantification of the microdefect distributions in silicon crystals grown by the Czochralski process, J. Electrochem. Soc. **152**, G781–G786 (2005)
38.54 M.S. Kulkarni, J.C. Holzer, L.W. Ferry: The agglomeration dynamics of self-interstitials in growing Czochralski silicon crystals, J. Cryst. Growth **284**, 353–368 (2005)
38.55 D. Turnbull, J.C. Fisher: Rate of nucleation in condensed systems, J. Chem. Phys. **17**, 71–73 (1949)
38.56 A.S. Michaels: *Nucleation Phenomena* (American Chemical Society, Washington 1966)
38.57 D. Kashchiev: *Nucleation, Basic Theory with Applications* (Butterworth-Heinemann, Oxford 2000)
38.58 V.V. Voronkov, R. Falster: Vacancy and self-interstitial concentration incorporated into growing silicon crystals, J. Appl. Phys. **86**, 5975–5982 (1999)
38.59 V.V. Voronkov, R. Falster: Intrinsic point defects and impurities in silicon crystal growth, J. Electrochem. Soc. **194**, G167–G174 (2002)
38.60 S.M. Hu: Nonequilibrium point defects and diffusion in silicon, Mater. Sci. Eng. **R13**(3/4), 105–192 (1994)

38.61 P.M. Fahey, P.B. Griffin, J.D. Plummer: Point defects and dopant diffusion in silicon, Rev. Mod. Phys. **61**, 289–384 (1989)

38.62 U. Gösele, D. Conrad, P. Werner, Q.Y. Tong, R. Gafiteanu, T.Y. Tan: Point defects, diffusion and gettering in silicon, Mater. Res. Soc. Symp. Proc. **469**, 13 (1997)

38.63 R. Falster, V.V. Voronkov, F. Quast: On the properties of the intrinsic point defects in silicon: a perspective from crystal growth and wafer processing, Phys. Status Solidi (b) **222**, 219–244 (2000)

38.64 V.V. Voronkov, R. Falster: Parameters of intrinsic point defects in silicon based on crystal growth, wafer processing, self- and metal-diffusion, ECS Trans. **2**(2), 61–75 (2006)

38.65 J. Esfandyari, C. Schmeiser, S. Senkader, G. Hobler, B. Murphy: Computer simulation of oxygen precipitation in Czochralski-grown silicon during HI-LO-HI anneals, J. Electrochem. Soc. **143**, 995–1001 (1996)

38.66 V.V. Voronkov, R. Falster: Grown-in microdefects, residual vacancies and oxygen precipitation bands in Czochralski silicon, J. Cryst. Growth **204**, 462–474 (1999)

38.67 V.V. Voronkov, R. Falster: Strain-induced transformation of amorphous spherical precipitates into platelets: application to oxide particles in silicon, J. Appl. Phys. **89**, 5965–5971 (2001)

38.68 V.V. Voronkov, R. Falster: Nucleation of oxide precipitates in vacancy-containing silicon, J. Appl. Phys. **91**, 5802–5810 (2002)

38.69 K. Sueoka, M. Akatsuka, M. Okui, H. Katahama: Computer simulation for morphology, size, and density of oxide precipitates in CZ silicon, J. Electrochem. Soc. **150**, G469–G475 (2003)

38.70 M.S. Kulkarni: Defect dynamics in the presence of oxygen in growing Czochralski silicon crystals, J. Cryst. Growth **303**, 438–448 (2007)

38.71 K. Nakai, Y. Inoue, H. Yokota, A. Ikari, J. Takahashi, K. Kitahara, Y. Ohta, W. Ohashi: Oxygen precipitation in nitrogen-doped Czochralski-grown silicon crystals, J. Appl. Phys. **89**, 4301–4309 (2001)

38.72 W. von Ammon, R. Hölzl, J. Virbulis, E. Dornberger, R. Schmolke, D. Gräf: The impact of nitrogen on the defect aggregation in silicon, J. Cryst. Growth **226**, 19–30 (2001)

38.73 A. Karoui, F.S. Karoui, G.A. Rozgonyi, M. Hourai, K. Sueoka: Structure, energetics, and thermal stability of nitrogen-vacancy-related defects in nitrogen doped silicon, J. Electrochem. Soc. **150**, G771–G777 (2003)

38.74 V.V. Voronkov, R. Falster: The effect of nitrogen on void formation in Czochralski crystals, J. Cryst. Growth **273**, 412–423 (2005)

38.75 M.S. Kulkarni: Defect dynamics in the presence of nitrogen in growing czochralski silicon crystals, J. Cryst. Growth **310**, 324–335 (2008)

38.76 Y. Yatsurugi, N. Akijama, Y. Endo, T. Nozaki: Concentration, solubility, and equilibrium distribution coefficient of nitrogen and oxygen in semiconductor silicon, J. Electrochem. Soc. **120**, 975–979 (1973)

38.77 V.V. Voronkov, M. Porrini, P. Collareta, M.G. Pretto, R. Scala, R. Falster, G.I. Voronkova, A.V. Batunina, V.N. Golovina, L.V. Arapkina, A.S. Guliaeva, M.G. Milvidski: Shallow thermal donors in nitrogen-doped silicon, J. Appl. Phys. **89**, 4289–4293 (2001)

38.78 J. Takahashi, K. Nakai, K. Kawakami, Y. Inoue, H. Yokota, A. Toshikawa, A. Ikari, W. Ohashi: Microvoid defects in nitrogen- and/or carbon doped Czochralski-grown silicon crystals, Jpn. J. Appl. Phys. **42**, 363–370 (2003)

38.79 M.S. Kulkarni: Lateral incorporation of vacancies in Czochralski silicon crystals, J. Cryst. Growth **310**, 3183–3191 (2008)

39. Models for Stress and Dislocation Generation in Melt Based Compound Crystal Growth

Vishwanath (Vish) Prasad, Srinivas Pendurti

A major issue in the growth of semiconductor crystals is the presence of line defects or dislocations. Dislocations are a major impediment to the usage of III–V and other compound semiconductor crystals in electronic, optical, and other applications. This chapter reviews the origins of dislocations in melt-based growth processes and models for stress-driven dislocation multiplication. These models are presented from the point of view of dislocations as the agents of plastic deformation required to relieve the thermal stresses generated in the crystal during melt-based growth processes. Consequently they take the form of viscoplastic constitutive equations for the deformation of the crystal taking into account the microdynamical details of dislocations such as dislocation velocities and interactions. The various aspects of these models are dealt in detail, and finally some representative numerical results are presented for the liquid encapsulated Czochralski (LEC) growth of InP crystals.

39.1 **Overview** 1335

39.2 **Crystal Growth Processes** 1336
39.2.1 Czochralski Technique 1336

39.3 **Dislocations in Semiconductors Materials** 1337
39.3.1 Deleterious Effects of Dislocations 1337
39.3.2 Origin of Dislocations 1338

39.4 **Models for Dislocation Generation** 1339
39.4.1 CRSS-Based Elastic Models 1340
39.4.2 Viscoplastic Models 1342

39.5 **Diamond Structure of the Crystal** 1343

39.6 **Deformation Behavior of Semiconductors** 1346
39.6.1 Stage of Upper and Lower Yield Points 1347

39.7 **Application of the Haasen Model to Crystal Growth** 1350

39.8 **An Alternative Model** 1351
39.8.1 Different Types of Dislocations 1351
39.8.2 Dislocation Glide Velocity 1352
39.8.3 Dislocation Multiplication 1355
39.8.4 Work Hardening 1357
39.8.5 The Initial Dislocation Density 1358

39.9 **Model Summary and Numerical Implementation** 1360
39.9.1 Summary of the Model 1360
39.9.2 Numerical Implementation 1361

39.10 **Numerical Results** 1362
39.10.1 Strength of Convection in the Melt and Gas 1362
39.10.2 Temperature Boundary Condition . 1362
39.10.3 A Sample Case 1363
39.10.4 Effect of Gas Convection and Radiation 1368
39.10.5 Melt Convection and Rotation Reynolds Numbers 1369
39.10.6 Control of Encapsulation Height ... 1371
39.10.7 The Cool-Down Period 1371
39.10.8 The [1$\bar{1}$1] Growth Axis 1372
39.10.9 Summary of the Calculations and Some Comparisons 1373

39.11 **Summary** 1374

References 1375

39.1 Overview

Semiconductors are the most important materials of our age. Currently, silicon-based technology accounts for more than 95% of the total semiconductor market. In silicon technologies, high operating speed is

achieved primarily by scaling down device dimensions. Smaller dimensions enable greater packing on a single chip, which enhances its capability and efficiency. However, reducing the size of the device has many limitations. Firstly, the parasitic capacitance of the device does not scale down linearly with the device dimension. Secondly, interconnects may pose a problem, and lithographic techniques can place a limit on the size reduction as well. Lastly the minimum size of a device is ultimately limited by the induced electric fields, which cause the breakdown of the p–n junction. An alternative is the use of III–V materials in high-speed electronics. These materials have higher electron and hole mobilities than Si; for example, at room temperature, the electron mobility in GaAs is five times larger than that in Si, and hence it has a better high-frequency response. Therefore devices using III–V materials are expected to be faster than those using Si. Apart from high-frequency response, III–V-based devices have better thermal characteristics and lower threshold voltages than other devices. Better thermal characteristics would enable the manufacture of identical components in smaller form factors. For example, a power amplifier based on InP would be smaller, more compact, and more efficient than power amplifiers based on other technologies, and its lower threshold voltage would allow the amplifier to work at a lower supply voltage.

In the area of optoelectronics and photonics, III–V materials, particularly InP, are the industry leaders. The development of low-transmission-loss quartz optical fibers, with optimum transmission characteristics in the 1.1–1.6 μm wavelength region, has led to growing demand for InP bulk material, as InP is increasingly being used as the substrate material for GaInAsP lasers, which produce light of this wavelength. The high-speed and optoelectronic characteristics of InP facilitate integrating high-speed transistors, with photoreceivers and diodes, in close proximity on a single chip. This technology would eliminate interconnects and enable the creation of communications networks operating at 40 Gb/s and beyond. At such high speeds the monolithic integration of transistors and photoreceivers is crucial as the presence of even small interconnects would seriously degrade integrated circuit (IC) performance. Thus clearly, III–Vs are the materials of the future. The present limited use of InP is not due to limitations in device technology, but rather to the lack of availability of high-quality defect-free InP substrates of large diameter. Commercially, dislocation-free Si substrates are available up to a diameter of 300 mm. GaAs substrates are available up to 150 mm, and InP up to 75 mm; however, the GaAs and InP substrates suffer from a variety of defects, the most serious of which are line defects or dislocations.

39.2 Crystal Growth Processes

The bulk of modern-day requirements of Si and III–Vs are met by techniques involving pulling single crystals from their melt – Czochralski and its variants, Bridgman and its variations, float-zone melting, etc. A brief description of the Czochralski technique is provided below.

39.2.1 Czochralski Technique

This technique was developed by *Czochralski* [39.1] and later perfected by *Teal* and *Little* [39.2]. The required equipment consists of a rotating crucible (Fig. 39.1), which contains the charge material to be crystallized surrounded by a heater capable of melting the charge and maintaining it in a molten condition. A pull rod, with a seed crystal attached to its bottom and rotating, usually in an opposite direction to that of the crucible, is mounted coaxially, and lowered until the end of the seed touches the melt. The melt temperature is carefully adjusted until a meniscus is supported by the end of the seed. Once a thermal steady state is achieved the pull rod is carefully rotated and lifted, while the melt crystallizes on the seed. The diameter of the crystal is increased from that of the seed to a desired value by careful variation of the heater power and pulling rate. The crystal and the crucible are rotated during the whole process to maintain radial homogeneity. The whole crystal growth assembly is placed in a closed water-cooled chamber, which is evacuated and filled with an inert gas (generally argon) at low pressure; this is to ensure that the system is shielded from the effect of harmful atmospheric gases and temperature fluctuations of the ambient. III–V growth presents additional problems in the form of the decomposition of the melt, since the V element is volatile and escapes from the hot melt. The problem of decomposition has been overcome through a novel modification – the LEC technique (Fig. 39.2). If a high pressure is not required to maintain

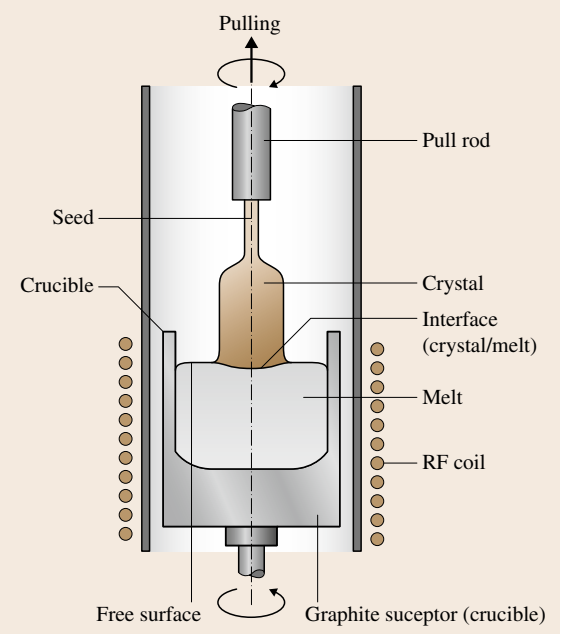


Fig. 39.1 Schematic of a low-pressure Czochralski system

the chemical equilibrium (i. e., the vapor pressure of the volatile component is not high), only minor modifications of the standard Czochralski system are required. In the case of GaAs, for example, it involves charging the crucible with a quantity of boric oxide, a low-melting-point glass, in addition to the polycrystalline gallium arsenide. On heating the crucible, the boric oxide softens and flows over the charge. Since it wets the crucible and is immiscible with and lighter than the melt, it encapsulates the charge. An inert gas is introduced into the growth chamber, at a pressure higher than the dissociation pressure of the GaAs melt, preventing loss of

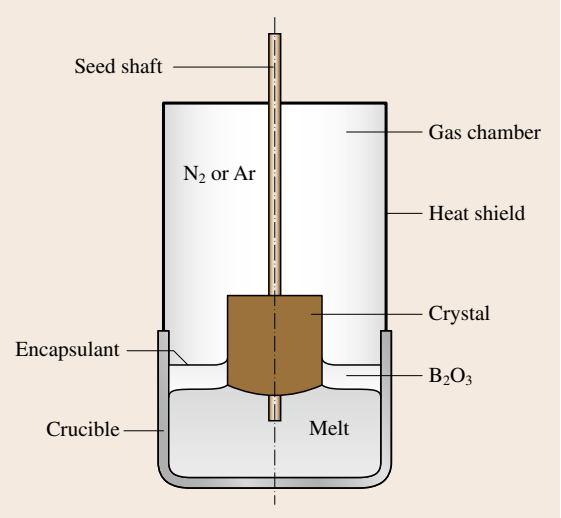


Fig. 39.2 Schematic of a high-pressure liquid-encapsulated Czochralski system

As from the melt surface. This suppresses arsenic bubble formation. Being essentially insoluble in the boric oxide, arsenic cannot diffuse through it. The crystal growing is performed in the usual manner by lowering the seed crystal through the encapsulant layer, until it contacts the melt interface; pulling is then initiated in the usual way.

The pressure of the inert gas in the growth chamber depends on the dissociation pressure of the melt and may range from about 1 atm for GaAs, through 40 atm for InP, to 70 atm for compounds such as GaP, and much higher pressures for ZnO, etc. The complexity and technological challenges associated with the LEC technique increase greatly with the pressure, which is high – 40 atm – for InP.

39.3 Dislocations in Semiconductors Materials

Dislocations are line defects in semiconductors, which adversely affect the characteristic semiconductor properties such as carrier density, mean lifetime of minority carriers, and carrier mean free path. Large single crystals of traditional elemental semiconductor materials such as Si and Ge can be grown virtually dislocation free using the necking technique pioneered by *Dash* [39.3]. However, growth of large-diameter single crystals of III–V and II–VI compound crystals is prone to the problem of dislocations regardless of the technique used. This problem is particularly severe in the growth of large crystals, and is one of the limiting factors in the growth of large-diameter high-quality GaAs and InP single crystals.

39.3.1 Deleterious Effects of Dislocations

The presence of dislocations introduces insurmountable problems in the use of such defective crystals for device applications. A high dislocation density in the crystal

introduces a large number of etch pits on the surfaces of wafers made from it. The use of wafers for device manufacture involves the etching and deposition of transistors, the dimensions of which are of the same order as that of the etch pits. If a transistor is engraved near an etch pit, its characteristic behavior is very different from that of a transistor etched in nondefective material. Hence it is imperative for device applications that wafers are uniform and largely free of defects.

Dislocations are detrimental for optoelectronic applications. Dislocations have been directly implicated in the case of degradation of GaAs-based light-emitting devices such as the GaAs–AlGaAs light-emitting diode (LED) [39.4]. The imperfect lattice near the dislocation core enhances nonradiative recombination between majority and minority carriers. Such recombinations near dislocation cores, instead of producing photons, only produce local lattice excitations and thermal energy, reducing luminescence. For example *Roedel* et al. [39.5] plotted luminescence efficiency versus dislocation density for 45 individual $Al_xGa_{1-x}As$:Si LEDs and obtained a curve directly documenting the decrease in efficiency with etch pit density on the original wafer from which the device was fabricated. The degradation is of three different broad categories, and the following are directly quoted from [39.4]

> *(1) A rapid degradation that occurs after a short duration of operation, (2) a slow degradation occurring during long term operation; and (3) a catastrophic degradation process that occurs at high optical power densities because of mirror facet damage.*

The first two modes of degradation are closely linked to nonradiative recombination near the dislocation cores. Moreover the thermal energy generated during such recombinations leads to faster dislocation rearrangements and movements involving dislocation climb and glide, and leads to the development of more dislocation structure, more defective material, and degradation of luminescence with time. Reference [39.4] provides a more detailed discussion on this topic.

39.3.2 Origin of Dislocations

In most works thermal stresses are cited as the chief reason behind dislocations. This is because the stresses drive the multiplicative mechanisms of dislocations, which begin to operate as the small number of dislocations initially present begin to glide under the influence of thermal stresses. While this is certainly true, this is a matter of dislocation multiplication from *existing dislocations* and does not explain the nucleation of the first *existing dislocations* from which others multiply. Hence one has to differentiate between *dislocation multiplication* and *dislocation nucleation*. Dislocation nucleation is not possible solely due to thermal stresses; it has been estimated that stresses of the order of $\mu/2\pi$, where μ is the theoretical shear modulus of the material, are required for the nucleation of dislocations solely under their influence. The crystal growth process does not involve thermal stresses of this magnitude. The nucleation of dislocations is generally attributed to the following mechanisms, many or all of which might be operating in tandem.

Agglomeration of Point Defects

During crystal growth, there exists a high equilibrium concentration of point defects in a given material element, when it is near the crystal–melt interface, because of the high temperature close to the melting point. During the growth process, the material element moves away from the interface, and cools down, giving rise to supersaturation of point defects. These point defects agglomerate into clusters, which in order to minimize their elastic strain energy collapse into various structures. Chapter 38 in this volume details the dynamics of the agglomeration and the defects formed therein, under various conditions, for Czochralski growth of Si. One of the minimum-energy structures into which the agglomerations of point defects can collapse is a planar disc in a (111) plane. Such a disc is equivalent to a stacking fault bound by a Frank partial dislocation loop. The disc grows by the climb of the bounding partial dislocation, and once it reaches a critical size, a Shockley partial may be spontaneously nucleated, removing the stacking fault and transforming the Frank partial into a loop of a perfect dislocation [39.6, Chap. 10]. For example, a Frank dislocation loop in the (111) plane dissociates as

$$\tfrac{a}{3}[111] + \tfrac{a}{6}[\bar{2}11] \rightarrow \tfrac{a}{2}[011]\,.$$

The steps of point defect agglomeration – collapse into discs and Shockley nucleation – are governed by kinetics determined by the thermal history and the material. The perfect dislocation loop formed by the above process is not a glide dislocation, since its Burgers vector does not lie in the plane of the dislocation loop. However, a fraction of these loops become mobile glide dislocation loops by the escape of a segment onto a glide plane, which contains the Burgers vec-

tor. Refer to [39.7] for a more detailed explanation. In compound semiconductors, the situation is complicated by the sublattices of the diamond structure being occupied by atoms of different elements. In these compounds, the formation of a vacancy (interstitial) disc involves equal numbers of vacancies (interstitials) of both sublattices, which is difficult, since the concentration of point defects on the two sublattices is generally different, depending on factors such as stoichiometry, etc. *Lee* et al. [39.8] have identified mechanisms by which the above problem may be overcome, to nucleate glide dislocations. *Lagowski* et al. [39.9] have experimentally revealed the effect of doping on the nucleation of glide dislocations, through aggregation of point defects.

Surface Damage

Damage on the surface of the growing crystal due to indentation, scratching, grinding, and impingement of hard particles leads to the nucleation of dislocations. All of these processes involve the action of large local stresses causing the relative displacement of two layers of the crystal: the nucleation of a dislocation. This mode of nucleation might not be important in crystal growth, but it has been the experience of crystal growers that small particles from the melt on the surface of the growing crystal produce such dislocations [39.10].

Foreign Particles or Precipitates

The presence of a second phase in the form of foreign particles, precipitates, etc., leads to misfit due to differential thermal expansion of the particle and the surrounding matrix. This gives rise to very high local stresses and consequently misfit dislocations. Usually, these misfit dislocations are in the form of glide dislocations in the glide plane of maximum shear stress, as discussed by *Ashby* and *Johnson* [39.11] and *Alexander* [39.7]. Ashby and Johnson show that the radius of the original nucleated loop should be greater than $R_c = \mu b/\tau$, where τ is the effective shear stress and b is the Burgers vector, so that the loop can continue to grow and expand under the influence of the shear stress. In III–V compounds, inclusions from the B_2O_3 encapsulant or precipitates of one phase due to lack of stoichiometry might act as foreign particles, thus nucleating glide dislocations.

Chemical Inhomogeneity

Chemical inhomogeneity may arise due to doping or impurity atoms and nonstoichiometry. These might lead to lattice mismatch, giving rise to misfit dislocations. Sometimes, the phenomenon of constitutional supercooling might arise in the melt near the solidification interface, giving rise to instabilities in the interface. In such cases, the planar interface splits into cells. Impingement of different parts of the interface (cells) might give rise to misfit dislocations. However, constitutional supercooling is expected to be rare in semiconductor single-crystal growth, due to the low concentration of impurity/dopant.

Use of a Defective Seed

If the seed crystal contains dislocations, these are likely to glide into the newly grown crystal. These seed dislocations would then multiply under the effect of thermal stresses. The use of a defective seed can be overcome by the use of a narrow necking (*Dash* process) [39.3].

Thermal Stresses

The most important cause for the presence of dislocations in as-grown semiconductor crystals is thermal stresses. Since the presence of temperature gradients in the crystal is necessary, crystal growth is an inherently nonequilibrium process, involving heat transfer through the crystal. The crystal acts as a medium through which heat, both released heat of fusion and that from the melt, is transferred to the ambient. The heat enters the crystal at the crystal–melt interface, and leaves through the external surfaces by radiation and convection. Thus the core of the crystal is hotter than the periphery, and the bottom of the growing crystal is hotter than the top. These temperature gradients give rise to differential expansion of different parts of the crystal, and consequently to thermal strains and stresses. These thermal stresses are the cause of plastic deformation, through the movement and multiplication of existing dislocations, resulting in unacceptable final dislocation densities in as-grown crystals.

39.4 Models for Dislocation Generation

Before we dwell on the mathematical models, a word on notation: vectors are represented as bold-face italic and second-order tensors as bold-face upright; indicial notation for vectors and tensors and the Einstein sum-

mation convention are not used, except in Sect. 39.2 or unless specifically mentioned.

Since thermal stress in the growing crystal is considered as the primary reason for the presence of dislocations in as-grown crystals, models for dislocation generation take a solid mechanics approach. Towards this, efforts are made to characterize the mechanistic response of the crystal to loading through appropriate constitutive equations. Depending on the type of constitutive behavior, the modeling approach falls into two broad categories: the crystal is considered as (1) an elastic crystalline medium and (2) a viscoplastic crystalline medium.

39.4.1 CRSS-Based Elastic Models

In this approach the crystal is assumed to be a linear elastic continuum. The crystal usually being a cubic material, three elastic coefficients are prescribed to characterize its elastic behavior. Isotropic behavior is also assumed sometimes. The thermal loading during the growth is specified through a coefficient of thermal expansion and the consequent thermal strain. With the help of minimal boundary conditions designed to prevent rigid-body motion, the elastic problem can be solved at any given thermal loading (which can be obtained from the temperature field) during the growth of the crystal. Solution of the elastic problem yields information on the displacements, the total, thermal, and elastic strains, and the elastic stresses within the crystal. Once the elastic stresses are obtained, these are used to estimate the plastic deformation and whether a particular area of the crystal will have a high or low dislocation density.

Dislocations in III–V materials with zincblende structure glide in the family of four {111} planes along the three ⟨110⟩ directions. Hence plastic deformation takes place along the 12 $\{111\}-\langle 110\rangle$ slip systems. *Jordan* et al. [39.12] postulated that plastic deformation starts if the resolved elastic shear stress on any of these 12 systems exceeds a critical value, which they termed the critical resolved shear stress (CRSS). Thus this model is similar in spirit to the rate-independent nonhardening plasticity models, which have a fixed yield surface in stress space. The equations and the methodology applied by *Jordan* et al. [39.12] are summarized next.

The total strain (the symmetrical part of the gradient of the displacements) at any given instant in the growth history is split into an elastic component, which is a function of the stress, and a thermal component, which is function of the temperature as follows

$$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}(\boldsymbol{\sigma}, T) = \tfrac{1}{2}\left(\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^{\mathrm{T}}\right) , \tag{39.1}$$

and

$$\boldsymbol{\varepsilon}(\boldsymbol{\sigma}, T) = \boldsymbol{\varepsilon}^{\mathrm{el}}(\boldsymbol{\sigma}) + \boldsymbol{\varepsilon}^{\mathrm{th}}(T) . \tag{39.2}$$

The elastic strain using the regular equations of linear elasticity is related to the stresses as

$$\boldsymbol{\sigma} = \mathbf{C} : \boldsymbol{\varepsilon}^{\mathrm{el}} , \tag{39.3}$$

where $\mathbf{C}$ is the fourth-order elastic modulus tensor. Similarly the thermal strain is expressed as

$$\boldsymbol{\varepsilon}^{\mathrm{th}} = \alpha(T - T_0)\boldsymbol{\delta} , \tag{39.4}$$

where α is the coefficient of thermal expansion, T_0 is some reference temperature, and $\boldsymbol{\delta}$ represents the Kronecker delta tensor. The principle that completes the equations is the mechanical equilibrium in the crystal, which assuming negligible effects of body forces such as gravity is expressed as

$$\nabla \cdot \boldsymbol{\sigma} = 0 . \tag{39.5}$$

For a given temperature field and assuming boundary conditions just sufficient to prevent rigid-body motion, (39.1–39.5) represent the familiar linear elastic problem and can be solved using any numerical technique to obtain the total strains, elastic strains, and stresses in the crystal.

The thermal–elastic stresses thus calculated are then resolved along the $\{111\}\langle 110\rangle$ systems. These slip systems are summarized in Table 39.1, and are defined in

Table 39.1 Summary of the slip systems in elemental and III–V semiconductors

Slip system number (k)	$\sqrt{2}s_{\mathrm{k}}$	$\sqrt{3}n_{\mathrm{k}}$
1	[101]	$[11\bar{1}]$
2	[011]	$[11\bar{1}]$
3	$[1\bar{1}0]$	$[11\bar{1}]$
4	[110]	$[1\bar{1}1]$
5	[011]	$[1\bar{1}1]$
6	$[10\bar{1}]$	$[1\bar{1}1]$
7	[101]	$[1\bar{1}\bar{1}]$
8	[110]	$[1\bar{1}\bar{1}]$
9	$[01\bar{1}]$	$[1\bar{1}\bar{1}]$
10	$[10\bar{1}]$	[111]
11	$[0\bar{1}1]$	[111]
12	$[1\bar{1}0]$	[111]

terms of the slip plane normal vector $\boldsymbol{n}$ and the slip direction vector $\boldsymbol{m}$, thus

$$\tau^i = \boldsymbol{m}_i \cdot \boldsymbol{\sigma} \cdot \boldsymbol{n}_i \ ,$$

where τ^i is the resolved shear stress on slip system i.

If the resolved shear stress along a slip system exceeds the critical resolved shear stress (CRSS), then that system is deemed active, contributing to the plastic deformation and multiplication of dislocation density. Towards this the *excess shear stress* on a given slip system can be calculated from

$$\tau^i_{\text{excess}} = \left|\tau^i\right| - \tau_{\text{CRSS}} \ . \tag{39.6}$$

Jordan et al. [39.12] postulated that the total dislocation density in a crystal is proportional to the sum of the excess shear stress on the first five slip systems, when the excess shear stresses are arranged in descending order. This follows from the fact that, although the plastic strain tensor has nine components, its symmetry and the incompressibility of the plastic strain mean that only five of the components are independent. Hence slip on five slip systems is sufficient to describe the plastic strain. In regions where this excess shear stress is zero, the dislocation density will be zero. Thus the total dislocation density N at a given location in the crystal is described as

$$N \propto \sum_{i=1}^{5} \tau^i_{\text{excess}} \ . \tag{39.7}$$

Jordan et al. [39.12] presented the first detailed stress calculations after calculating the thermal field from a quasisteady heat transfer model. This analysis provided the first fundamental description of dislocation generation in III–V compounds. Jordan et al. in their heat transfer analysis made a few fundamental assumptions for the calculation of the temperature field, the most important of which are as follows:

1. The crystal is modeled as a cylindrical boule, which is pulled at a constant rate from the melt held at a constant temperature T_f.
2. The solid–liquid interface is planar and is at the temperature T_f.
3. The temperature of the ambient around the boule is a constant $T_\text{a} < T_\text{f}$.
4. Heat loss from the lateral and top surfaces by radiation and natural convection follows Newton's law of cooling.

An analytical solution for the temperature field in the crystal was obtained by solving the conduction equation in the crystal, along with the above assumptions. The plane-strain solution of classical thermoelasticity was then adopted to determine the stresses and calculate the dislocation densities. *Jordan* et al. [39.12] compared their calculated dislocation densities with a computerized scan of the etch pit density of a KOH-etched (100) GaAs wafer; the main feature of the dislocation distribution such as fourfold symmetry in the etch pit density, observed experimentally, was replicated by their calculations. Similar calculations were also performed for a ⟨111⟩ LEC-pulled InP wafer. In this case the sixfold defect distributions determined experimentally somewhat resembled the dislocation contours predicted by the model. These calculations firmly established the role of thermal stresses in dislocation generation and multiplication in crystals grown from the melt.

Following the pioneering work of *Jordan* et al. [39.12], *Kobayashi* and *Iwaki* [39.13] performed similar calculations without neglecting the axial displacements; their calculations are slightly closer to experimental data than the calculations of Jordan and coworkers. *Duseaux* [39.14] repeated the calculations with a more realistic geometry by including seed and shoulder regions of the growing crystal. These early calculations along with those by *Lambropoulos* [39.15], *Schvezov* et al. [39.16], and *Meduoye* et al. [39.17, 18] assumed a known geometry of the crystal, usually with a planar melt–crystal interface, and the thermal calculations were performed with an assumed convection coefficient for heat transfer from the melt to the crystal and from the crystal to the gas. Calculations by *Motakef* et al. [39.19], *Dupret* et al. [39.20], and *Bornside* [39.21], while still neglecting melt convection, used global radiation and conduction models which included both the melt and the crystal. The interface shape in these calculations was not assumed, but was obtained from the calculations. *Zou* et al. [39.22] coupled a finite element technique, performing an isotropic linear thermal–elastic analysis on the growing crystal, with a finite volume technique for the thermal and flow analysis of the growth process. This work incorporated a number of features, including convection in the melt and the gas, radiation, etc. It should be noted that most of these works treat the crystal as linear elastic and, after calculating the stresses, performed some sort of excess shear stress calculations, following Jordan et al. Thus, there is a large amount of literature devoted to calculating dislocation densities through a thermal–elastic model, using the approach of Jordan and coworkers.

This approach, though original and successful, has several shortcomings. For example, it treats the crystal as an elastic/perfectly plastic continuum, which is untenable for elemental and compound semiconductors. The deformation of semiconductors is a highly rate-dependent process; the response of such a material to stress is not instantaneous, but depends on the kinetics of the dislocation movement and multiplication, which are strong, time-dependent functions of temperature. The so-called CRSS proposed by Jordan is not a material property and is a function of the local dislocation configuration, temperature, doping, and other factors. The dislocation configuration and temperature are functions of growth history, and change within a given material element during its growth history. All the above analyses take one temperature field as an input; it is not very clear which temperature at which instant of the growth history is the appropriate one. It is however true that a CRSS such as locking stress does exist in certain materials such as GaAs [39.23] and Si due to the presence of impurities, which lock the dislocations. A certain minimum shear stress is then required to *unlock* these dislocations, but once the unlocking takes place the material again behaves in a rate-dependent manner.

39.4.2 Viscoplastic Models

The alternative approach to modeling dislocation density evolution is the use of viscoplastic constitutive equations. Once the thermal history of the crystal from seed to as-grown state is obtained, the procedure is to integrate these viscoplastic constitutive equations in time, while maintaining equilibration of stresses at all times to obtain the final dislocation density. The advantage of this approach is that the final dislocation densities and state of the as-grown crystal is a functional of the initial state of the seed, and the entire temperature history during the growth and the cool-down period of the crystal after growth. This approach also has the potential to give an estimate of the residual stresses in the cooled crystal.

While specific constitutive equations to be used will be discussed in the following sections, a brief description of the basic viscoplasticity formulation, described in detail in *Lubliner* [39.24], with some adaptations to present requirements is given here. As is customary in standard plasticity formulations, the internal structure, or the state of the crystal, at any point, is hypothesized as being represented by a set of internal variables denoted by ξ_α or the vector $\boldsymbol{\xi}$. The mechanical strains at a given position are completely determined by the troika of internal variables, stress components, and temperature at a given position.

$$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}(\boldsymbol{\sigma}, \boldsymbol{\xi}, T) \,. \tag{39.8}$$

Additionally, for small-strain viscoplastic constitutive formulations, the following tenets are usually adapted.

Additive Decomposition of Strain Tensor

One assumes that the strain tensor can be decomposed into an elastic, plastic, and thermal part, according to the relation

$$\boldsymbol{\varepsilon}(\boldsymbol{\sigma}, T, \boldsymbol{\xi}) = \boldsymbol{\varepsilon}^{\mathrm{el}}(\boldsymbol{\sigma}) + \boldsymbol{\varepsilon}^{\mathrm{pl}}(\boldsymbol{\xi}) + \boldsymbol{\varepsilon}^{\mathrm{th}}(T) \,, \tag{39.9}$$

where the elastic strain components are assumed to be a function of the stress tensor only, the plastic strain components are assumed to be a function of the internal variables only, and finally the thermal strain component is a function of temperature only.

(Elastic) Stress Response

The stress tensor $\boldsymbol{\sigma}_{ij}$ is related to the elastic strain $\varepsilon_{ij}^{\mathrm{el}}$ through the existence of a stored energy functional W, according to the hyperelastic relationship

$$\boldsymbol{\sigma} = \frac{\partial W(\boldsymbol{\varepsilon}^{\mathrm{el}})}{\partial \boldsymbol{\varepsilon}^{\mathrm{el}}} \,.$$

For linear elasticity, the stored energy functional W assumes a quadratic form in the elastic strain, i. e., $W = \frac{1}{2}\boldsymbol{\varepsilon}^{\mathrm{el}} : \mathbf{C} : \boldsymbol{\varepsilon}^{\mathrm{el}}$, where $\mathbf{C}$ is the fourth-order tensor of elastic moduli, which is assumed constant. Then the relations between the stress components and the elastic strain tensor reduce to

$$\boldsymbol{\sigma} = \mathbf{C} : [\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^{\mathrm{pl}} - \boldsymbol{\varepsilon}^{\mathrm{th}}] \,. \tag{39.10}$$

Evolution of Internal Variables and Plastic Strain

Evolution equations are proposed for the internal variables and plastic strain tensor over time. These evolution equations for $\{\boldsymbol{\varepsilon}^{\mathrm{pl}}, \boldsymbol{\xi}\}$ are termed the flow and hardening rules, respectively. The relations are of the form

$$\dot{\boldsymbol{\varepsilon}}^{\mathrm{pl}} = \mathbf{r}(\boldsymbol{\sigma}, \boldsymbol{\xi}, T) \,, \tag{39.11a}$$

$$\dot{\boldsymbol{\xi}} = \boldsymbol{h}(\boldsymbol{\sigma}, \boldsymbol{\xi}, T) \,. \tag{39.11b}$$

The relation (39.11b) is more basic and (39.11a) can be derived from it, because of the functional dependence of plastic strain solely on the internal variables according to (39.9).

Isotropic Thermal Strain Response

The thermal strain response is assumed to be isotropic and of the form

$$\boldsymbol{\varepsilon}^{\mathrm{th}} = \alpha(T - T_0)\boldsymbol{\delta}\,, \tag{39.12}$$

where α is the coefficient of thermal expansion, T_0 is some reference temperature, and $\boldsymbol{\delta}$ is the Kronecker delta tensor.

Equations (39.9–39.12) form the outline of an internal-variable-based viscoplastic constitutive model, set in a small deformation framework, and assuming linear elastic response coupled with isotropic thermal expansion. If the internal variables adopted are some measures of dislocation densities, and the evolution (39.11) is based on dislocation dynamics, the outline presented above has the potential to predict the final defect densities as integrated functions of the growth history. The procedure is essentially one of integrating (39.11) through the growth history, while ensuring the equilibration of stresses and fulfillment of boundary conditions on the surface of the crystal. Specific instances of constitutive relations to model the deformation behavior of semiconductor materials are presented in subsequent sections of this chapter. Prior to these, a brief description of the structure of the diamond crystal will be provided, as this has a great bearing on the deformation characteristics of semiconductor materials, at least, those that crystallize in this structure.

39.5 Diamond Structure of the Crystal

The peculiarities in the plasticity of elemental and III–V semiconductor materials arise from their structures. These materials crystallize in the diamond cubic structure or sphalerite structure for compound III–V semiconductors. These structures are illustrated in Figs. 39.3 and 39.4 (adapted from [39.25]), which show three-dimensional views of the diamond and sphalerite structure, respectively. It can be seen that each atom in these structures is tetrahedrally coordinated, i. e., each atom has exactly four neighbors, and the bonds between these neighbors are in the ⟨111⟩ directions. A projection of the three-dimensional structure on a $(1\bar{1}0)$ plane is shown in Fig. 39.5. In this view the (111) atomic planes are stacked in the sequence *AaBbCcAaBbCc*... It can be seen from a straightforward comparison of Figs. 39.3 and 39.5 that the double and single lines in Fig. 39.5 mean one or two interatomic bonds. There are two different interatomic plane distances between the (111) planes – from the tetrahedral geometry, it can be calculated that the perpendicular distance between the *a* and the *A* planes is thrice the distance between the *A* and the *b* planes. The diamond lattice can also be described as just two interpenetrating face-centered cubic (fcc) lattices, or alternatively as an fcc structure with a basis of two atoms per lattice point. If the stacking sequence in the first fcc lattice is *ABCABCABC*..., then the stacking sequence in the second interpenetrating lattice is *abcabcabc*... Since the structure is just a doubled-up fcc lattice, the glide planes as in the fcc case, belong to the {111} family, and the perfect dislocations have Burgers vector $\frac{1}{2}\langle 110\rangle$. Due to the strong directional covalent bonding, the Peierls barrier in these materials is very high, and the resulting deep troughs in the Peierls barrier cause the dislocations to align primarily along the ⟨110⟩ directions, when the dislocation density is low. Thus the dislocations are primarily screw or 60° dislocations.

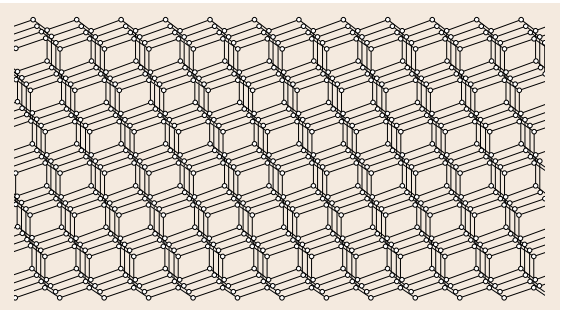


Fig. 39.3 The diamond lattice (after [39.25])

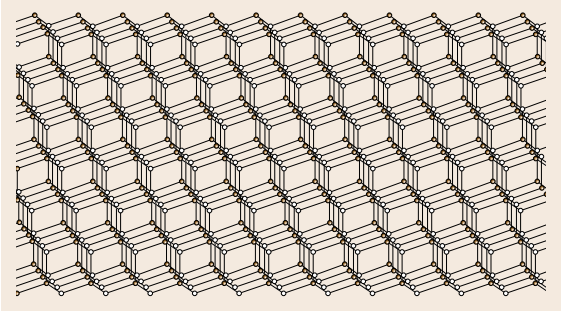


Fig. 39.4 The sphalerite lattice. *Dark* and *light* atoms show different species of atoms (after [39.25])

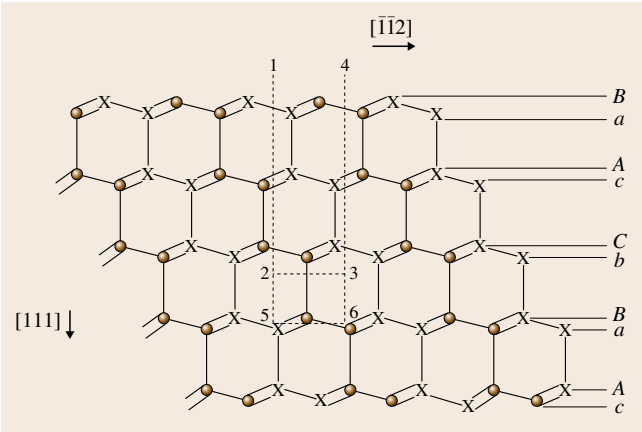


Fig. 39.5 The diamond lattice projected onto the $(1\bar{1}0)$ plane. *Bullets* represent atoms in the plane of the paper; Xs represents atoms below the plane of the paper (after [39.6])

Owing to the double-layered atomic arrangement of the diamond cubic lattice, there are two inherently different sets of dislocations, with one member of each having the same Burgers vector. These two different sets are termed the glide set and the shuffle set of dislocations. The difference is clarified by an examination of Fig. 39.5. A 60° dislocation of the glide set is formed by cutting out the material bounded by the surface 1–5–6–4, and then displacing the sides of the cut, until they are joined. The feature of these dislocations is that there is a dangling bond along surface 5–6, while all the other cut bonds on surfaces 1–5, 6–4 are made up after displacement. Thus the extra plane of atoms terminates between two layers of the different letter index, for example a and B in Fig. 39.5. Similarly, a 60° dislocation of the shuffle set is formed by cutting material along the surface 1–2–3–4 and then rejoining as before. In this case the dangling bond is along the surface 2–3, while all the bonds along surfaces 1–2 and 3–4 are filled after reconstruction. In this case the extra plane of atoms terminates between layers of the same index, such as B and b in Fig. 39.5. Three-dimensional views of the 60° glide ($\boldsymbol{BC}$) and shuffle dislocations ($\boldsymbol{B'C'}$) are given in Figs. 39.6 and 39.7, respectively. It can be seen that the shuffle and glide dislocations can transform into each other any time by the emission or absorption of point defects. The same distinction between glide and shuffle dislocations applies to dislocations of screw, edge, and intermediate orientations. To describe the dislocations the Thompson tetrahedron notation is used. Since there are two fcc lattices, there are two Thompson tetrahedrons used, one to describe the glide set (for example, dislocation $\boldsymbol{BC}$ above) and one to describe the shuffle set (for example, dislocation $\boldsymbol{B'C'}$ above).

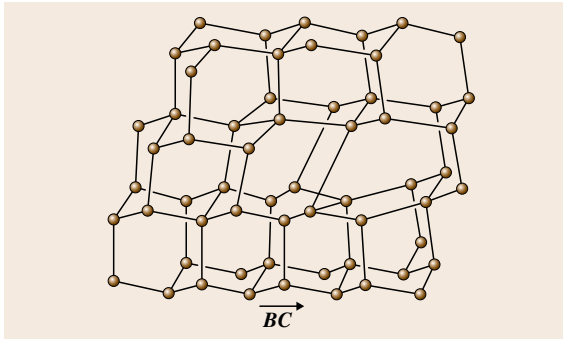


Fig. 39.6 The 60° glide dislocation (after [39.6])

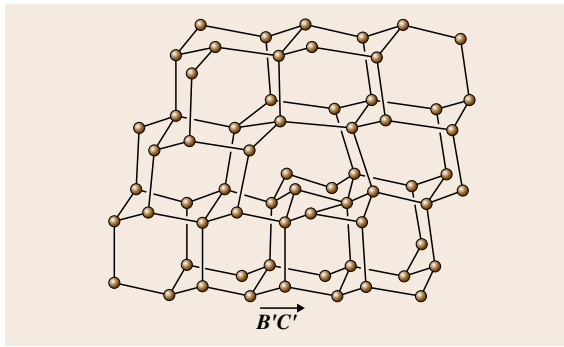


Fig. 39.7 The 60° shuffle dislocation (after [39.6])

Dislocations in the glide set can undergo dissociation into partial dislocations, in the same manner as dislocations in fcc metals. Figure 39.8 is a three-dimensional view of the glide dislocation $\boldsymbol{BC}$ dissociated into $\delta\boldsymbol{C}$ (30° partial) and $\boldsymbol{B}\delta$ (90° partial). These partial dislocations in the glide set are glissile, and because the dislocation glides between the layers b and C, the dissociation into $\delta\boldsymbol{C}$ and $\boldsymbol{B}\delta$ results in an intrinsic stacking fault between the two partials. The dissociation of shuffle dislocations in the set $A'B'C'D'$ is not direct, because such dissociation would produce a high-energy fault of the type $CcAaB|aBbCc$.

Hornstra [39.26], and later *Alexander* [39.27], gave an alternate description of an extended shuffle dislocation, which may be described in two completely equivalent ways: either as an association of a stacking-fault ribbon to a shuffle dislocation, or as a dissociated glide dislocation that has emitted or absorbed a line of atoms in the core of one of its partials. This is consistent with the relationship between perfect shuffle and

glide dislocations that can be converted into each other by emission or absorption of point defects.

Dislocations of the screw, edge, and other intermediate types, of both the shuffle and glide variety, may be generated in ways equivalent to that of the 60° dislocation. For these configurations, dislocations of the glide and shuffle type may be dissociated into partial dislocations in equivalent ways. For example, a screw dislocation of the glide type may be dissociated into two 30° partial dislocations, with a stacking fault between them. These can be converted into the shuffle type by emission or absorption of point defects. A pertinent question in the literature is whether the dislocations are extended and, if so, whether they exist in the shuffle or the glide set. It was previously assumed that moving dislocations were undissociated and existed in the shuffle mode, since the number of covalent bonds to be broken is three times smaller than in the glide mode. However, it has been demonstrated that dislocations are dissociated when at rest in both the diamond and sphalerite structure [39.28,29], and remain dissociated while moving [39.30].

Since the movement of partial dislocations in the shuffle set is more difficult than those in the glide set, it is more probable that the glide set is prevalent. However, it must be pointed out that the coexistence of both sets is not ruled out in the literature, and sometimes the as yet unexplained generation of point defects during deformation of semiconductors is attributed to the existence of, and interchange between, both sets [39.31].

The situation is further complicated in the case of compound semiconductors such as InP, GaAs, InSb, and other III–V and II–VI compounds, which exist in the sphalerite or the zincblende structure (Fig. 39.4). The sphalerite structure is the same as the diamond structure, the only difference being that the two atoms at each lattice point are of two different elements in the sphalerite structure, while they are of the same element in the diamond structure. Identical to the diamond structure, the glide planes are {111}, the direction of closest packing, while the slip directions, or Burgers vector of the perfect dislocations, are again ⟨110⟩, the shortest translation vectors in the lattice. In the following paragraphs the chief differences between the two structures are illustrated.

α and β Dislocations

Apart from the complexities of the glide and shuffle sets already existing in the diamond cubic structure, there are further complexities in the sphalerite system, i. e., there are two different {111} systems. For example, in Fig. 39.5, the A, B, and C layers contain one atom, say In for InP, while the layers a, b, and c contain the P atoms. Opposite-sign dislocations in the sphalerite structure have the extra plane of atoms ending in different atoms. For example, in a 60° shuffle dislocation (Fig. 39.5), formed by removing material bounded by surface 1–2–3–4, the extra plane of atoms lies below the surface 2–3 and ends on the B layer in the diagram, which is an In layer. If instead a shuffle dislocation of exactly the opposite sign was formed at exactly the same place, the extra plane of atoms would lie above the surface 2–3, and end on the b layer, which is a P layer. Dislocations with the edge of their extra half-planes ending on In (atoms of lower valency) layers are called β dislocations, while those with the edge of their extra half-planes ending on P (atoms of higher valency) layers are called α dislocations. This terminology, although illustrated for the shuffle set, can also be applied to glide dislocations. Since, the core structures of α and β dislocations are different, they have different properties, the most important of which is the significant difference of mobility between the two.

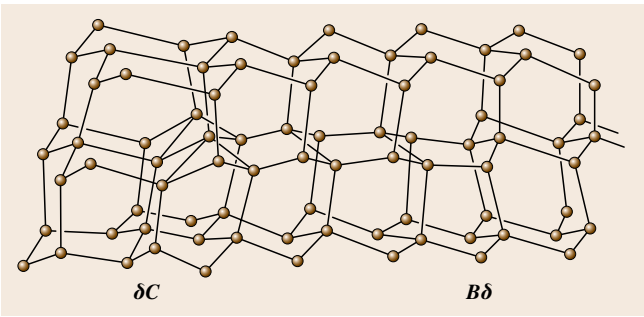


Fig. 39.8 Dissociation of a prefect 60° glide dislocation into 30° and 90° partials (after [39.6])

As pointed out before, the dislocations in compound or elemental semiconductors lie along the ⟨110⟩ directions of the glide plane, at low dislocation densities. A perfect dislocation loop would thus consist of two screw dislocations and four 60° dislocations. The geometry of a typical loop is shown in Fig. 39.9a, which shows that a glide loop possesses two types of dislocation segments: screw and 60° dislocations (60° being the angle between the dislocation line and the Burgers vector). The 60° dislocations are of two different types, α and β, since the extra half-planes of these are respectively down and up. The screw dislocations are not classified since they do not have an extra half-plane. Figure 39.9b–d show the dissociation of the α, β, and screw dislocations into partial dislocations in the glide

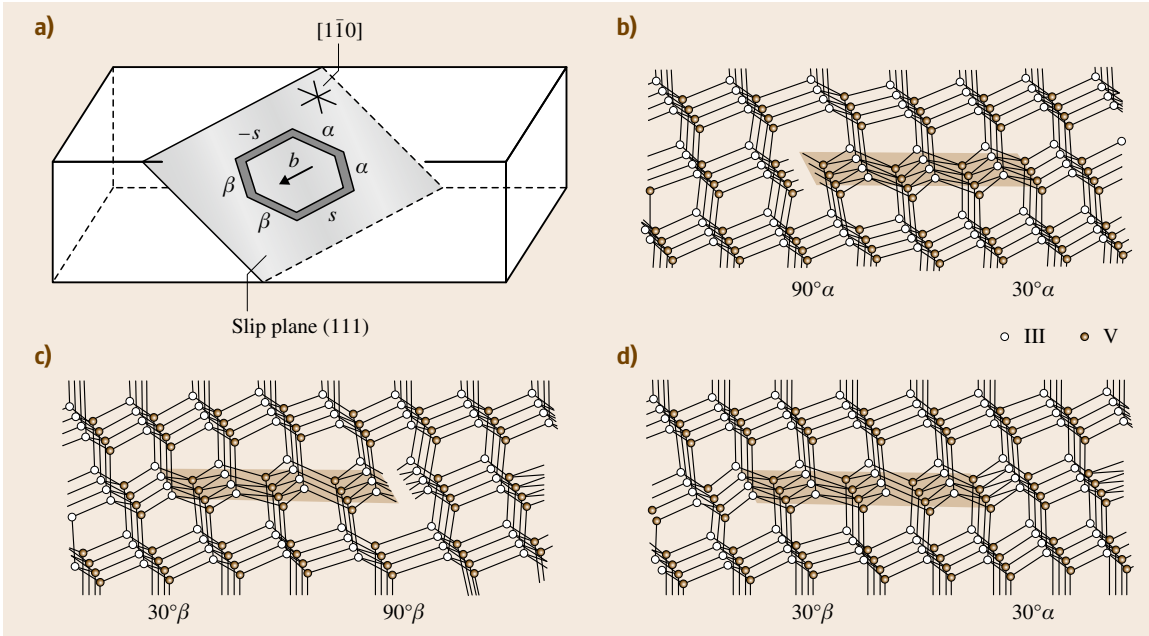


Fig. 39.9 **(a)** A hexagonal dislocation loop on a glide plane, **(b)** core structure of a 60° α dislocation, **(c)** 60° β dislocation, and **(d)** screw dislocation. The *shaded planes* represent the stacking faults (after [39.25])

set, with 60° dislocations split into 30 and 90° partials, and each screw split into a 30° partial dipole. Each of the partials can be associated with α or β character as shown.

39.6 Deformation Behavior of Semiconductors

Elemental and compound semiconductors are brittle at low temperatures, but acquire a substantial capacity for plastic deformation at temperatures above 30% of their melting point. The brittleness is a result of strong directional covalent bonds, which result in a deep Peierls potential. The plasticity of these materials is strongly strain rate and temperature dependent, again owing to the strong Peierls potential acting against dislocation movement. The dislocations glide in these materials at low speeds, in contrast to fcc metals whose plasticity is almost rate independent. The deformation behavior of these materials has been studied extensively through the stress–strain behavior of elemental and compound semiconductors such as Si, Ge, and InSb during uniaxial compression or tensile testing conducted at constant strain rate on a crystal oriented for single slip, and creep tests at constant stress for samples again oriented for single slip. The most common way of obtaining these curves is the constant-strain-rate compression test. This is done in a constant-displacement-rate machine, in which the specimen is kept between two rigid crossheads while one of them moves at a fixed speed. The axial force applied on the crosshead to compress the specimen is recorded at any instant. The force is recorded as a function of the crosshead displacement, which can then be converted into curves of shear stress versus shear strain based on the specimen dimensions. All elemental and compound semiconductors exhibit common behavior in uniaxial compression constant-strain-rate tests, oriented for single-slip conditions, when the tests are done at temperatures where the mobilities of the dislocations are similar for all materials. This commonality of behavior indicates that the underlying microscopic processes behind the deformation have a common origin. Figure 39.10 elucidates the generic features common to all elemental and III–V compound semiconductors. In this figure, the shear stress in the main slip system is plotted against the shear strain in the same system, as the deformation evolves. The striking feature of this curve is the presence of

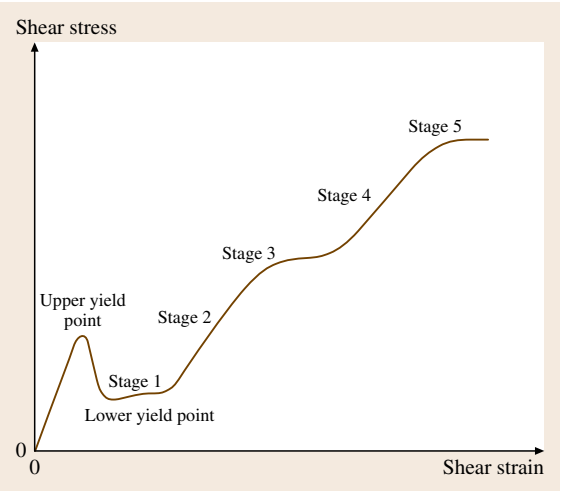


Fig. 39.10 Generic plot of shear stress versus shear strain for deformation of semiconductors in single slip

an upper and a lower yield point. The stress increases rapidly with strain at the beginning of the deformation, reaches a maximum, then decreases with increasing strain and reaches a minimum, from where the stress increases with increasing stain. The stress maximum and minimum are termed the upper and lower yield point, respectively. At low temperatures, three deformation stages following the upper and lower yield points are observed: a stage of low work hardening (I), a stage of strong work hardening (II), and a stage of increasing softening (III), where recovery processes operate in the crystals exposed to an external stress – dynamical recovery. At high temperatures, two further stages appear: a new hardening stage (IV) and a second recovery stage (V).

39.6.1 Stage of Upper and Lower Yield Points

The phenomena of upper and lower yield points was also observed in LiF. *Johnston* and *Gilman* [39.32] explained this phenomenon in terms of microscopic dislocation dynamics. In a constant-strain-rate uniaxial deformation test, the total strain rate can be split into its elastic and plastic components as

$$\dot{\varepsilon} = \dot{\varepsilon}_{\mathrm{pl}} + \dot{\varepsilon}_{\mathrm{el}} \,.$$

At the beginning of the deformation, the plastic strain rate is low, due to the low initial dislocation density in semiconductors. The total strain rate is then mostly elastic in origin, leading to a high initial rise in stress. This rising stress multiplies the dislocation density many-fold, leading to the plastic strain rate catching up and finally overtaking the total strain rate. As a result the stress actually falls, leading to a maximum: an upper yield point. The stress continues falling until dislocation interaction becomes important, leading to the hardening of the material, after a stress minimum: the lower yield point. *Alexander* and *Haasen* [39.33] reformulated the theory by Johnston, for semiconductors, to explain the yield phenomenon.

The central tenet of their theory is the introduction of a variable N to denote the dislocation density in the crystal. The plastic shear strain rate on the only active slip system is given by the Orowan relation

$$\frac{\mathrm{d}\gamma_{\mathrm{pl}}}{\mathrm{d}t} = bNv \,, \tag{39.13}$$

where b is the Burgers vector of dislocations, v is the mean velocity of the dislocations, and γ_{pl} is the plastic shear strain on the only active slip system. Since there is only one active slip plane in the crystal, the axial component of the macroscopic plastic strain tensor, $\varepsilon_{\mathrm{pl}}$, is given by $\varepsilon_{\mathrm{pl}} = \phi\gamma_{\mathrm{pl}}$, where ϕ is a geometrical factor relating the shear and the axial strains.

The velocity of dislocations in the slip system v is dependent on the resolved shear stress on that system and has been found experimentally [39.34] to be of the form

$$v = v_0 \left(\frac{\tau_{\mathrm{eff}}}{\tau_0}\right)^m \exp\left(-\frac{Q}{k_{\mathrm{B}}T}\right) , \tag{39.14}$$

where m, τ_0, Q, and v_0 are empirical constants depending on the semiconductor material and its doping, k_{B} is the Boltzmann constant, τ_{eff} is the effective resolved shear stress on a dislocation and is given by

$$\tau_{\mathrm{eff}} = \tau - A\sqrt{N} \,, \tag{39.15}$$

τ is the externally applied shear stress on the active slip system, and $A\sqrt{N}$ is the back-stress on a given dislocation due to the neighboring ones, which resist its movement through interaction. The square-root dependence of the back-stress on N is a classical one and was first derived by *Taylor* [39.35], based on an arrangement of parallel dislocations on a slip plane. Though the actual arrangement of the dislocations (the microstructure) was found to be much more complicated than that assumed by Taylor, this form of the back-stress has been found to be valid under a wide range of conditions and is widely used in work-hardening theory [39.36]. The

dislocations multiply as they move, resulting in evolution of the dislocation density. It is reasonable to assume that the multiplication rate is proportional to the dislocation density and the velocity of their movement. Thus the evolution of dislocation density can be modeled by

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \delta N v \,. \tag{39.16}$$

Based on experimental evidence from [39.37] δ was assumed to be directly proportional to the effective stress τ_{eff}. *Alexander* and *Haasen* [39.33] opine that the multiplication mechanism is mainly due to the action of long jogs on screw dislocations, and the number of these jogs present is proportional to the effective stress, giving rise to the factor of τ_{eff}, therefore, $\delta = k\tau_{\mathrm{eff}}$. Combining (39.13–39.16), the following two equations are obtained

$$\frac{\mathrm{d}\varepsilon_{\mathrm{pl}}}{\mathrm{d}t} = \phi b N v_0 \left(\frac{\tau - A\sqrt{N}}{\tau_0}\right)^m \exp\left(-\frac{Q}{k_{\mathrm{B}}T}\right), \tag{39.17}$$

and

$$\frac{\mathrm{d}N}{\mathrm{d}t} = kNv_0 \left(\frac{\tau - A\sqrt{N}}{\tau_0}\right)^{m+1} \exp\left(-\frac{Q}{k_{\mathrm{B}}T}\right). \tag{39.18}$$

Equations (39.17) and (39.18) form a complete set of phenomenological models, which given the externally applied shear stress τ, can model the evolution of the plastic strain and dislocation density with time. They describe the yield region in the curve shown in Fig. 39.10 rather well, and can account for the occurrence of the upper and lower yield stress. In the stress–strain curve shown in Fig. 39.10, the total strain rate, which is the sum of elastic and plastic strain rates, is maintained constant. Hence

$$\dot{\varepsilon} = \phi\frac{\dot{\tau}}{G} + \dot{\varepsilon}_{\mathrm{pl}} \quad \text{or} \quad \dot{\varepsilon} = \phi\frac{\dot{\tau}}{G} + \phi b N v \tag{39.19}$$

is held constant, where G is the shear modulus.

Equations (39.17–39.19) can be solved numerically to derive the stress–strain curve for constant-strain-rate compression tests. The results of these numerical solutions have confirmed the experimental observation that the upper and lower yield stresses are very sensitive to the strain rate, the temperature, and in some cases the initial dislocation density in the material. They replicate the experimental observation of the upper and lower yield stresses of the form

$$\tau_{\mathrm{uy}} \propto \dot{\varepsilon}^{\frac{1}{n}} \exp\left(\frac{U}{k_{\mathrm{B}}T}\right)$$

and

$$\tau_{\mathrm{ly}} \propto \dot{\varepsilon}^{\frac{1}{n}} \exp\left(\frac{U_1}{k_{\mathrm{B}}T}\right),$$

where τ_{uy} is the upper yield stress and τ_{ly} is the lower yield stress. *Alexander* and *Haasen* [39.33] through some approximations arrive at the results

$$\tau_{\mathrm{uy}} \propto \dot{\varepsilon}^{\frac{1}{3}} \exp\left(\frac{Q}{3k_{\mathrm{B}}T}\right)\left[\ln\dot{\varepsilon} + \left(\frac{Q}{k_{\mathrm{B}}T}\right) - \tfrac{3}{2}\ln N_0 + \text{const.}\right]^{1/3}$$

and

$$\tau_{\mathrm{ly}} \propto \dot{\varepsilon}^{\frac{1}{3}} \exp\left(\frac{Q}{3k_{\mathrm{B}}T}\right)\left(A^2 b B_1\right)^{1/3},$$

where $B_1 = v_0/\tau_0$, Q, A, v_0, and τ_0 are parameters from (39.14) and (39.15) and m has been assumed to be 1. N_0 is the initial dislocation density. From the above analysis, the following conclusions can be drawn:

1. The numerical calculations from (39.17–39.19) predict the upper and lower yield stresses for a given deformation experiment. The predictions are quite accurate for a number of materials except that the lower yield stress is slightly underpredicted.
2. A log–log plot of the lower yield stress versus $1/T$, at constant strain rate, would yield a straight line, the slope of which is related to the quantity activation energy for dislocation velocity Q from (39.14). The same quantity Q can be estimated from dislocation velocity measurements and deformation tests. Values estimated independently from both methods have been shown to coincide for a number of semiconductor materials, thus giving an indirect validation of the above formulation [39.33].
3. The upper yield stress is shown to depend on the initial dislocation density. This is true for elemental semiconductors. In compound semiconductors such as InP and GaAs, the upper yield stress is experimentally seen to be independent of the initial dislocation density [39.38, 39]. This is because the dislocations present initially are locked in their position by impurities, and the surface acts as an effective source of dislocations. Hence the initial dislocation density does not play a big role in the deformation.
4. The lower yield stress is predicted to be independent of the initial dislocation density of the material. This is found to be true experimentally, with the

lower yield stress showing only weak dependence on the initial dislocation density. *Alexander* and *Haasen* [39.33] postulated that the state of the crystal at the lower yield point is a stationary state. In this state the dislocation density is at an optimum value, which gives a minimum stress for a given strain rate. The value of this optimum dislocation density was derived to be equal to $(4/9)\tau_{\mathrm{ly}}^2/A^2$, which is independent of the initial dislocation density. Substituting the above result into (39.17), a state of plasticity of the form $f(\dot{\varepsilon}, T, \tau_{\mathrm{ly}}) = 0$ can be derived, independent of the state of the material at the beginning of the deformation.

5. The results of the uniaxial strain rate tests were compared with uniaxial, constant stress creep tests, with the material oriented for single slip. The generic shape of strain versus time is shown in Fig. 39.11. It can be seen that the inflexion point with strain ε_{w} has the largest strain rate. This again corresponds to a state of plasticity; the dislocation density at this point is again at the same optimum value, and the material is softest, corresponding to a maximum strain rate at this point. From the theory outlined above, this state of plasticity should satisfy the same relation as above: $f(\dot{\varepsilon}_{\mathrm{w}}, T, \tau) = 0$, as has been verified experimentally by *Völkl* [39.40].

Thus some level of understanding has been reached for the behavior in the yield regions, though there are still limitations to the theory as will be shown later. After the stage of the yield points, the stress–strain curve

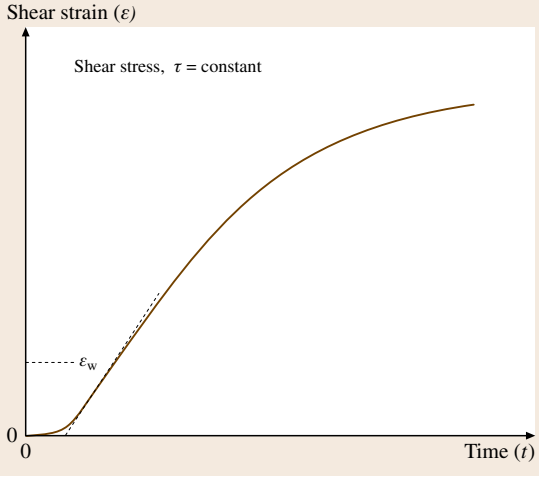


Fig. 39.11 Creep curve with specimen in single-slip orientation and constant applied stress

is followed by five more distinct stages at higher strains. Some of these stages are found only at higher temperatures and represent alternate stages of hardening and softening of the material due to phenomena such as cross-slip and climb of dislocations.

Stage I

This is also known as the stage of easy glide. In this stage, the dislocation structure changes from a predominance of screw dislocations in the yield regions to a predominance of edge dislocations, dipoles, and multipoles. This stage is characterized by low hardening.

Stage II

This stage is characterized by a constant hardening rate $\theta = \mathrm{d}\tau/\mathrm{d}\varepsilon$, which is almost invariant with respect to temperature and strain rate changes. This value is also close to the hardening coefficients of stage II of fcc metals ($\approx 2.4\times10^{-3}G$, where G is the shear modulus of the material). The deformation in this region is characterized by inhomogeneity of dislocation distribution and a number of locally activated glide systems.

Stage III

This is characterized by the softening behavior of the crystal, during which the hardening coefficient decreases. This stage is also exhibited in fcc crystals, which might lead to assumptions about similar dislocation mechanisms operating in this range. However a detailed analysis of the temperature and strain rate dependence, together with metallographical investigations, reveals that stage III softening in semiconductors is due to climb of dislocations, while in metals it is due to cross-slip of dislocations. *Siethoff* and coworkers have extended the experimental data for this stage for Si and Ge [39.41], GaAs [39.42], and InP [39.43], and in [39.44] have fitted this data to the dislocation climb model of *Mukherjee* et al. [39.45] and the jog-dragging model of *Barret* and *Nix* [39.46]. They found the data to be more consistent with the jog-dragging model, which proposed the softening to be a consequence of interaction of point defects with dislocations when screw dislocations drag long jogs along with them. However, more light needs to be shed on this stage, before the final word can be said on deformation mechanisms of stage III.

Stages IV and V

These stages were discovered in Ge by *Brion* et al. [39.47] and occur only in a narrow tempera-

ture range below the melting point T_m. Stage IV, like stage II, is characterized by linear hardening, where the hardening coefficient is invariant with temperature and depends slightly on strain rate. Stage V is similar to stage III and is characterized by softening or dynamic recovery. Investigations have revealed the recovery to be due to the cross-slip of screw dislocations, and *Siethoff* [39.48] have fitted the data obtained for stage V to *Escaig*'s theory for cross-slip of screw dislocations [39.49].

An important observation from experiment is the shrinkage of the stages with a rise in temperature. Stages IV and V do not occur at low temperatures. As the temperature is raised, each of the stages begins earlier, i. e., at lower strain, in the deformation curves. This suggests that dislocation mechanisms that lead to stages II–V begin to operate earlier at high temperatures. The deformation, due to thermal stresses, in crystal growth processes is not expected to exceed stage I, but in areas of high temperature close to the melting point, such as the crystal–melt interface, mechanisms prevalent in the later stages such as cross-slip, climb, and interaction with point defects may be important.

39.7 Application of the Haasen Model to Crystal Growth

Völkl [39.40] was the first to use the above model to study dislocation generation in the Czochralski system. In this analysis, the entire thermal and thermal–elastic stress history of a given particle in the crystal, from the particle's evolution from the melt–crystal interface to the end of the growth process, is determined. The thermal–elastic stress history so determined is used to determine the resolved shear stress history on the highest stressed slip system; an unstated assumption by Völkl is that the same slip system is the most highly stressed throughout the growth history of the particle. The resolved shear stress history is designated $\tau_t^{gs}(t)$. Völkl assumes that, as the plastic deformation proceeds, the resolved shear stress relaxes or is dissipated by the plastic strain, and the residual elastic stress $\tau_{er}(t)$ is the driving force behind plastic deformation. The relaxation process can be described by

$$\frac{d\tau_{er}(t)}{dt} + \frac{G}{\varphi}\frac{d\varepsilon_{pl}}{dt} = \frac{d\tau_t^{gs}(t)}{dt} . \qquad (39.20)$$

In conjunction with (39.20), (39.17), and (39.18) with τ replaced by $\tau_{er}(t)$ this leads to

$$\frac{d\varepsilon_{pl}}{dt} = \phi b N v_0 \left(\frac{\tau_{er}(t) - A\sqrt{N}}{\tau_0}\right)^m \exp\left(-\frac{Q}{k_B T}\right) , \qquad (39.21)$$

$$\frac{dN}{dt} = kNv_0 \left(\frac{\tau_{er}(t) - A\sqrt{N}}{\tau_0}\right)^{m+1} \exp\left(-\frac{Q}{k_B T}\right) . \qquad (39.22)$$

Since the thermal–elastic history $\tau^{gs}(t)$ has already been determined from the thermal history of the crystal, through a finite element code, (39.20–39.22) can be integrated simultaneously to obtain the final dislocation density N. The dislocation density, in the volume element, after it just emerges out from the melt is taken as a fixed constant N_0, i. e., the initial condition

$$N\big|_{t=0} = N_0 .$$

However, according to Völkl the final values of N are reasonably independent of the N_0 chosen. It is apparent that a choice of $N_0 = 0$ will lead to a singularity. Völkl justifies his assumption of single slip by observing that an exponential dependence of plastic deformation on stress will cause the most highly loaded slip system to act and thus release the stresses on all other slip systems. *Maroudas* and *Brown* [39.50] used an isotropic generalization of the plastic strain response in the Alexander–Haasen model and performed what they called an integrated analysis of crystal growth. *Tsai* [39.51] has also applied the Alexander–Haasen model to bulk crystal growth. He generalized the Alexander–Haasen model to an isotropic plastic response instead of assuming deformation on the most highly stressed slip system. *Tsai* et al. [39.52] later formulated a multislip generalization of the Alexander–Haasen model, involving all 12 slip systems active in the semiconductor crystal. *Lambropoulos* and *Wu* [39.53] applied the model to Czochralski growth of GaAs, and investigated the effect of the shape of the interface on the final dislocation generation, while *Miyazaki* and *Kuroda* implemented this model for Czochralski (CZ) growth, using a finite element technique [39.54].

The driving force behind deformation in the growing crystal is the thermal strain. Since thermal strains are small, the deformations are likely to fall within the yield region in Fig. 39.10, thus justifying the use

of the Alexander–Haasen model. However, there is no evidence that the thermal loading would favor the single-slip scenario. Moreover, as noted earlier, stages IV and V in Fig. 39.10 appear only at high temperatures. Higher temperatures are characterized by shrinkage of the stress–strain curves – the various stages occur at lower strains. This indicates that phenomena such as cross-slip and climb of dislocations, which dominate the later stages, become important at lower strains, for deformation at high temperatures. Thus, at temperatures approaching the melting point of the crystal – for which little data is available – these phenomena might be important even in the yield-point region. A multislip generalization of the Alexander–Haasen model is therefore needed, with provisions for high-temperature phenomena such as cross-slip, climb, etc. A model by *Moosbrugger* [39.55] and *Moosbrugger* and *Levy* [39.56] considers several of these features for CdTe crystals, and is the most advanced of the Alexander–Haasen variants.

39.8 An Alternative Model

The Alexander–Haasen model, though impressive, suffers from certain drawbacks for its application to compound semiconductors, and even elemental semiconductors. It is well known that the structure of the lattice in semiconductors gives rise to different types of dislocations, for example, screw, α dislocations, β dislocations, etc. All of these have different mobilities, and each one of them is expected to follow a different relation of the type (39.14). Perhaps, the adoption of a single dislocation variable N and the characterization of the dislocation velocity by a single equation of the type (39.14) is insufficient. Moreover, equations of the type (39.14) have been experimentally verified only for certain range of temperatures and stresses, which do not quite cover the entire gamut of these occurring in the crystal growth process. The specification of the initial dislocation density N_0 is also problematic. What is N_0 in the growing crystal? *Völkl* [39.40] as well as other researchers, as mentioned in the previous section, assumed a certain low N_0 along with the claim that this does not affect the final result much. This claim might be true, but such a treatment means that the final grown crystal, even for Si, always contains dislocations, since the dislocations move and multiply even at small stresses. Such a situation is incompatible with daily experience, as silicon single crystals are grown without any dislocations.

The deformation experiments looked at so far are uniaxial creep or constant-strain-rate tests, where the crystal has been oriented for single slip. In multislip deformations, or in applications such as crystal growth, dislocations in many slip systems are expected to be active. Thus what is need is a generalization of the Alexander–Haasen model, retaining the basic structure and ideas, but expanding it to include notions of multislip, different types of dislocations, modified multiplication laws, the possible effects of dopants, and some information on the initial dislocation sources in the crystal. Let us examine each of these issues to build a model.

39.8.1 Different Types of Dislocations

Semiconductors have a high Peierls energy, resulting in deep Peierls wells along the ⟨110⟩ directions. Consequently, the dislocations tend to align themselves along the ⟨110⟩ directions to minimize the energy. Thus typical dislocation loops tend to assume a hexagonal shape. Transmission electron microscopy (TEM) studies have shown that this condition holds when the dislocation density is low and the impurity concentration in the crystal is not too high [39.25, 27]. If the dislocation density is high, the local back-stresses between the dislocations will distort the hexagonal shape, and the dislocations may become curved, no longer lying in ⟨110⟩ directions. The conditions of low dislocation density might be expected to hold in crystal growth. X-ray topography by *Dudley* et al. [39.57, 58] shows long segments of screw dislocations aligned along the ⟨110⟩ directions, and hexagonal loops. The geometry

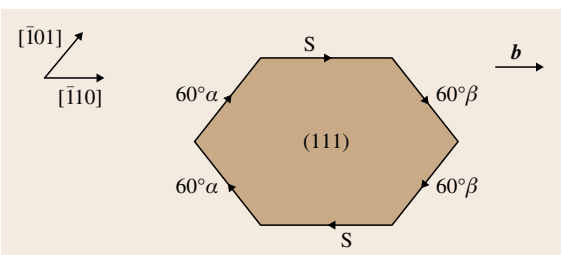


Fig. 39.12 Geometry of a hexagonal dislocation loop on the (111) plane.

of such a loop is shown in Fig. 39.12. Since there are three different dislocation types, the shear strain in this (111)[$\bar{1}$10] slip system is given by a straightforward extension of the Orowan equation

$$\dot{\gamma}_{\mathrm{pl}} = \left(N_\alpha \bar{v}_\alpha + N_\beta \bar{v}_\beta + N_\mathrm{s} \bar{v}_\mathrm{s}\right) b \,, \tag{39.23}$$

where N_α, N_β, and N_s are the densities of α, β, and screw dislocations, respectively, in the (111) plane with the Burgers vector in the [$\bar{1}$10] direction. Here v_α, v_β, and v_s are the corresponding velocities, with the overbar representing an average. As the loop in the figure expands, it can be seen that the screw segments expand due to the movement of α and β segments, while these segments expand primarily by the movement of the screw dislocations. If the lengths of these segments are denoted by L_α, L_β, and L_s, their rate of change can be approximately given by

$$\dot{L}_\alpha = 2\bar{v}_\mathrm{s}, \quad \dot{L}_\beta = 2\bar{v}_\mathrm{s} \quad \text{and} \quad \dot{L}_\mathrm{s} = 2\left(\bar{v}_\alpha + \bar{v}_\beta\right) .$$

Since the evolution of the dislocation densities of α and β types are governed approximately by the same quantity, a good approximation would be

$$N_\alpha = N_\beta$$

and

$$\frac{N_\alpha + N_\beta}{N_\mathrm{s}} = \frac{2\bar{v}_\mathrm{s}}{\bar{v}_\alpha + \bar{v}_\beta} \,.$$

Substituting the above approximations into (39.23), we obtain the following approximation for the evolution of the shear strain

$$\dot{\gamma}_{\mathrm{pl}} = 2N_\mathrm{s} \bar{v}_\mathrm{s} b \,.$$

Thus the evolution of the shear strain in a slip system is governed chiefly by the velocity and density of the screw dislocations in that system. Treatment similar to the above was first proposed by *Steinhardt* and *Haasen* [39.59] to explain results of creep experiments on GaAs, and was more clearly expounded by *Yonenaga* et al. [39.60], whom the above follows closely.

Since there are 12 {111}⟨$\bar{1}$10⟩ active in the lattice, the plastic shear strains on all of these have to be dyadically combined to get the total plastic strain rate, which is a second-order tensor. This can be expressed as

$$\dot{\boldsymbol{\varepsilon}}^{\mathrm{pl}} = \sum_{i=1}^{12} N_\mathrm{s}^i \bar{v}_\mathrm{s}^i b (\boldsymbol{n}_i \otimes \boldsymbol{s}_i + \boldsymbol{s}_i \otimes \boldsymbol{n}_i) \mathrm{sign}(\tau_i) \,, \tag{39.24}$$

where index i represents the slip system, $\boldsymbol{s}_i$ represents the slip direction of the slip system i, $\boldsymbol{n}_i$ represents the normal to the slip system i, N_s^i represents the density of screw dislocations on the slip system i, and $\bar{v}_\mathrm{s}^i$ is the average velocity of these dislocations. sign(τ_i) is the sign of the resolved shear stress along the slip system. Equation (39.24) represents a three-dimensional generalization of the Orowan equation (39.13), taking into account the geometry of a dislocation loop in compound semiconductors. The slip and slip plane normal vectors for the 12 slip systems are given in Table 39.1.

39.8.2 Dislocation Glide Velocity

The glide velocity of dislocations was investigated primarily for segments of the screw and 60° type mainly by the technique of producing dislocation half-loops through deliberate scratches on the surface of a semiconductor. Once the scratches are polished off, the material is subjected to loading (four-point bending, or tension or compression) and the glide velocities of the resulting dislocations are measured. This technique has been used for a wide range of temperature and stresses, and for various elemental and compound semiconductors. Following the work of *Chaudhuri* et al. [39.34] data has been collected for Si [39.61–63], and III–V semiconductors such as GaAs [39.59], InP [39.64–66], etc. A consensus exists regarding the results for measurements in the so-called central range of stress and temperature. *George* and *Rabier* [39.31] define this central range as

$$0.45 T_\mathrm{m} < T < 0.65 T_\mathrm{m} \,,$$
$$5 \times 10^{-5} \mu < \tau < 10^{-3} \mu \,,$$

where T_m is the melting point of the material and μ is the shear modulus of the material.

The temperature and shear stress dependence of the glide velocity of screw and 60° segments is as follows.

In the central range of measurements, the temperature dependence of the dislocation velocities at constant stress is given by an Arrhenius-type law of the form

$$v \propto \exp\left(-\frac{Q(\tau)}{k_\mathrm{B} T}\right) ,$$

where $Q(\tau)$ is a stress-dependent activation energy, τ is the shear stress acting on that slip system, and k_B is the Boltzmann constant. The activation energy Q is of the order of 1–2 eV in most semiconductors. The stress dependence of Q has been found to fall under two different regimes. For 60° dislocations in Ge, it is weakly stress dependent for stresses above 20 MPa, and is a strongly decreasing function of stress below this

stress. A similar transition stress of 10 MPa is found even for screw dislocations in Ge. In Si, the transition stress for 60° dislocations is 10 MPa, while none has been found for screw dislocations. The Arrhenius law has also been observed to breakdown in Si and Ge at some critical temperatures. *Farber* and *Nikitenko* measured 60° dislocation glide velocities up to temperatures of $0.9T_m$ in Si [39.63] and have observed some discrepancies. The velocities still depict an Arrhenius-type behavior, but the activation energy and prefactor are changed at temperatures above $0.75T_m$.

The dependence on stress of the dislocation glide velocity at constant temperature is conveniently described by a power law of the form

$$v \propto \exp(\tau^m) .$$

Combining the above two dependencies, an overall dependence of the form

$$v = v_0 \left(\frac{\tau}{\tau_0}\right)^m \exp\left(-\frac{Q(\tau)}{k_B T}\right) \qquad (39.25)$$

is obtained. This form is applicable for higher stresses, 20–100 MPa in Ge and > 30 MPa in Si, with constant m and Q. Below this range m and Q cannot be taken as constants. A double-logarithmic plot of v versus τ at constant temperature shows a bend. This is partly due to the stress dependence of $Q(\tau)$ below the critical stress mentioned above. *George* et al. [39.61] opine that, below the critical stress, Q can be expressed in the form

$$Q = Q_0 - E_i \ln\left(\frac{\tau}{\tau_0}\right) .$$

Alexander [39.27] has explored the stress range above the critical stress for Si, and found that the parameter m depends not only on the type of dislocation but also on the glide system to which the dislocation belongs, and nonshear stress components etc. These complications are not well understood. However, *Sumino* [39.25] opines that the above complications are a result of impurities immobilizing the dislocations at low stresses, and also the many problem associated with the technique of measuring dislocation velocities described above. *Alexander* discusses these in reference [39.27]. *Imai* and *Sumino* [39.62] used a live x-ray topography technique to measure the velocity of dislocations in highly pure Si in the temperature range 600–800 °C and stress range 1–40 MPa. They found perfect agreement with (39.25), with $m = 1$, and $Q = 2.20$ eV for 60° dislocations and $Q = 2.35$ eV for screw dislocations.

Hence the use of (39.25) lies on shaky foundations with several opinions expressed regarding its form and range of usage for different materials. Several, theoretical models have been put forward to explain the phenomena of glide, a few of which will be reviewed, before a brief summary of literature about dislocation glide velocities in InP is presented.

Dislocation glide in diamond structured elemental and compound III–V semiconductors takes place through the formation and migration of double kinks. The theory has been formulated by *Hirth* and *Lothe* [39.6, Chap. 15]. They derived the nucleation rate of double kinks on a dislocation as

$$J = \nu_0 \frac{\tau b h}{k_B T} \frac{\exp(-2F_k - W_m)}{k_B T} ,$$

where ν_0 is the Debye frequency, τ is the applied shear stress, $2F_k$ is the formation energy of a double kink under the action of a shear stress τ, W_m is the migration energy of the kink, b is the magnitude of the Burgers vector, and h is the kink height, or the distance between two Peierls valleys. For small kink densities or short dislocation segments, the double kinks traverse the entire length of the dislocation, without being annihilated, and the dislocation glide velocity is proportional to the length of the segment L. Hence

$$v = hLJ \text{ or } v = \frac{L\tau b h^2}{k_B T} \frac{\nu_0 \exp(-2F_k - W_m)}{k_B T} .$$

For long dislocation segments, the kinks are annihilated when they meet kinks of opposite sign, before they reach the ends of the segment. The mean free path X of the kinks in this case is given by

$$\frac{1}{\frac{X}{2}J} = \frac{X}{2v_k} \quad \Rightarrow \quad X = 2\sqrt{\frac{v_k}{J}} ,$$

where v_k is the kink velocity given by $v_k = (\tau b h)/(k_B T)\nu_0 a^2 \exp(-W_m/k_B T)$, where W_m is kink migration energy, and a is the jump distance of a kink.

In the case of long dislocation segments, the velocity is given as

$$v = hJX = 2h(Jv_k)^{\frac{1}{2}} = 2\frac{\nu_0 \tau b h^2}{k_B T} \frac{a \exp(-F_k - W_m)}{k_B T} .$$

The above expressions suggest that the activation energy of dislocation glide would vary with the dislocation length; there exists a transition region between the two regimes. *Louchet* [39.67] confirmed this. He was able to observe that the velocity of long dislocation segments was independent of length, while the velocity

of short segments was proportional to their length. He determined the critical length for $\tau = 90\,\mathrm{MPa}$ as $L_c = 0.4\,\mu\mathrm{m}$, which was interpreted as the mean free path of the kinks before annihilation. From these measurements, W_m and F_m were calculated separately, and both were of the magnitude of 1 eV. It was previously assumed that, analogously to fcc metals, the W_m in semiconductors was very small, and the presence of local obstacles along the dislocation was postulated to explain the discrepancy of a high W_m obtained. Theoretical treatments with this assumption presented by *Celli* et al. [39.68] and *Rybin* and *Orlov* [39.69] yield a very high density of obstacles, or a very high activation energy for dislocation glide velocities. The theory of local obstacles along the dislocation, opposing kink motion, is not accepted any longer, and it is now recognized that the high W_m is a result of a high secondary Peierls potential. The existence of a high Peierls means that migration of kinks is as difficult as their nucleation. This is related to the core structure of the partial dislocations, comprising the dislocation segment. It is believed that, in elemental semiconductors, the bonds on partial dislocations and kinks are reconstructed, except at certain defects along the core of the partials called solitons, or antiphase defects. It is only at these solitons that dangling bonds are present. The solitons play a direct role in the nucleation and migration of kinks, accounting for the same order of magnitude of the kink migration and nucleation energies.

Several analyses of the core structure of 90 and 30° partial dislocations, and estimates of W_m and F_k, have been performed [39.70, 71]. These analyses are based on atomistic simulations, using some empirical interatomic potentials, or ab initio methods, which are able to reproduce the quantum mechanical effects. These techniques, along with recently developed electron microscopy techniques [39.72] for the direct observation of movement and nucleation of kinks, should clarify W_m and F_k in the future. It should be pointed out that the above formalism is based on nucleation and migration of kinks in whole dislocations, while in reality the whole dislocation segments are split into partial dislocations. *Möller* [39.73] performed an analysis taking into account the presence of partial dislocations. He predicted the existence of a critical stress below which dislocation glide takes place through the correlated motion of the partials. At stresses above the critical stress, the glide took place with the movement of the two partials being independent. The correlated movement of partial dislocations has higher activation energy than the uncorrelated movement. This analysis somewhat reproduced the experimental observations in Ge and Si about the existence of two different regimes for the stress-dependence of the activation energy Q, decribed earlier in this section (Sect. 39.8.2). Möller's analysis was performed under the framework of local obstacles along the dislocation, and there is a need to adapt this to more recent ideas of a high secondary Peierls potential.

The influence of doping on the glide velocity of dislocations in semiconductors is well recorded. This is realized chiefly through electronic and metallurgical interaction. The metallurgical effect is classical and is relevant to any type of dopant or impurity in the semiconductor. It is a local effect and is due to the interaction between the impurity atom and the dislocation core. Another manifestation of this is the so-called locking stress, which certain impurities cause in certain semiconductors. At shear stresses below the locking stress, the dislocations are locked and do not glide and, once the shear stress exceeds the locking stress, the dislocations glide smoothly. Examples of this phenomenon are O, P, and S in Si. The electronic interaction is prevalent only for electrically active dopants or impurities when their concentration exceeds the intrinsic carrier concentration of the material. It has been found that III or V doping in elemental semiconductors affects the glide velocities in the same way irrespective of the size and chemical nature of the dopant. The primary influence of these electrically active dopants is that they change the Fermi energy level in the semiconductor. There are two classes of theories to explain the effect on dislocation glide velocity. *Haasen* [39.74] postulate that the line charge on a dislocation in a semiconductor depends on the position of the Fermi energy level. A charged dislocation has the formation energy of a double kink reduced, because of the electrostatic repulsion between the two kinks. This affects the dislocation glide velocity. *Hirch* [39.75] postulated that the formation of a double kink leads to the generation of acceptor E_{ka}, or donor E_{kd} levels in the band gap. Charged as well as neutral kinks can exist on a dislocation, while the concentration of the neutral kinks is constant, the concentration of charged kinks depends on the position of the Fermi level (E_F), and the charged kink's own level in the spectrum. The concentrations of negatively and positively charged kinks are given by

$$\frac{C_k^+}{C_k^0} = \exp\left(\frac{(E_{kd} + \mathrm{eV}) - E_F}{k_B T}\right)$$

and

$$\frac{C_k^-}{C_k^0} = \exp\left(\frac{E_F - (E_{ka} + \mathrm{eV})}{k_B T}\right),$$

where V is the gate voltage. The above formulae are derived by the application of Fermi–Dirac statistics. The concentration of neutral kinks is constant, and hence the dislocation glide velocity is directly dependent on the concentration of charged links. If the kink's level is near the middle of the energy gap – close to the Fermi level – increasing the n-type doping would raise the Fermi level and increase the concentration of negatively charged kinks and consequently dislocation velocity. Similarly increasing p-type doping will decrease the Fermi level and raise the concentration of positively charged kinks and consequently the dislocation velocity. Consequently both n-type as well as p-type materials will have higher dislocation velocity than the intrinsic material, as in the case of silicon. On the other hand, if the kink's level is near the valence band – much below the Fermi level – the intrinsic material will have negatively charged kinks. Increasing the n-type doping will raise the Fermi level and increase the concentration of negatively charged kinks raising the dislocation velocity, while a p-type doping will lower the Fermi level and decrease the concentration of negatively charged kinks and lower the dislocation velocity. Thus in materials where the kink's level is near the valence band, the velocity should increase from p-type to n-type material, as in Ge. This gives some understanding of the doping effect.

Finally the complications introduced by the splitting into partial dislocations and the existence of two different types of atomic species in compound semiconductors should be recognized. A 60° dislocation splits into a 30 and a 90° partial, and the mobility of each partial is different. The velocity of the whole 60° dislocation depends on which partial is leading in the direction of the movement. Thus the same dislocation might show asymmetry in velocity depending on the direction of the movement. In compound semiconductors, the velocity of α and β dislocations of the same type are expected to be different because of the different core structures. Thus, a review of the theoretical and experimental aspects of dislocation glide velocity reveals that, at this stage, the theory is not yet comprehensive enough to yield concrete relations for dislocation glide velocities that can be used in a microdynamical theory of plasticity. The best that can be done currently is to adapt an expression of the type (39.14), with constant m and Q, while recognizing its inherent limitations – that the parameters m and Q are stress and temperature dependent.

There is some data available on the direct measurement of the velocity of dislocations for InP. For undoped and sulphur-doped material, the velocity of α dislocations was reported by *Nagai* [39.65] in the temperature range 250–400 °C. He fitted the velocities to a relation of the type (39.14), with $m = 2.7$, for both doped and undoped crystal. The value of Q was found to be constant, with $Q = 1.1$ eV for S-doped crystal (6.5×10^{18} cm^{-3}), and 1.0 eV for undoped InP. S doping was found to increase the velocity of α dislocations. *Maeda* and *Takeuchi* [39.66] set out to measure the effect of recombination-enhanced glide in InP crystal under the effect of minority carrier injection. In the course of their investigation, they measured the velocity dependence of β dislocations in S-doped crystal (5×10^{18} cm^{-3}) on temperature, in the range 220–410 °C. They found a Q value of 1.6 eV. *Yonenaga* and *Sumino* [39.64, 76] investigated the velocity of all three types of dislocations in InP crystals – undoped, and with a variety of dopings. The temperature range of their measurements was 300–500 °C, and stresses between 2 and 20 MPa. They fitted their data to an equation of type (39.14), with $\tau_0 = 1$ MPa. For undoped InP, they obtained the parametric fits presented in Table 39.2.

Table 39.2 Dislocation velocity data for undoped InP (after [39.64, 76])

Dislocation	v_0	m	Q (eV)
α	40 000	1.4	1.6
β	500 000	1.8	1.7
screw	40 000	1.7	1.7

Yonnenaga and *Sumino* [39.64] have also obtained the corresponding parameters for material doped with Zn (p-doping), S (n-doping), and the isovalent impurities Ga and As. The dislocation velocity data in Table 39.2 is used in the modeling effort presented later.

39.8.3 Dislocation Multiplication

Dislocations being line defects, any movement of these will involve self-multiplication during motion and increase their lengths. If some parts of the dislocation loop are less mobile than others, the more mobile parts will move around in a spiral, increasing their length. In the loop shown in Fig. 39.12, the β segments are more mobile than the α segments, which are more mobile than the screw segments. Under the influence of stress, the β segments will move quickly through the material while screws will lag behind, resulting in an expansion of the screw segment's length. Another typical case is jogs, found on screw dislocation. These jogs are less mobile than the main dislocation; hence these lag behind,

Table 39.3 The type of interactions between dislocations of different slip systems (N: pair belong to cross-slip systems, G: pair form glissile junctions, H: Hirth locks, S: sessile junctions, C: pair belong to coplanar systems. The numbers in parentheses are f_{ij})

i \ j	1	2	3	4	5	6	7	8	9	10	11	12
1		C(0)	C(0)	S(1)	G(0)	H(1)	N(0)	G(1)	G(1)	H(1)	S(1)	G(0)
2	C(0)		C(0)	G(1)	N(0)	G(1)	G(0)	S(1)	H(1)	S(1)	H(1)	G(0)
3	C(0)	C(0)		H(1)	G(0)	S(1)	G(0)	H(1)	S(1)	G(1)	G(1)	N(0)
4	S(1)	G(0)	H(1)		C(0)	C(0)	G(1)	N(0)	G(1)	G(0)	S(1)	H(1)
5	G(1)	N(0)	G(1)	C(0)		C(0)	S(1)	G(0)	H(1)	G(0)	H(1)	S(1)
6	H(1)	G(0)	S(1)	C(0)	C(0)		H(1)	G(0)	S(1)	N(0)	G(1)	G(1)
7	N(0)	G(1)	G(1)	G(0)	S(1)	H(1)		C(0)	C(0)	H(1)	G(0)	S(1)
8	G(0)	S(1)	H(1)	N(0)	G(1)	G(1)	C(0)		C(0)	S(1)	G(0)	H(1)
9	G(0)	H(1)	S(1)	G(0)	H(1)	S(1)	C(0)	C(0)		G(1)	N(0)	G(1)
10	H(1)	S(1)	G(0)	G(1)	G(1)	N(0)	H(1)	S(1)	G(0)		C(0)	C(0)
11	S(1)	H(1)	G(0)	S(1)	H(1)	G(0)	G(1)	G(1)	N(0)	C(0)		C(0)
12	G(1)	G(1)	N(0)	H(1)	S(1)	G(0)	S(1)	H(1)	G(0)	C(0)	C(0)	

while the more mobile part of the screw dislocation spirals around the jog, multiplying its length. Jogs can be formed either by the cross-slip mechanism, or by cutting of a screw dislocation by a forest dislocation in another slip plane. An example of this was shown by live topography of a dislocation multiplication process in silicon by *Sumino* and *Harada* [39.77], where a jog was formed on a screw dislocation by a 60° dislocation on a different slip plane cutting through it. The jog subsequently acted as a spiral center for multiplication. Dislocation multiplication can also occur by the Frank–Read mechanism, as was shown by an image of a Frank–Read source obtained by *Völkl* et al. [39.40]. The evolution of the dislocation density, through the dislocation multiplication, was modeled by the equation $\mathrm{d}N/\mathrm{d}t = k\tau_{\mathrm{eff}}Nv$ in the Alexander–Haasen model. The basic premise of this equation was that the total length of dislocation created $\mathrm{d}N$ is proportional to:

1. The distance moved by the dislocation.
2. The length of already existing dislocations, since the longer the existing dislocations, the greater the probability of jogs forming on them.
3. The effective stress τ_{eff} in the equation is to represent the stress required by the spiraling segment in the parallel slip plane – which has been created by a jog – to pass the screw dislocation in the original slip plane.

Sumino and *Yonenaga* [39.76] retained the basic structure of the Haasen law and added to it features of multislip deformation and the different types of dislocations found in compound semiconductors. They introduced a dislocation multiplication law of the form

$$\frac{\mathrm{d}N_{\mathrm{s}}^{i}}{\mathrm{d}t} = KN_{\mathrm{s}}^{i}\bar{v}_{\mathrm{f}}^{i}\tau_{\mathrm{eff}}^{i} + K^{*}N_{\mathrm{s}}^{i}\bar{v}_{\mathrm{f}}^{i}\tau_{\mathrm{eff}}^{i}\sum_{j\neq i}N_{\mathrm{s}}^{j}\,, \qquad (39.26)$$

where the index i represents one of the 12 slip systems, and the bar over the velocity denotes an average.

The above law models the evolution of screw dislocation density in all 12 possible slip systems in III–V compound semiconductors. The basic proportionality of the evolution rate to the dislocation density and velocity is retained. The average velocity of the fastest dislocation v_{f}^{i} is used, since the expansion of a spiral and a loop is controlled by the velocity of the fastest segment, which in the case of InP is the β segment. The second term models the formation of jogs on screw dislocations and the consequent expansion through spiral formation. The rate of formation of jogs is proportional to the number of times dislocations in different slip systems cut each other, which can be modeled by a product of dislocation densities on the two slip systems.

Equation (39.26) can therefore be generalized as

$$\frac{\mathrm{d}N_{\mathrm{s}}^{i}}{\mathrm{d}t} = KN_{\mathrm{s}}^{i}\bar{v}_{\mathrm{f}}^{i}\tau_{\mathrm{eff}}^{i} + K^{*}N_{\mathrm{s}}^{i}\bar{v}_{\mathrm{f}}^{i}\tau_{\mathrm{eff}}^{i}\sum_{j\neq i}f_{ij}N_{\mathrm{s}}^{j}\,. \qquad (39.27)$$

The f_{ij} are closely related to the formation of jogs. Following reference [39.78], the interaction between the dislocations of any pair of slip systems as categorized in Table 39.3 are:

1. Pairs belonging to two cross-slip systems such as 11 and 9, denoted by "N" in Table 39.3. When the dislocations of such systems cross each other, two kinks are formed, and no jogs result.
2. Pairs forming a Hirth lock. When the dislocations of such systems cross each other, jogs are formed on both dislocations. "H" denotes such interactions in Table 39.3.
3. Pairs that are in the same plane, and whose resulting Burgers vector is therefore on the same plane as the original ones. These interactions do not result in jogs on either system, and are denoted by "C" in Table 39.3.
4. Pairs forming glissile junctions, such as systems 1 and 12 in Fig. 39.1. The interaction between such systems produces a jog on one and a kink on the other. For example, if the dislocations of the two slip systems 1 and 12 pass, a jog is formed on slip system 12, while a kink is formed on slip system 1. The interactions between systems forming glissile junctions are denoted by "G" in Table 39.3.
5. Finally, the last type of interaction is between the pairs of systems forming sessile junctions, such as 4 and 1. When the dislocations of two such systems interact, they form stable Lomer–Cottrell locks, and impede the motion of both original dislocations. Such pairs are denoted by "S" in Table 39.3.

Alternatively, if the dislocations cross each other, a jog is formed on each one. The f_{ij} are assigned values of either 1 or 0 in Table 39.3 based on whether a jog is formed on a screw dislocation of system i when cut by a forest dislocation of system j. Equations (39.24) and (39.27) form a complete viscoplastic system with 12 internal variables. The meaning of the effective stress τ^i_{eff} is clarified next.

39.8.4 Work Hardening

The flow stress can be classically split into a rate-dependent term and a hardening term, independent of the strain rate and temperature. The rate-dependent part of the applied stress is used to overcome the frictional resistance of the lattice to the glide of a dislocation, which is temperature sensitive. The rate-independent part is the resistance from interaction with other dislocations, both in the same slip system as well as threading dislocations in other slip systems, and finally impurities. In the single-slip model described before, the work-hardening term was given by $A\sqrt{N}$. *Taylor* [39.35] first introduced this form, drawing from

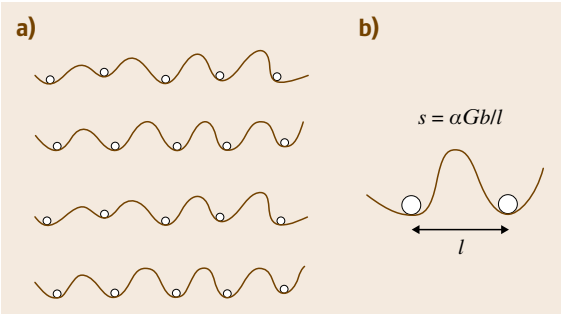


Fig. 39.13 (a) A series of dislocations facing barriers in the glide plane, and (b) a dislocation link stuck in an obstacle of strength s

theoretical treatment of the elastic interaction between parallel dislocations. The elastic theory of dislocations gives the magnitude of A as $Gb/(2\pi(1-v))$, where G is the shear modulus and v is the Poisson ratio. *Moulin* et al. [39.79] conducted a mesoscopic simulation, using discrete dislocation dynamics, for the yield point of Si, in a single-slip framework. From a fit of their data, they found the work hardening to be adequately defined by an equation of the type $A\sqrt{N}$, but the value of A was $1.48Gb/(2\pi(1-v))$. Interaction between dislocations of different slip systems gives rise to complicated hardening behavior. Apart from long-range elastic interactions, stable locks may form between dislocations of noncoplanar slip systems. *Hirth* and *Lothe* [39.6] discuss different types of locks, the most prominent of which are the Lomer–Cottrell lock and the Hirth lock.

Franciosi and *Zaoui* [39.78] have classified the different types of locks formed, and their efficiency in locking. These locks act as obstacles to other dislocations in the glide plane. Figure 39.13 shows a dislocation stuck between such obstacles. If a link of dislocation is stuck between two such obstacles of distance l (Fig. 39.13b) the shear stress s in the glide plane required to release the dislocation is of the order $s = \alpha Gb/l$, where G is the shear modulus, b is the Burgers vector, and α is some numerical factor.

Hence the resistance offered to the dislocation motion on a given glide plane is inversely proportional to $1/l$. Since l is the distance between obstacles, it is inversely proportional to the number of obstacles on the glide plane. The number of obstacles on a glide plane, due to forest dislocations on another glide system, say the α glide system, is proportional to $\sqrt{N_\alpha}$. Hence the total rate-independent resistance offered to dislocation motion in glide system β can be denoted heuristically

as

$$\sum_{\alpha\neq\beta} a_{\beta\alpha}\sqrt{N_\alpha} + \frac{Gb}{2\pi(1-\upsilon)}\sqrt{N_\beta}\,, \tag{39.28}$$

where the second term is the self-hardening term of the Alexander–Haasen model. An equation of the type (39.28) has been used in the literature to justify experimental observations on latent hardening in Cu [39.78].

The coefficients $a_{\beta\alpha}$ are phenomenological and might depend on many factors including (1) the geometry of the forest dislocation configurations, which determines the distance between the intersections of these with the primary glide plane, and (2) the strength of the locks formed, which might depend on other factors such as the stacking fault energy. *Franciosi* and *Zaoui* [39.78] classified the $a_{\beta\alpha}$ based on the strength of the locks which are formed. They classify them according to whether the dislocations belong to the cross-slip systems (interaction coefficient a_1), form Hirth locks (interaction coefficient a_2), belong to coplanar systems (interaction coefficient a_3), form glissile junctions (interaction coefficient a_4), or form sessile Lomer–Cottrell locks (interaction coefficient a_5). They found the a coefficients to increase linearly with the stacking fault energy, and the anisotropy in the a coefficients to increase with decreasing stacking fault energy.

Alexander and *Crawford* [39.80] determined the interactions of different slip systems through latent hardening experiments. The procedure they used was as follows: A Ge single crystal was deformed at constant strain rate in single-slip orientation up to stage I. The lower yield stress was noted. The crystal sample was cut into smaller samples, which were then deformed at the same strain rate. Each sample was aligned for single slip, in such a way that the other secondary glide systems become the most stressed systems. The lower yield stress in each of these cases was noted, and compared with the lower yield stress in the initial sample. The results were then analyzed in the framework of the Alexander–Haasen model, with an equation similar to (39.28) adopted as the formula for the back-stress. Normalizing the $a_{\alpha\beta}$ by Gb, they found the normalized coefficients to be: 0.37 for systems forming Lomer–Cottrell locks, 0.24 for systems forming Hirth locks, 0.35 for glissile locks, and 0.64 for the cross-slip system. Thus, the cross-slip system was found to have the largest hardening coefficient. These results are at variance with those of *Franciosi* and *Zaoui* [39.78] cited above. *Alexander* and *Crawford* [39.80] explain their observation by noting that the structure in the primary glide system at the beginning of easy glide consists primarily of edge dipoles, and from the geometry it can be deduced that these intersect the cross-slip plane most effectively. In the case of crystal growth, the deformation structure is dominated by long screws, hence these coefficients should be expected to differ from those determined by *Alexander* and *Crawford* [39.80]. Also the stacking fault energy of InP is nearly four times smaller than that of Ge, hence the values might be expected to be more asymmetric.

Summarizing, an equation of the form (39.29) can be used for the hardening. Since the interaction coefficients are not available, one hopes to determine them from fitting deformation data in multislip and single-slip experiments. The effective stress in thus given by

$$\tau^i_{\text{eff}} = \left|\tau^i\right| - \frac{Gb}{\beta}\left(N^i_{\text{s}}\right)^{1/2} - \frac{Gb}{\beta^*}\sum_{j\neq i} g_{ij}\left(N^j_{\text{s}}\right)^{1/2}\,. \tag{39.29}$$

The g_{ij} are set as a_1, a_2, a_3, a_4 or a_5, depending on whether the entry in Table 39.3 is N, H, C, G or S.

39.8.5 The Initial Dislocation Density

One of the main issues in the application of the Alexander–Haasen model to crystal growth is the specification of the initial dislocation densities. An important question is: what is the dislocation density in just-solidified, virgin material near the interface? Dislocations can arise at the interface due to certain phenomena at the crystal–melt interface, such as the formation of inclusions. Such dislocations are known as growth dislocations. The phenomena giving rise to them are manifestations of instability in the crystal–melt interface, and are described qualitatively in [39.81, Sect. 6.2]. Postgrowth analysis has shown that growth dislocations usually arise at inclusions formed at the growth interface, intersect the interface, and are propagated along with the movement of the interface. When the crystal is grown under stress-free conditions, the growth dislocations intersect the interface at certain specific angles, depending on their Burgers vector, in order to minimize their elastic energy. Thus growth dislocations arise from inclusions, and under stress-free conditions propagate in straight lines, intersecting the interface at a predetermined angle [39.82]. This configuration is disturbed under Czochralski growth at high temperatures, where stresses and considerable plastic deformation are prevalent. Postgrowth analysis of InP single crystals has shown the absence of growth dislo-

cations. X-ray topographs by *Dudley* et al. [39.57, 58] show that most of the dislocations are glide or slip dislocations intersecting the sides of the crystal. They rarely intersect the crystal–melt interface. This can be interpreted to be due to the absence of the generation of dislocations due to interfacial phenomena in InP.

As discussed before, another mode for the nucleation of dislocations is the condensation of excess point defects into prismatic dislocation loops, or voids. This effect is aggravated in compounds such as GaAs, InP, and other binary compounds due to the presence of a narrow homogeneity region. In these systems, the stoichiometric compounds do not have the highest melting point; hence the first to solidify composition would be rich in one of the components. Thus, in the case of GaAs, the first to form solid is As rich, this As precipitates out during cool down, and precipitates of As are formed. These precipitates can give rise to dislocations. The modeling of point defect dynamics, and the clustering of these defects, is a complex process with many issues still unresolved.

Another important source of dislocations is surface damage. When the growing InP crystal emerges from the boric oxide encapsulant, the volatile phosphorous species can escape from the sides of the crystal, leading to extensive surface damage. This surface damage is a local distortion of the lattice and is a good site for dislocation nucleation. Under the influence of thermal stresses, the dislocations thus nucleated at the surface can move to the interior of the crystal and multiply. The rate of evaporation of the phosphorous vapor is closely related to the ambient pressure in the furnace.

Another source of dislocations is the initial seed crystal used in the CZ process. Dislocations present in the seed may move into the freshly formed material and multiply there under the influence of stresses. The necking process invented by *Dash* [39.3] prevents the movement of dislocations from the seed to the growing crystal. The term *necking* arises from the shape of the crystal being grown. The diameter of the growing crystal is decreased from the seed to a minimum value before increasing the diameter again. Thus the crystal has a neck of minimum diameter. Attempts have been made to extend the necking technique to sulfur-doped LEC-grown InP by *Dudley* et al. [39.83]. The InP crystals so grown were analyzed using synchrotron white-beam x-ray topography (SWBXT). Figure 39.14a,b presents the topographs of (110) wafers of two different as-grown InP crystals – sample A corresponding to a crystal grown with a narrow neck, and sample B corresponding to a crystal grown without necking. Figure 39.14c shows a geometric sketch of the dislocations in both the crystals. An analysis of these topographs reveals that most of the dislocations in LEC-grown InP crystals originate from the seed–crystal interface, propagate via slip planes, and

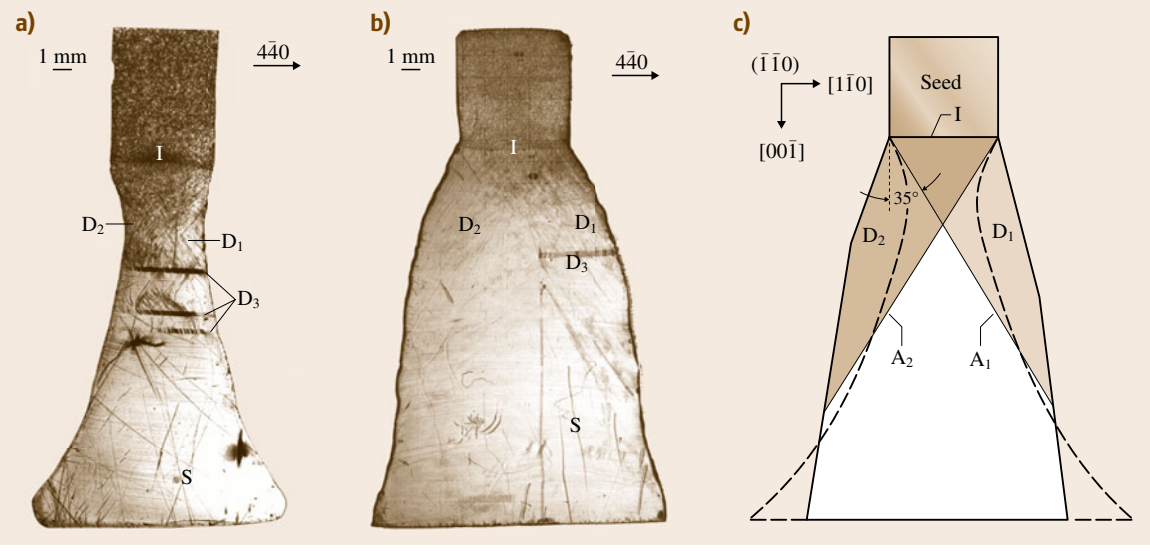


Fig. 39.14 **(a)** Sample A: X-ray topograph of (110) wafer of sulfur-doped InP crystal grown with a narrow neck. **(b)** Sample B: X-ray topograph of (110) wafer of sulfur-doped InP crystal grown without a narrow neck. **(c)** Schematic showing geometry and distribution of dislocations in as-grown crystals, for both shapes of crystals shown in **(a,b)** (after [39.83])

eventually exit the crystal through the periphery of the necked region. Based on this mechanism, dislocations exist only in the outer regions (dislocation regions) of the grown crystal separated by the four {111} planes which are extended from the four edges of the seed bottom. In the cross-sectional view of Fig. 39.14c, two of these planes, $(11\bar{1})$ and $(1\bar{1}1)$, are represented by lines A_1 and A_2, respectively. The regions of the crystal with dislocations are shaded, while those that are dislocation-free are unshaded.

As shown in Fig. 39.14c, the narrow necked portion of sample A makes the dislocation region much smaller than that of sample B. This means that the dislocations emerging from the seed–crystal interface in sample A slipped out of the neck in much shorter distances than that in sample B. A pyramid-shaped crystal, which is nearly dislocation free, was quickly formed below the dislocation regions. The above experiment and conclusions show the necessity of considering the movement of dislocations from dislocated regions to regions containing no dislocations initially. Such a formulation can be established using the continuum theory of dislocations, and would resemble the strain-gradient theories in structure. Such a formulation would involve spatial gradients of internal variables, and is not amenable to solution on the length scales involving a typical crystal wafer. Hence this line has not been pursued.

39.9 Model Summary and Numerical Implementation

39.9.1 Summary of the Model

A complete set of equations for the boundary-value problem using the constitutive model derived are summarized here, followed by a brief description of the numerical procedure.

Equilibrium of forces

$$\nabla \cdot \boldsymbol{\sigma} = 0 \,. \tag{39.30}$$

Compatibility

$$\boldsymbol{\varepsilon} = \tfrac{1}{2}\left(\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^{\mathrm{T}}\right) . \tag{39.31}$$

Split of total strain

$$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^{\mathrm{el}} + \boldsymbol{\varepsilon}^{\mathrm{pl}} + \boldsymbol{\varepsilon}^{\mathrm{th}} \,. \tag{39.32}$$

Constitutive relation for thermal strain

$$\boldsymbol{\varepsilon}^{\mathrm{th}} = \alpha \boldsymbol{\delta}(T - T_0) \,. \tag{39.33}$$

Constitutive relation for elastic strain

$$\boldsymbol{\varepsilon}^{\mathrm{el}} = C^{-1} : \boldsymbol{\sigma} \,. \tag{39.34}$$

Evolution equation for plastic strain and dislocation densities

$$\dot{\boldsymbol{\varepsilon}}^{\mathrm{pl}} = \sum_{i=1}^{12} N_{\mathrm{s}}^{i} \bar{v}_{\mathrm{s}}^{i} b (\boldsymbol{n}_i \otimes \boldsymbol{s}_i + \boldsymbol{s}_i \otimes \boldsymbol{n}_i) \mathrm{sign}(\tau_i) \,, \tag{39.35}$$

$$\frac{\mathrm{d}N_{\mathrm{s}}^{i}}{\mathrm{d}t} = K N_{\mathrm{s}}^{i} \bar{v}_{\mathrm{f}}^{i} \tau_{\mathrm{eff}}^{i} + K^{*} N_{\mathrm{s}}^{i} \bar{v}_{\mathrm{f}}^{i} \tau_{\mathrm{eff}}^{i} \sum_{j \neq i} f_{ij} N_{\mathrm{s}}^{j} \,. \tag{39.36}$$

Auxiliary equations

$$\tau_{\mathrm{eff}}^{i} = \left|\tau^{i}\right| - \frac{Gb}{\beta}\left(N_{\mathrm{s}}^{i}\right)^{\frac{1}{2}} - \frac{Gb}{\beta^{*}} \sum_{j \neq i} g_{ij} \left(N_{\mathrm{s}}^{j}\right)^{\frac{1}{2}}$$

and

$$\tau^{k} = \boldsymbol{n}_{\mathrm{k}} \cdot \boldsymbol{\sigma} \cdot \boldsymbol{s}_{\mathrm{k}} \,.$$

Related parameters

The values of K, K^{*}, β, β^{*}, and the interaction coefficients g_{ij} are obtained by fitting the data to results from the single-slip uniaxial compression tests at a constant strain rate. These tests were simulated using the model, and fitted to data in [39.76, 84]. The values obtained were: $K = 13 \times 10^{-6}\,\mathrm{Pa^{-1}m^{-1}}$, $K^{*} = 8 \times 10^{-15}\,\mathrm{m/Pa}$, $\beta = 3.3$, $\beta^{*} = 1.0$, $a_1 = 0.33$, $a_2 = 0.33$, $a_3 = 0.33$, $a_4 = 1.2$, and $a_5 = 1.3$. More details can be found in [39.85].

The elasticity moduli C_{ijkl} (indicial notation is used here) are based on [39.86] with the assumption that the continuum behaves as a cubic crystal with only three independent parameters. The relevant values are $C_{ijkl} = 1.02 \times 10^{11}\,\mathrm{Pa}$ when $i = j = k = l$, $C_{ijkl} = 4.4 \times 10^{10}\,\mathrm{Pa}$ when $i = k \neq j = l$ or $i = l \neq j = k$, $C_{ijkl} = 5.7 \times 10^{10}\,\mathrm{Pa}$ when $i = j \neq k = l$, and $C_{ijkl} = 0.0$ in all other cases. The coefficient of thermal expansion α is assumed to vary in quadratic fashion from $4.869 \times 10^{-6}\,\mathrm{K^{-1}}$ at 0 K and $4.410 \times 10^{-6}\,\mathrm{K^{-1}}$ at 100 K to $5.75 \times 10^{-6}\,\mathrm{K^{-1}}$ at 1000 K [39.86]. The magnitude of the Burgers vector b is $4.15 \times 10^{-10}\,\mathrm{m}$ [39.87, Table 1].

39.9.2 Numerical Implementation

A finite element technique can be adopted to obtain solutions of the boundary-value problem described above (indicial notation is used in this section). Briefly, this task involves the integration of the constitutive equations over time while ensuring equilibration of stresses in the body. Reference [39.88] gives a good description of the numerical implementation of viscoplastic equations. Assume that, at time t_n, the body is at equilibrium and that the set of displacements at each of the finite element nodes $\{\boldsymbol{u}_i\}_n$, and stresses $\{\boldsymbol{\sigma}_{ij}\}_n$, plastic strains $\{\boldsymbol{\varepsilon}^{\mathrm{pl}}_{ij}\}_n$, and the 12 internal variables $\{N^i_{\mathrm{s}}\}_n$ at the finite element Gaussian points are all known. However between time t_n and the subsequent time t_{n+1}, the temperature has changed and hence the thermal strain has changed. The task is now to find the new displacements, stresses, plastic strain, and dislocation densities at time t_{n+1}. This is accomplished by the following procedure:

1. The finite element technique runs either a global Newton–Raphson iteration or a quasi-Newton iteration for the updated displacements $\{\boldsymbol{u}_i\}_{n+1}$ at time t_{n+1}. This global iteration is based on a discretized version of (39.30).
2. A guess is made at every nodal point in the finite element domain for the set of updated displacements at the k-th global Newtonian iteration $\{\boldsymbol{u}_i\}^k_{n+1}$. The updated strains, at every Gaussian point in the finite element domain, for the k-th iteration can be calculated using the discretized version of (39.31), and the thermal strain at time t_{n+1} can be calculated from (39.33), using the thermal fields at time t_{n+1}.
3. To update the stresses, plastic strains, and the dislocation densities, (39.35) and (39.36) are integrated for the time step $\Delta t = (t_{n+1} - t_n)$ with the help of (39.31–39.34) and the auxiliary equations, at every Gaussian point. The integration method can be either an implicit Euler scheme, or any first-order ordinary differential equation (ODE) integration scheme such as one of the Runge–Kutta methods. The algorithmic elasto-plastic tangent modulus $\left\{D_{ijmn} = \partial\boldsymbol{\sigma}_{ij}/\partial\boldsymbol{\varepsilon}_{mn}\right\}^k_{n+1}$ can also be obtained for a given integration scheme. Refer to [39.85] for a detailed development using an implicit Euler scheme and a (2)3 Fehlberg pair, which is an explicit integration scheme.
4. Once the integration is completed, one obtains the stresses $\{\boldsymbol{\sigma}_{ij}\}^k_{n+1}$ and the algorithmic elasto-plastic tangent modulus at every Gaussian point in the finite element domain. These are then inserted into the global Newton–Raphson iteration, and the displacements at the next iteration $\{\boldsymbol{u}_i\}^{k+1}_{n+1}$ are obtained.
5. Steps 3 and 4 are repeated until equilibrium is attained in the body to the desired tolerance level. The variables have now all been updated to time t_{n+1}.

Steps 1–5 are repeated at each new time point from the seed to the completely grown and cooled crystal. At each time point the thermal loading (strains) in the crystal has to be updated based on the temperature field prevailing in the crystal. This temperature data can be obtained either from experiment or more realistically through computational fluid dynamics (CFD) and heat-transfer-based computational models of the entire high-pressure Czochralski process.

A second complication that has been so far ignored is the addition of new material during the growth process. This addition of virgin material can be modeled as the addition of a new layer of finite elements at certain time points. The frequency of the addition of these elements depends on the general length scale of the finite element grid. If the length scale of the grid is say 'l' mm then a new layer of finite elements, with average dimension 'l' mm, can be added at time intervals corresponding to growth of 'l' mm of material. Figure 39.15 shows a layer of new elements that has just been added to the original grid. This new layer of elements corresponds to the material that has just crystallized. Now a relevant question is: what are the values of the plastic strains $\boldsymbol{\varepsilon}^{\mathrm{pl}}_{ij}$ and the dislocation densities $N^{\mathrm{i}}_{\mathrm{s}}$ in the newly crystallized material. Since the material has just crystallized, the obvious answer would be to set both the plastic strain components as well as the dislocation densities to zero. However, there are a couple of problems with this approach.

Firstly, setting the initial dislocation densities to zero would ensure that this material is always dislocation free, as the dislocation density evolution rate would always be zero according to (39.36). To resolve this, the approach of *Völkl* [39.40] is followed and the initial dis-

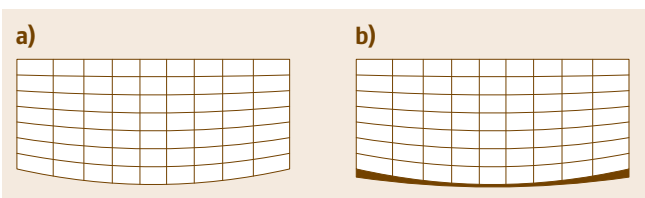


Fig. 39.15 **(a)** Two-dimensional projection of the grid system in the seed. **(b)** The grid system after 2 mm of growth. Shaded region represents the newly pulled material

location density is set to some finite but small positive value.

Secondly, setting the plastic strains to zero would give rise to huge stresses in the just-crystallized material, the reason being that nonfreshly crystallized (parent) parts of the crystal have significant plastic deformation, and the nodes on the interface between the two parts have significant displacement vectors to reflect this plastic strain. These displacements on the interface translate to large strains in the freshly crystallized material and large stresses, since this material behaves elastically due to the lack of plastic strain. This phenomenon as explained below is just a technicality. Before the crystallization of the new layer of elements, the present interface between the freshly crystallized material and the parent crystal was the crystal–melt interface, which is traction free. Hence the new layer of elements crystallize on a traction-free surface. Thus no stress from the parent crystal can propagate across this surface into the freshly crystallized material. Hence if the thin layer of freshly crystallized elements is isothermal (which is a reasonable assumption) then the new layer of elements representing the freshly crystallized material should be stress free. So what leads to the high stresses described before? The answer is that the reference state corresponding to zero plastic strain for the freshly crystallized material is different from that for the parent crystal. Hence, in order to use the same reference state for the parent crystal as well as the freshly crystallized material, some plastic strain has to be added to the freshly crystallized material.

In conclusion, the freshly crystallized elements are added in a stress-free state. To accomplish this the plastic strains at the Gaussian integration points of the freshly added elements are set such that the stresses at these Gaussian integration points are zero. This approach has been used in all the results presented in Sect. 39.10.

39.10 Numerical Results

To be of any use, the viscoplastic model needs the complete thermal history of the crystal during the growth. Such detailed information has not yet been obtained experimentally for any specific case, hence the results have been derived from the temperature field history predicted by a CFD-based model of the high-pressure Czochralski process termed MASTRAPP (multizone adaptive scheme for transport and phase change processes) [39.89]. MASTRAPP is a two-dimensional axisymmetric CFD-based algorithm applied to a simplified domain of the furnace. It simulates the temperature and flow field history during the growth process; it includes phenomena such as heat conduction in the crystal, thermal radiation, convection in the melt and gas, radiation, moving crystal–melt interface, and latent heat generation, and can handle presence of different domains such as the gas, crystal, melt, and encapsulant, and the movement of the boundaries of these different domains during the growth process. The following inputs are required in MASTRAPP as macroscopic parameters specifying the growth process.

39.10.1 Strength of Convection in the Melt and Gas

The fluid flow and temperature distributions are a result of natural convection due to buoyancy forces, and forced convection due to the rotation of the crystal and the crucible. The relevant nondimensional parameters for natural convection are the melt and gas Grashof numbers, and for forced convection are Re_s (the Reynolds number characterizing the crucible rotation) and Re_c (the Reynolds number characterizing the crystal rotation). For the Czochralski system shown in Fig. 39.2 with the diameter of the crucible being 100 mm and the height of the system – from the bottom of the crucible to the top of the heat shield – 170 mm, the Grashof numbers are of order 10^9, and the Reynolds numbers are of order 10^4. In papers dealing with numerical simulations of crystal growth, Grashof and Reynolds numbers an order of magnitude lower are used; the lower Grashof number can be justified by saying that they correspond to microgravity situations, and the lower Reynolds number to much lower revolution rates of the crystal and crucible than are typically used.

39.10.2 Temperature Boundary Condition

The reduced domain of Fig. 39.2 interacts with the other parts of the furnace through the specification of appropriate conditions at the boundaries of the reduced domain. Towards this the temperature on the crucible plus the heat shield is prescribed as a function of the distance from the base of the crucible in Fig. 39.2. The base of the crucible, and portions occupied by the melt, are maintained at 1435 K, and the temperature is lin-

early ramped down along the wall of the heat shield to a temperature of 427 K at the top. All the simulation results presented assume the boundary condition just described.

39.10.3 A Sample Case

A simulation of the growth of a flat-top 50 mm-diameter InP crystal from a height of 7–45 mm with a 20 mm height of encapsulation is used as a sample case. The gas and melt Grashof numbers are fixed at 10^6, and the rotational Reynolds numbers at 500. The growth axis is along the [001] direction. The results presented below correspond to the end of the growth period, before the cooling of the crystal to room temperature.

Temperature Field in the Crystal

The primary reason for the thermal stress and the consequent plastic deformation is the incompatibility of the thermal strain arising from the temperature field. The evolution of the axisymmetric temperature field and its gradients is shown in Figs. 39.16 and 39.17 respectively. At the beginning of the growth, when the seed has just thermally stabilized in the furnace, the isotherms show that heat enters the crystal from the sides and escapes from the top of the crystal. Hence the axial or vertical temperature gradient is higher in the upper parts of the crystal – of the order of 35 K/cm – than in the lower part. The radial temperature gradient is highest in the lower corners of the crystal, and is nearly zero in the central part of the crystal. As the crystal grows, the thickness of the encapsulant above the crystal decreases, thereby decreasing the thermal resistance to heat flow between the melt and the ambient gas; consequently the heat flux flowing through the crystal and the thermal gradients increase with growth.

The crystal just emerges out of the encapsulant at $t = 84$ min, and a direct pathway is established between

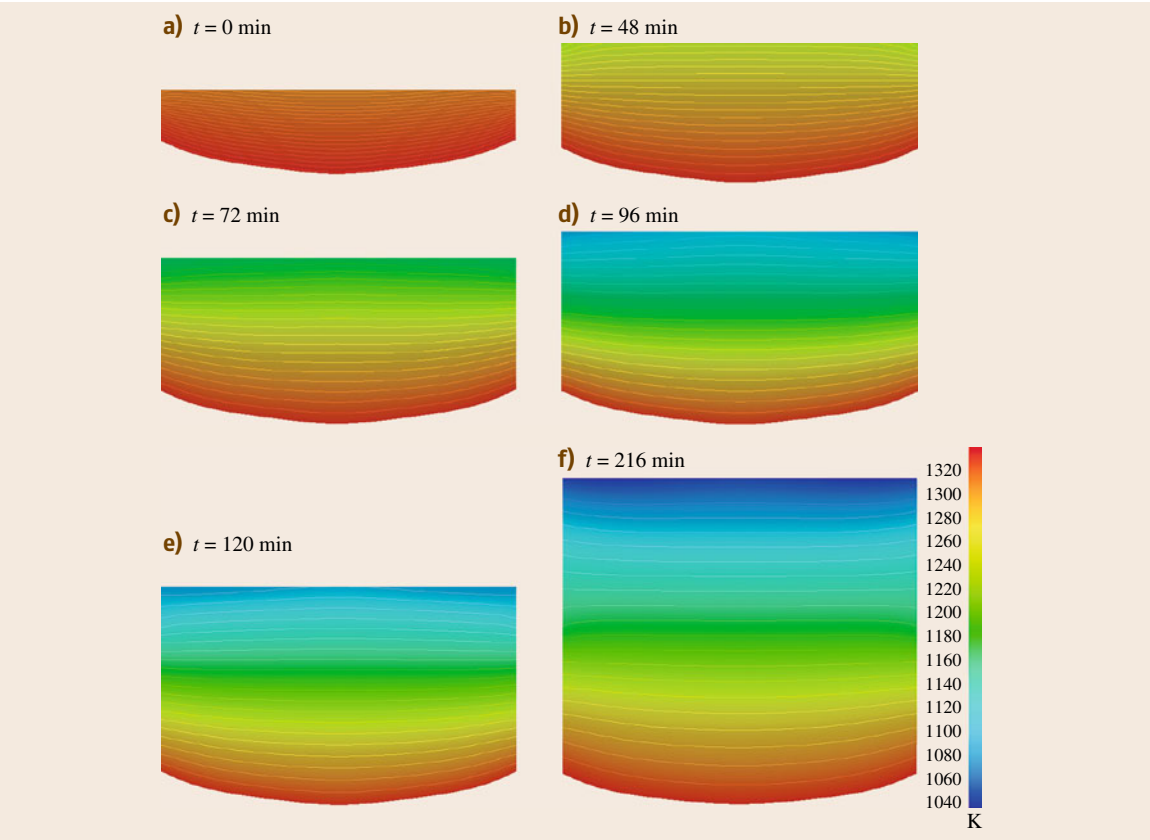


Fig. 39.16a–f Evolution of the temperature field in the growing crystal. The gas Grashof number is 10^6, the melt Grashof number is 10^6, the rotational Reynolds numbers are 500, and the encapsulant height is 20 mm

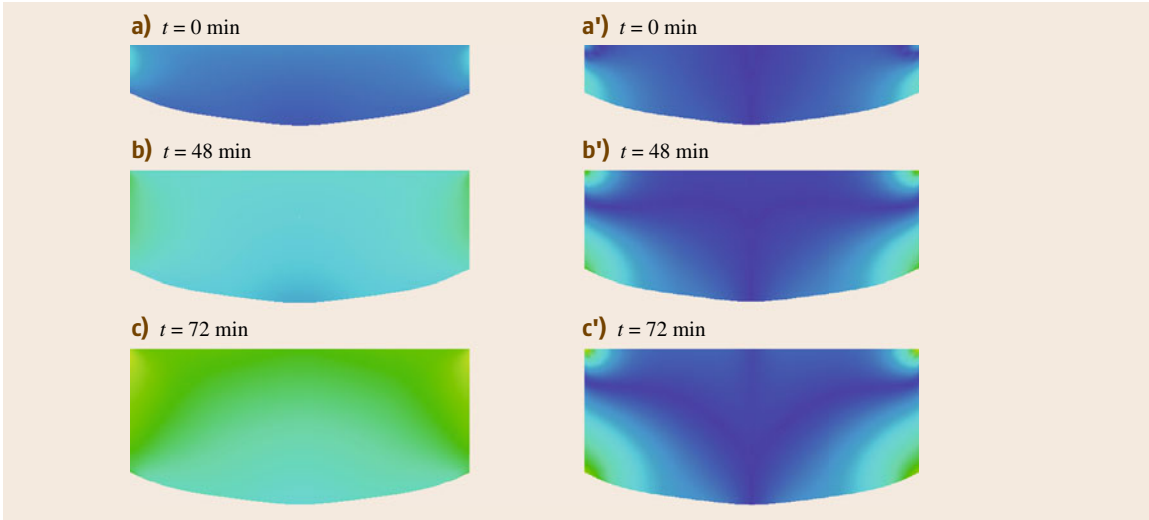


Fig. 39.17 History of thermal gradient field in the growing crystal. The gas Grashof number is 10^6, the melt Grashof number is 10^6, the rotational Reynolds numbers are 500, and the encapsulant height is 20 mm. Color scales for vertical and radial gradients are in **(f)** and **(f')**, respectively

the melt and the gas through the highly conducting crystal. Moreover, the lateral side of the part of the crystal above the encapsulant loses heat by convection and radiation to the cooler parts of the furnace, thus establishing a radially outward flow of heat, i. e., a radial temperature gradient. These give rise to a jump in the heat flux through the crystal. This is evident from comparing the frames at $t = 72$ min with the frames at $t = 96$ min in Fig. 39.17, when the top of the crystal is 1 mm below the encapsulant and 3 mm above the encapsulant respectively. In the latter frame the thermal gradients are appreciably higher. This change in the thermal profile and consequent high temperature gradients on emergence of the crystal from the encapsulant can be termed *thermal shock*.

As the crystal grows further, the length scale increases, while the temperature difference driving the heat flow between the melt and the gas remains the same; consequently vertical and radial temperature gradients decrease. The portion of the lateral surface of the crystal just above the encapsulant is hotter than the sidewalls of the furnace and loses heat by radiation and convection, leading to high radial temperature gradients in that region. Conversely, portions of the lateral surface of the crystal that are nearer to the top of the crystal are cooler than the side walls of the furnace and hence gain heat due to radiation from the sides of the furnace. *Zou* et al. report a similar behavior [39.22]. The radial temperature gradient is almost zero in most of the crystal except on the lateral surface at three specific locations: the bottom corner of the crystal near the interface due to a radially inward flow of heat from the melt into the crystal, just above the encapsulant due to cooling by convection and radiation, and finally the top of the crystal due to heating of the lateral surface by radiation.

Summarizing, the thermal history of the crystal can be divided into three stages:

1. Low thermal gradients, when the crystal is below the encapsulant and the gradients increase with growth
2. The thermal shock, when the crystal grows out from beneath the encapsulant, resulting in a jump in the thermal gradients
3. The later stages of the growth, characterized by decreasing temperature gradients and practically zero radial gradients.

Except for in three regions on the lateral surface of the crystal: one near the interface, the other just above the encapsulant, and the third at the top of the crystal (Fig. 39.17f). The crystal experiences maximum temperature gradients when it has just emerged from the encapsulant. The vertical temperature gradients obtained from the simulation, in the range of 50–100 K/cm, are of the same order as those reported in experiments with both GaAs [39.90] and InP [39.91].

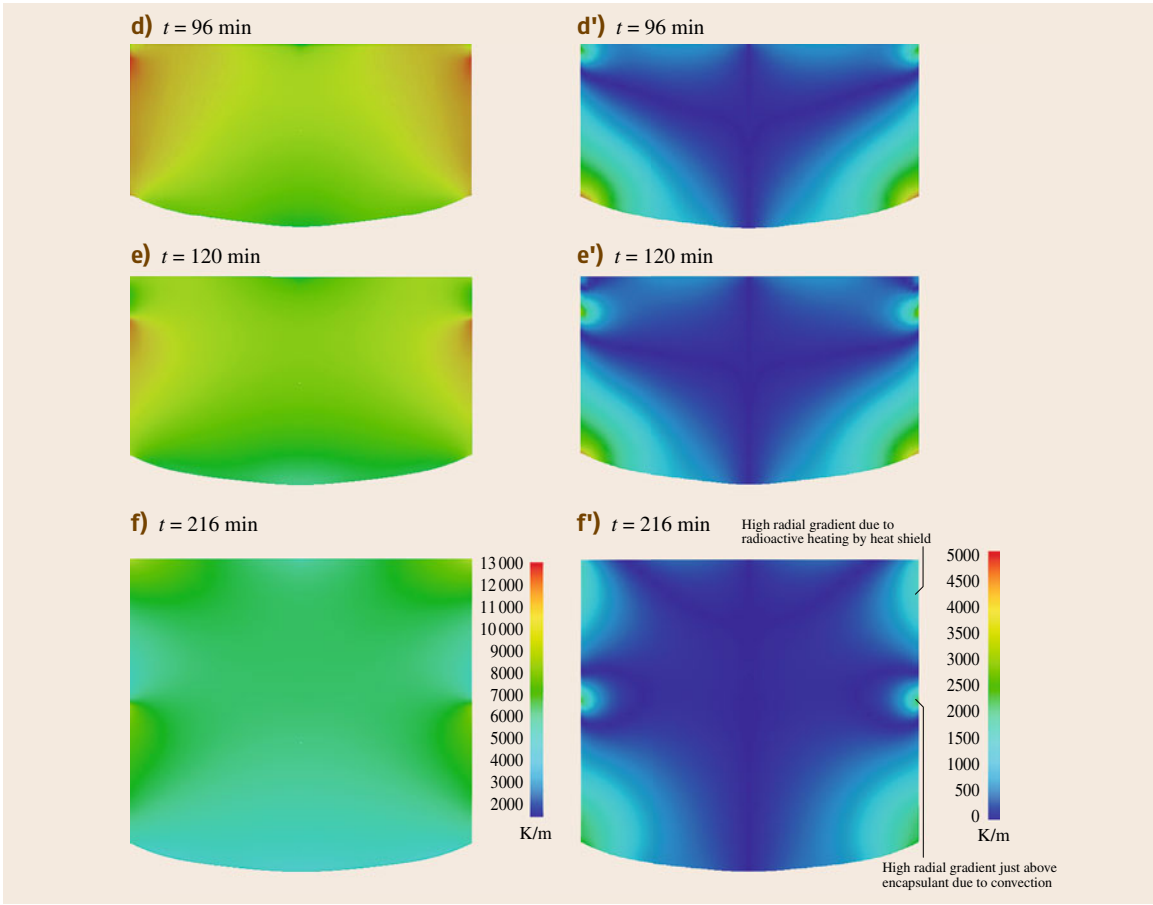


Fig. 39.17 (cont.)

Dislocation Densities after Growth Period

Figure 39.18 shows a fully grown crystal at the end of the growth period, with the colors representing the total dislocation density. The order of magnitude of the highest total dislocation density is around $1 \times 10^7\ \mathrm{m}^{-2}$ and occurs near the top of the crystal. For a closer inspection of the total dislocation density, Fig. 39.19 shows the distribution on (001) slices, taken at various heights of the crystal, the position of each being indicated in Fig. 39.18. On all the slices, except slice h, the dislocation density is highest near the circumference and lower towards the center. On slices a, f, and g (not shown) the dislocation density variation corresponds to the w-shaped variation mentioned often in

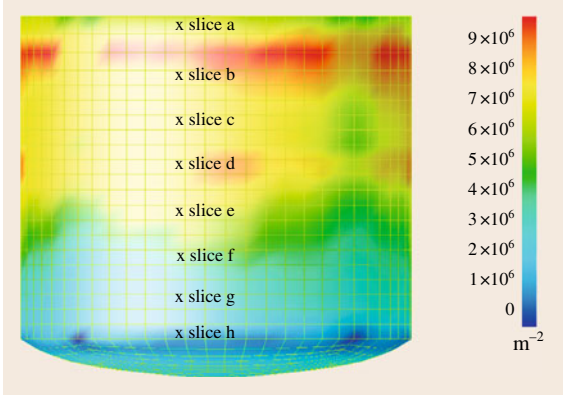


Fig. 39.18 Total dislocation density at the end of the growth period of a [001]-grown crystal. The vertical position of the various horizontal slices is indicated. The gas Grashof number is 10^6, the melt Grashof number is 10^6, the rotational Reynolds numbers are 500, and the encapsulant height is 20 mm ▶

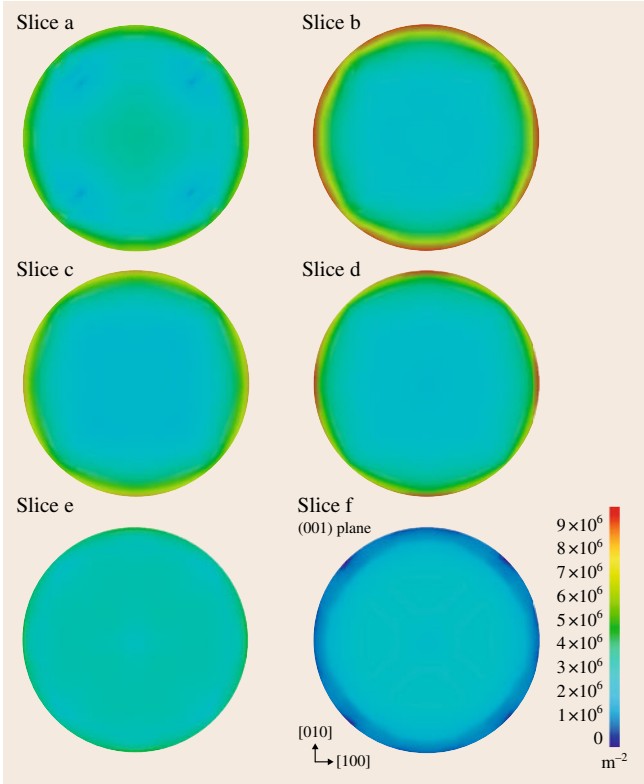


Fig. 39.19 Total dislocation densities at the end of the growth period projected on the (001) slices. The position of the various slices was indicated in Fig. 39.18. The gas Grashof number is 10^6, the melt Grashof number is 10^6, the Reynolds numbers are 500, and the encapsulant height is 20 mm

literature [39.12, 90, 92] – the dislocation density is highest on the circumference, initially decreases as one moves radially towards the center, and then increases. The dislocation density patterns on all the slices exhibit fourfold symmetry, and reflect the symmetry of the slip systems with respect to the [001] growth axis. All the slices, except slice h, exhibit distinct areas of low dislocation density aligned along the ⟨110⟩ directions and areas of highest density along the ⟨100⟩ directions, consistent with etch-pit observations [39.12, 93]. The values for total dislocation density are in the range 2×10^6–$1\times10^7\ \mathrm{m}^{-2}$. It should be pointed out that the values of dislocation density are not those in the final crystal but only the fully grown crystal at the end of the growth period, with the cooling-down period still remaining. Figure 39.20 shows a KOH-etched LEC-grown GaAs(001) wafer from the work of *Buchheit* et al. [39.94]. The etch pit configuration in this figure is

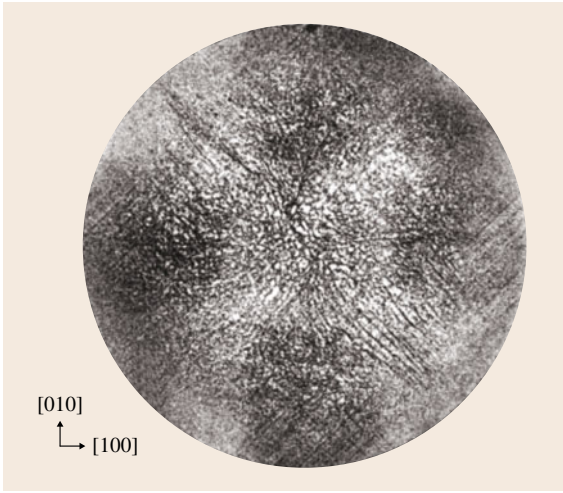


Fig. 39.20 KOH-etched (001) slice of GaAs (after [39.94], with permission)

similar to that of slice a in Fig. 39.19. A similar configuration can be seen in the work of *Chung* et al. [39.57,58] from x-ray topographs of LEC-grown (001)InP wafers.

Distribution of Dislocation Density on the Individual Slip Systems

Due to the symmetry of the [001] growth axis [39.95, Fig. 1], the dislocation density distributions of slip systems 1, 2, 5, 6, 7, 9, 10, and 11 on a (001) cross-section are inherently the same and can be transformed from one to the other by simple operations such as reflection and rotation by some angle. Similarly, by symmetry considerations the dislocation density distributions of slip systems 3, 4, 8, and 12, on any (001) cross-section are identical except that they are rotated with respect to each other by multiples of 90°. Hence, the dislocation density on slip systems 1 and 3 is presented, as each is representative of the other systems with which it shares symmetry.

Figure 39.21a shows the density of dislocations belonging to slip system 1, projected onto slice b. Notice the location of the maximum on the circumference. It was found by inspection that the maximum occurs on the circumference and at the same angular orientation in a majority of the slices. Similarly, Fig. 39.21b shows the density of dislocations belonging to slip system 3 on slice b. In this case too, the dislocation density maxima always occur on the circumference and at the same angular locations for most of the slices. The angular orientations of the maxima might reveal the dominant stress component responsible for the multiplication of

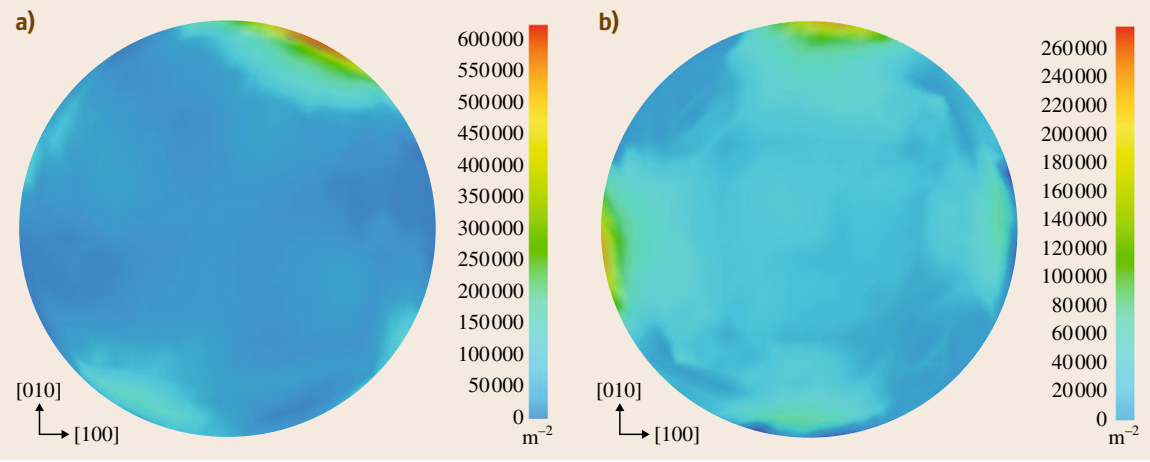


Fig. 39.21a,b Dislocation densities of slip systems 1 and 3 on slice b. The slip plane normal and slip direction for system 1 are $(11\bar{1})$ and [101], and for system 3 are $(11\bar{1})$ and $[1\bar{1}0]$. **(a)** Density of dislocations belonging to slip system 1. **(b)** Density of dislocations belonging to slip system 3

dislocation densities. Since the temperature field is axisymmetric, an assumption of isotropy (for the purpose of analysis) would imply that the elastic strain field driving the deformation would be axisymmetric, and it would be appropriate to think in terms of axisymmetric stress components in cylindrical coordinates – σ_{rr}, $\sigma_{r\theta}$ σ_{rz}, $\sigma_{\theta\theta}$, $\sigma_{\theta z}$, and σ_{zz}. A useful exercise would be to isolate the effects of each of these components, by setting one of them to unity and the rest to zero, and plot the variation of the resolved shear stress on systems 1 and 3 with the polar angle. This might help explain the angular orientations of the location of the maxima for dislocation densities of slip systems 1 and 3. The results for slip system 1 are shown in Fig. 39.22a, while those for slip system 3 are shown in Fig. 39.22b. In both these figures $\sigma_{\theta\theta}$ is set to 1 and the rest to 0. A comparison of these figures with the distribution of dislocation densities on the (001) slice shown in Fig. 39.21a,b shows that the circumferential component $\sigma_{\theta\theta}$ dominates the deformation at least near the outer circumference of the slice. The reason for this is that the maxima in the resolved shear components on slip systems 1 and 3 due to $\sigma_{\theta\theta}$ occur at the same polar angles, where the dislocation density on these slip systems is high. The exercise was repeated with components other than $\sigma_{\theta\theta}$, but the polar angles of the maxima did not correspond to the polar angles of dislocation density maxima in those cases.

The deformation is multislip in some regions of the crystal. For example, consider the maximum in total dislocation density present on the circumference of slice b in the [010] direction (Fig. 39.19). In this region slip systems 1, 3, 4, and 6 are almost equally active and their dislocation densities are found to be of the order of $2.5\times10^6\,\mathrm{m}^{-2}$ (Fig. 39.21a), $2\times10^6\,\mathrm{m}^{-2}$ (Fig. 39.21b), $2\times10^6\,\mathrm{m}^{-2}$, and $2.5\times10^6\,\mathrm{m}^{-2}$, respectively. By symmetry, the other three maxima in total dislocation density present on the circumference are also composed of four, almost equally active slip systems. Similarly, multislip activity is present in the central parts of slice b. The total dislocation density here (Fig. 39.19) is of the order of $2.5\times10^6\,\mathrm{m}^{-2}$. Dislocations belonging to each of the slip systems 1, 2, 5, 6, 7, 9, 10, and 11 have an approximate density of $2\times10^5\,\mathrm{m}^{-2}$ and those belonging to each of slip systems 3, 4, 8, and 12 have a density of almost $2.5\times10^5\,\mathrm{m}^{-2}$. Thus the total dislocation den-

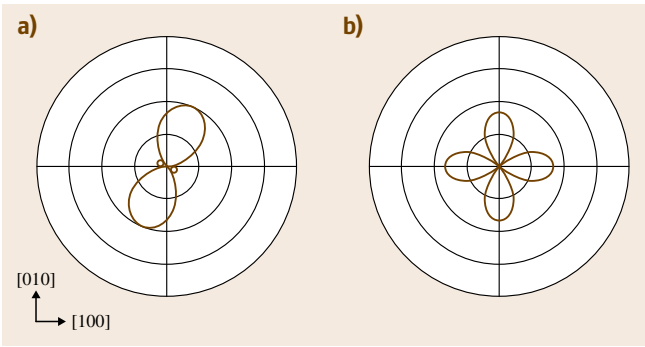


Fig. 39.22 Polar plot of resolved shear stress on slip systems 1 **(a)** and 3 **(b)**, with circumferential stress component set to 1 and all others to 0. The slip-plane normal and slip direction for system 1 are $(11\bar{1})$ and [101], and for system 3 are $(11\bar{1})$ and $[1\bar{1}0]$.

sity in the central parts of slice b is due to almost equal activity on all the slip systems.

39.10.4 Effect of Gas Convection and Radiation

The effect of radiative interaction of the lateral walls of the crystal with the furnace walls on the dislocation generation was isolated by artificially setting the radiative term to zero. The final total dislocation density field was observed to be practically the same as the case without radiation.

The effect of gas convection was investigated by setting the gas Grashof number to zero, while retaining radiation. Figure 39.23a shows the total dislocation density projected on a (010) surface after the end of the growth for this simulation. The total dislocation density in the case without gas convection is less than that in the case with gas convection – Grashof number 10^6 (Fig. 39.23b) – by more than a factor of 2. Thus a significant observation that can be made here is that radiation interaction of the sides of the crystal with the furnace does not significantly affect the dislocation density, while gas convection does.

Figure 39.23c shows a projection of the total dislocation density on the (010) plane at the end of the growth period (216 min) for a gas Grashof number of 10^7. The maximum dislocation density is now in the range of $3\times10^7\,\mathrm{m}^{-2}$. Similarly Fig. 39.23d shows the total dislocation density for the case of Grashof number 10^8. It is interesting to note that the maximum dislocation density doubled from 4×10^6 to $9\times10^6\,\mathrm{m}^{-2}$ for the case between no gas convection and Grashof number 10^6, while a further tenfold increase in the gas Grashof number to 10^7 tripled the maximum dislocation density to $3\times10^7\,\mathrm{m}^{-2}$, and another tenfold increase of the Grashof number to 10^8 further increased the maximum dislocation density to the range of $2\times10^8\,\mathrm{m}^{-2}$. This confirms that gas convection plays a significant role in the evolution of the dislocation densities.

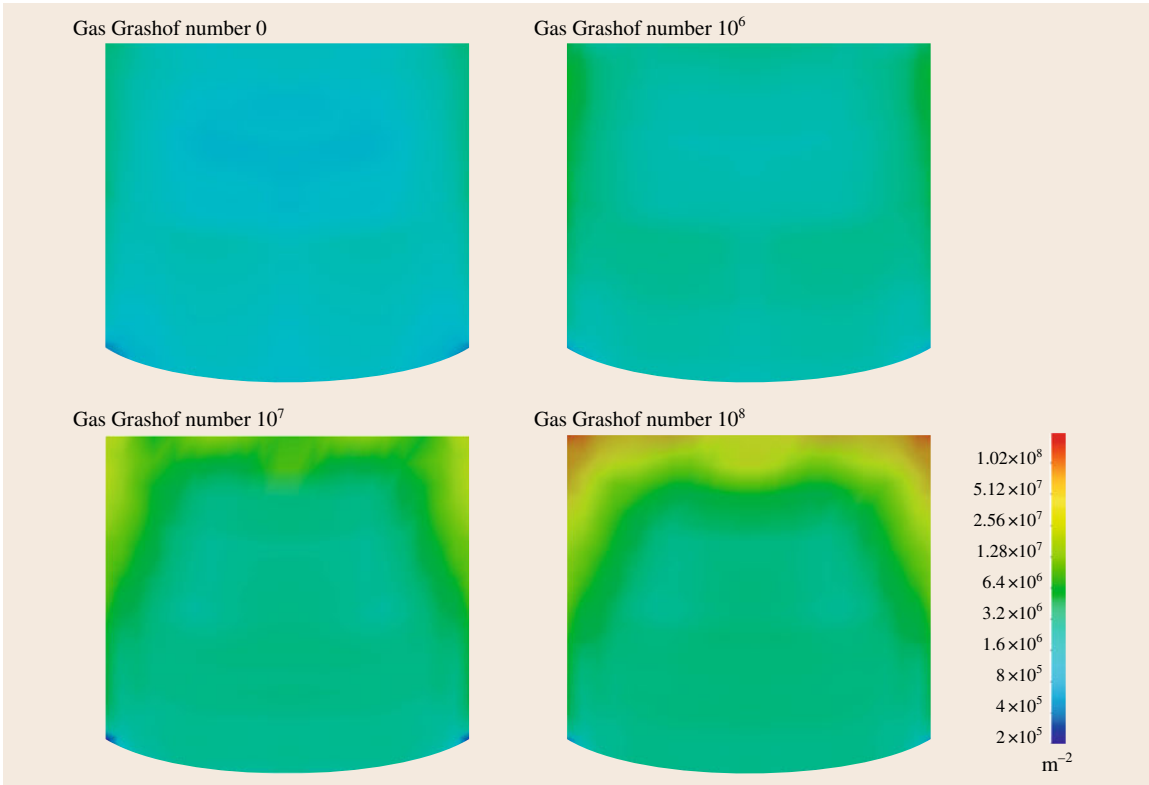


Fig. 39.23 Projection on the (010) plane passing through the growth axis, of the total dislocation density at the end of the growth period for different strengths of gas convection in the furnace. The melt Grashof number is 10^6, the Reynolds numbers are 500, and the encapsulant height is 20 mm in all cases

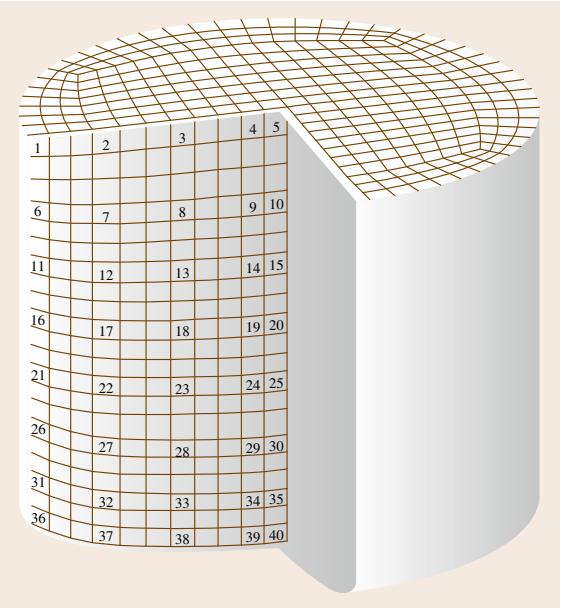


Fig. 39.24 Location of the various sample points, the history of which will be followed

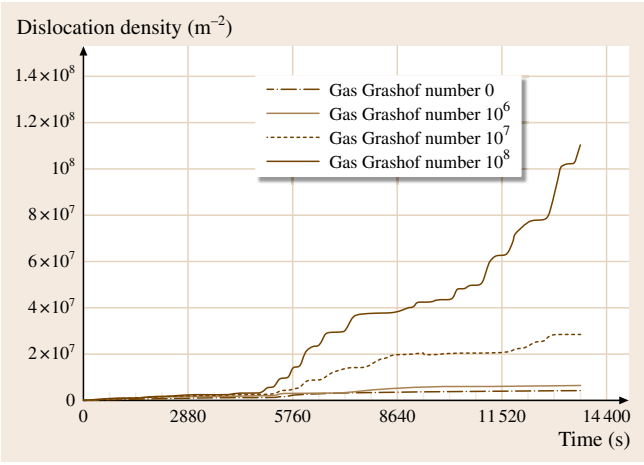


Fig. 39.25 Total dislocation density evolution for point 1, under different gas conditions. The melt Grashof number is 10^6, the rotational Reynolds numbers are 500, and the encapsulant height is 20 mm

The results for the total dislocation density at the end of the growth period for all the cases presented so far (growth with encapsulant height 20 mm with gas convection Grashof numbers of 0, 10^6 with radiation from the sides, 10^7, and 10^8) are now analyzed on the grid of points 1–40 located as shown in Fig. 39.24. The total dislocation density evolution history at point 1, for the cases of no gas convection and for gas Grashof numbers 10^6, 10^7, and 10^8, is presented in Fig. 39.25. The difference in dislocation density evolution history between the four cases again brings out the importance of gas convection. The dislocation density at point 1 is very low until a time of about 5500 s, and starts rising rapidly after this period. This timeframe corresponds to the period when the top of the crystal emerges from beneath the encapsulant (thermal shock).

39.10.5 Melt Convection and Rotation Reynolds Numbers

Melt Convection

In the calculations presented above, the Grashof number for melt convection was set at 10^6, while in real life it is closer to 10^8. Stronger melt convection might lead to increased heat flow through the crystal, and also affect the curvature of the crystal–melt interface. *Lambropoulos* [39.15], *Lambropoulos* and *Wu* [39.53], and *Chen* et al. [39.92] have proposed that the interface curvature will have a significant effect on the evolution of the total dislocation density.

The Grashof number in the melt was increased to 10^7 and then 10^8 to study the effect on the total dislocation density. In both of these simulations, the gas Grashof number was kept at 10^8. It was found that, although the interface becomes highly convex towards the melt for a melt Grashof number of 10^8, the final dislocation density does not change appreciably from the corresponding simulation when the melt Grashof number was 10^6. This result differs from observations in literature, which suggest that a curved interface is dangerous since it leads to high stresses – the CRSS at the interface is very low and high stresses there will lead to a high dislocation density. The situation in these simulations is however more subtle. The high temperature (near the melting point) at the interface ensures that the strain (or stress) generated due to the curved interface is quickly dissipated as plastic strain due to the high velocity of the dislocations, and the resulting additional dislocation density due to the enhanced curvature is not high enough to show up in the final results at the end of the growth period, since later in its history the same material element goes through more strenuous loading at lower temperatures, thus masking the effect of the additional interface curvature.

Rotation of Crucible and Crystal

Another important factor might be the rotation of the crucible and the crystal. These are usually rotated in opposite senses, and typically the angular velocities of rotation are of the order of 5–10 rpm, which results in Re_c and Re_s of the order of 8000–16 000. The rotation of the crucible creates a swirl component in the fluid motion, and as a result centrifugal and coriolis forces arise. The rotation of the crystal creates another forced convection term, and its effect on gas convection is akin to an Ekman flow above a rotating disk. The forces due to natural convection, crucible swirl, and crystal rotation

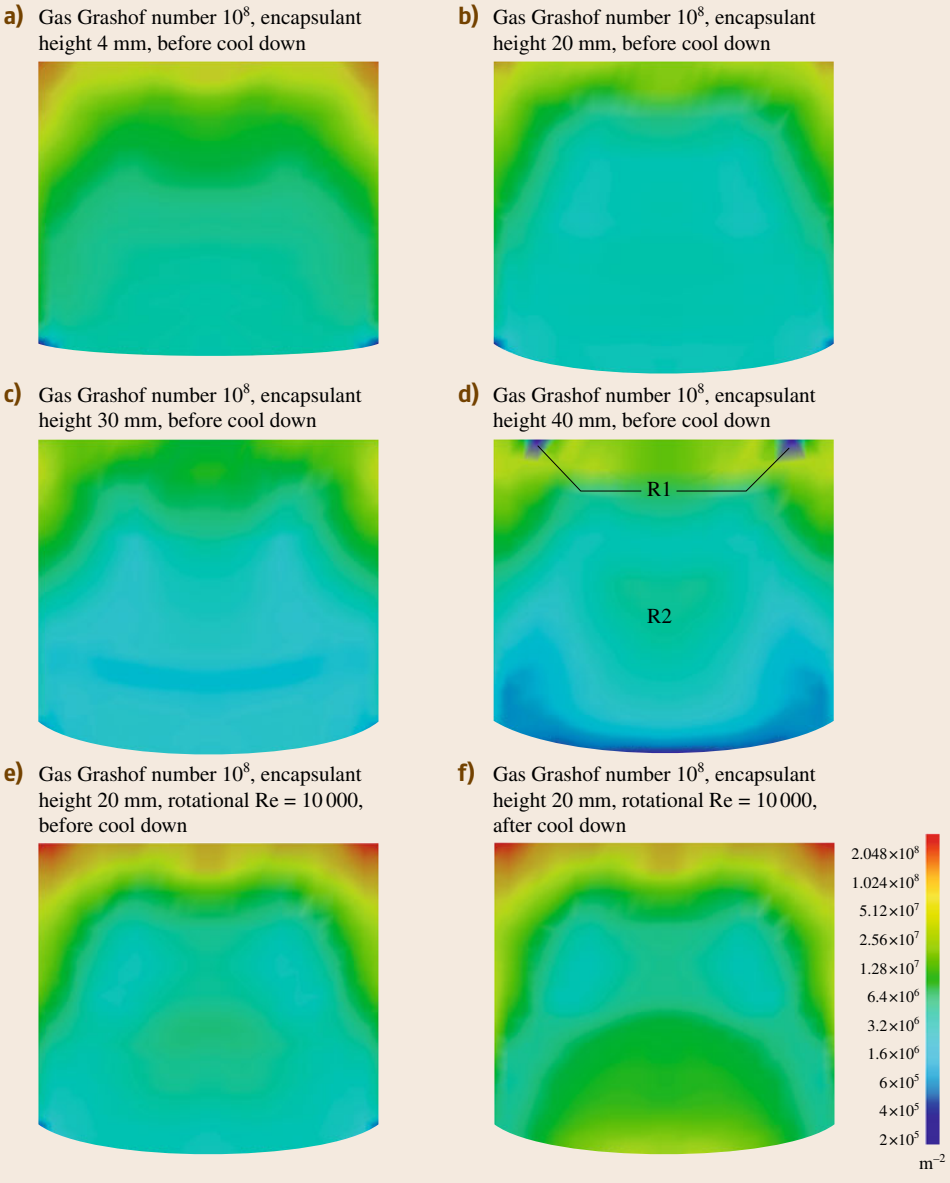


Fig. 39.26a–f Total dislocation densities projected on (010) plane passing through the growth axis. In all cases the melt Grashof Number is 10^6, and the rotational Reynolds numbers are 500 unless otherwise indicated

create a complicated pattern of gas flow, which might be oscillatory and three dimensional in nature.

An increase of the rotation Reynolds number to 2500 and 5000 did not cause much difference in the total dislocation density from the corresponding case with rotation Reynolds number of 500. However, a further increase in the rotation Reynolds number to 10 000 caused a doubling of the maximum total dislocation density to $4\times10^8\,\mathrm{m}^{-2}$, as can be seen in Fig. 39.26e. This might be consistent with the observations that at higher rotation rates the flow becomes unstable and promotes temperature and flow oscillations in the gas phase [39.96].

39.10.6 Control of Encapsulation Height

The height of encapsulation is expected to influence the total dislocation density by changing the temperature history. It will determine the instant at which the crystal will come out of the encapsulant and experience *thermal shock*, the duration of time for which the crystal is exposed to the hot gas, and the length of the crystal that is exposed to the gas. Figure 39.26a–d shows the total dislocation density projected on the (010) plane in the fully grown crystal for four different encapsulant heights (4, 20, 30, and 40 mm) with gas Grashof number of 10^8, melt Grashof number of 10^6, and rotational Reynolds numbers of 500. Though not shown in Fig. 39.26, a case for encapsulant height 50 mm under the same conditions was also simulated. This case corresponds to fully encapsulated growth, with the crystal remaining under the encapsulant throughout. The dislocation density in this case was one order of magnitude lower – less than $5\times10^6\,\mathrm{m}^{-2}$. *Elliot* et al. [39.97] report very low dislocation densities for fully encapsulated growth. The 50 mm case is to be contrasted with the case of 40 mm encapsulant height. In this case too, the crystal is under the encapsulant for 204 of the 216 min of its entire growth history, and at the end of its growth period, only the top 3 mm of the 43 mm long crystal is exposed to the gas convection. However, the dislocation levels at the top in this case are an order of magnitude higher than in the fully encapsulated case. This is further proof of the dominant role played by exposure of the crystal to gas convection in dislocation multiplication, at least as predicted by the modeling effort.

The highest dislocation density occurs in the case of encapsulant height 4 mm. The maximum occurs in the top corner of the crystal, and is of the order of $2\times10^8\,\mathrm{m}^{-2}$ in the case of 4 mm encapsulant, and decreases to around $8\times10^7\,\mathrm{m}^{-2}$ in the case of encapsulant height of 40 mm. In every case, the crystal seems to be clearly demarcated into two different zones: an upper zone where the total dislocation density is of the order of 3×10^7–$6\times10^7\,\mathrm{m}^{-2}$, and a lower region where the dislocation density is of the order of 5×10^6–$1\times10^7\,\mathrm{m}^{-2}$. The dislocation density tends to be higher towards the edges of the crystal. The area of high dislocation density is greatest in the case of encapsulant height of 4 mm.

39.10.7 The Cool-Down Period

The results presented so far are those at the end of the growth process. After the growth is the process of cooling down the fully grown crystal to room temperature. The temperature field in the crystal during the cool-down period is a function of the environment in the furnace, the rate at which the heaters in the furnace are ramped down, and other such details. These features are not available in the MASTRAPP CFD model, for a realistic simulation of the cool-down period. Therefore a reasonable approximation to the cool-down process is to ramp the temperature field in the crystal at the end of the growth period down to room temperature over a length of time. What this means is that the temperature gradients and thermal loads on the fully grown crystal are ramped down to zero.

Consequently, as the thermal loading is withdrawn, stresses develop due to the incompatibility of the plastic strain, leading to additional plastic deformation and dislocation multiplication, and finally considerable residual stresses in the cooled-down crystal. An important issue is the length of time over which the temperature ramp-down should take place. In a real Czochralski furnace, the crystal ingot is cooled down to room temperature over a few hours. Integrating the viscoplastic equations over a few hours of real time is beyond the computational resources available. Hence, the strategy adopted was to ramp down the thermal fields to room temperature over 12 min. In certain cases, we used three times this period to check for differences in the results. A clear disadvantage of the shortness of the ramp-down period is that the exaggerated unloading rate might lead to an overestimation of the final total dislocation density and residual stresses.

Figure 39.26f shows the final dislocation density after the cool-down period on a (010) slice containing the centerline, for the case with encapsulant height of 20 mm, gas Grashof number of 10^8, melt Grashof number of 10^6, and rotational Reynolds number of 10 000. To contrast with the situation the dislocation density be-

fore the cool-down period is presented in Fig. 39.26e. It can be seen that the dislocation density in the lowest part of the crystal has doubled or even tripled during the cool-down period. In fact the highest dislocation density in the crystal now occurs in the bottom and top parts of the crystal, with a distinct low-dislocation-density region in the middle part of the crystal. This seems to coincide with the experimental observation by *Chung* et al. [39.57, 58]. Though the figures are not displayed here, the cool-down calculations were performed for every case discussed previously. The results have a common feature: while the total dislocation density at the top and sides of the crystal develops during the growth period, the bottom of the crystal acquires a high dislocation density during the cool-down period. This phenomenon occurs due to the large temperature difference between the top and bottom of the crystal – roughly 300 K. Due to this, by the middle of the cooling period, the top portion of the crystal has cooled down sufficiently to be incapable of plastic deformation (since the threshold strain for plastic deformation is too high), and consequently generates considerable residual stresses. In contrast the bottom portion of the crystal is still hot (relative to the top portion) and is capable of plastic deformation. Since the crystal is a continuous medium, the residual stresses generated in the top part of the crystal generate a high loading in the lower part too, resulting in increased plastic deformation and dislocation generation there. Thus a new prediction by the modeling effort is that the dislocation density in the upper portions of the crystal develops during the growth period, while that in the lower part develops during the cool-down period. Though experimental data is scarce, this phenomenon has been reported by *Neubert* and *Rudolph* [39.98], in the context of vapor pressure controlled Czochralski (VCZ) growth of GaAs single crystals.

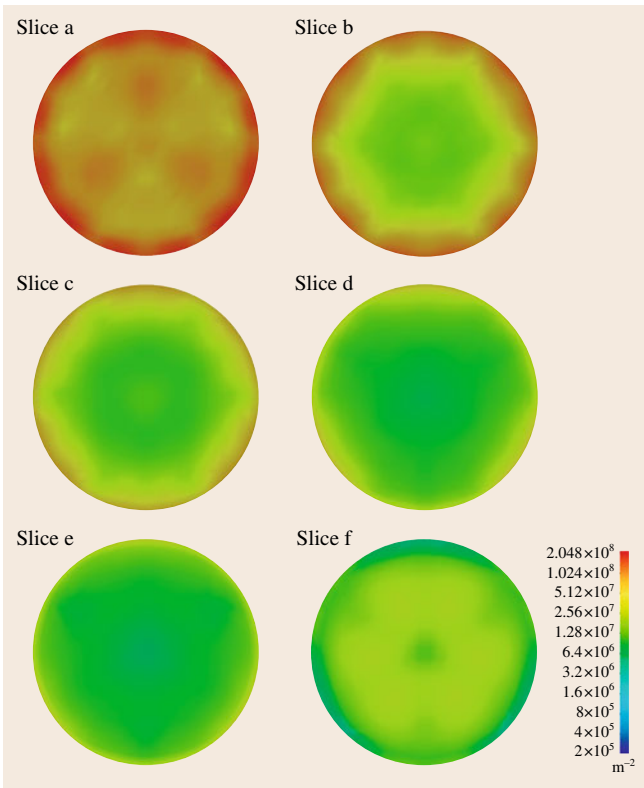


Fig. 39.27 Total dislocation density on $(1\bar{1}1)$ slices of a $[1\bar{1}1]$-grown crystal after the growth and cool-down period. The gas Grashof number is 10^8, the Reynolds numbers are 500, and the melt Grashof number is 10^6. The encapsulant height is 4 mm. The position of the slices is as indicated in Fig. 39.18

The cool-down period was tripled in the case of encapsulant height of 20 mm, gas Grashof number of 10^8, melt Grashof number of 10^6, and rotational Reynolds numbers of 500. A longer cool-down period means that parts of the crystal are at higher temperature for a longer time and capable of plastic deformation for a longer period. This gives rise to dislocation densities higher by a factor of 1.3 and slightly lower residual stresses at the end of the cool-down period than before.

39.10.8 The $[1\bar{1}1]$ Growth Axis

All the cases presented so far have been for growth along the [001] axis. In this section, the case for growth along the $[1\bar{1}1]$ axis is presented. The main difference is in the geometry of the slip systems. While in the case of [001] growth, the slip systems have fourfold symmetry with respect to the growth axis, in the case of $[1\bar{1}1]$ growth, threefold symmetry is expected.

Only one case is simulated for the $[1\bar{1}1]$ growth axis, with gas Grashof number of 10^8, rotation Reynolds numbers of 500, melt Grashof number of 10^6, and encapsulant height of 4 mm. Figure 39.27 presents the total dislocation density after growth and cool down in slices a–h (see Fig. 39.18 for locations of slices a–h) for the $[1\bar{1}1]$ case. The total dislocation density profiles in slices d and e exhibit threefold symmetry, and those in slices a, b, and c exhibit sixfold symmetry. Both of these symmetries are common in etch pit density observations on slices [39.40, 99]. A direct comparison of the slices showed that the total dislocation density after cool-down is slightly less for the $[1\bar{1}1]$ growth com-

pared with the [001] case. The explanation for this is not clear.

39.10.9 Summary of the Calculations and Some Comparisons

A CFD model for the high-pressure Czochralski growth of *In P* single crystals – MASTRAPP – was linked with the ABAQUS software to integrate the viscoplastic constitutive equations developed in previous sections under the thermal loading prevalent during the growth period and to predict the final total dislocation density and stress. After the growth period, the crystal temperature was ramped down from the temperature at the end of the growth period to room temperature to simulate the *cool-down* period of the crystal. A study was then conducted to delineate the effect of various parameters. The conclusions are summarized below:

1. The strength of gas convection has the most significant effect on the total dislocation density. The total dislocation density is very low – $400\,\mathrm{cm}^{-2}$ – for a gas Grashof number of zero, i. e., with no convection. On increasing the gas Grashof number to 10^6, the total dislocation density approximately doubles, and a further increase of the gas Grashof number to 10^7 raises the maximum total dislocation density to the order of $2000–3000\,\mathrm{cm}^{-2}$; finally, an increase of the Grashof number to 10^8 raises the maximum total dislocation density to the order of $20\,000\,\mathrm{cm}^{-2}$. In a real furnace, the Grashof numbers are of the order of 10^9, and hence dislocation densities one order of magnitude higher can be expected, and are found in commercial wafers. In contrast, thermal radiation was found not to play a significant role in dislocation generation.
2. The phenomenon of *thermal shock* has been delineated. The thermal shock is captured clearly in both the thermal and elastic stress histories. The total dislocation density in the crystal jumps multifold after the thermal shock.
3. An important finding is that the high dislocation density at the top of the crystal is developed during the growth process after exposure to gas convection, and that at the bottom of the crystal during the cool down.
4. The dislocation density pattern on select slices for [001]-grown crystal exhibits fourfold symmetry, and the total dislocation is higher at the edges of the wafer. The dislocation density is generally highest along the ⟨010⟩ directions and lowest along the ⟨110⟩ directions, in agreement with experimental findings. Similarly, for $[1\bar{1}1]$-grown crystal, the dislocation density pattern shows sixfold and threefold symmetry.
5. The effect of the height of encapsulation on the total dislocation density at the end of the growth period has been studied. Encapsulation heights of 4, 20, 30, 40, and 50 mm are used with gas Grashof number of 10^8, melt Grashof number of 10^6, and rotational Reynolds numbers of 500. The last case corresponds to fully encapsulated growth and the total dislocation density developed is only of the order of $2\times10^6–4\times10^6\,\mathrm{m}^{-2}$. This confirms experimental findings of fully encapsulated growth producing low-dislocation-density crystals.
6. Raising the melt Grashof number did not significantly affect the total dislocation density; however, an increase in the rotation Reynolds numbers increased the dislocation density.

If there is a common theme to the message from these calculations, it is that gas convection plays the most important role in dislocation multiplication. How does this square up with experimental observations? We report a few experimental papers which attempt to reduce the dislocation densities by adopting the same strategies explored numerically in the preceding sections – the reduction of thermal gradients – primarily achieved through certain artifices such as improving the design of the heat insulators and heaters, adjustment of the thickness of the boric oxide layer, reduction of the gas convection by reducing the ambient gas pressure as much as possible and design of appropriate baffles and convection shields, and finally stabilization of the melt flows. *Von Neida* and *Jordan* [39.100] achieved low thermal gradients in the LEC growth of GaAs by both adjusting the position of the heating source and modifying the B_2O_3 level. They noticed a reduction in dislocation densities by a factor of 10. *Elliot* et al. [39.101] achieved dislocation densities of $5000\,\mathrm{cm}^{-2}$, in the LEC growth of GaAs, by reducing the thermal gradients to around 6 °C/cm. *Katagiri* et al. [39.102] used a heat shield above the heater and crucible to decrease the axial temperature gradient from 100 °C, in a conventional LEC method, to 40 °C. The crystals grown in these conditions showed large areas free of dislocations. *Hirano* and *Uchida* [39.103] and *Hirano* [39.104] developed a new thermal baffle to obtain low-dislocation-density S-doped InP crystal of 50 mm diameter. The thermal baffle was designed to suppress gas convection. However, there are serious

problems associated with growth under low-thermal-gradient conditions – crystal diameter instability and instabilities at the crystal–melt interface, which might give rise to twinning. This problem is particularly severe in InP growth, which is prone to twinning because of the low stacking fault energy. These problems have been overcome by applying a vertical magnetic field to suppress the flow in the melt and stabilize the crystal–melt interface [39.105–108]. Another approach is to shield the crystal completely from gas convection by utilizing fully encapsulated growth. *Kohda* et al. [39.109] combined fully encapsulated growth and a vertical magnetic field to obtain 50 mm-diameter completely dislocation-free semi-insulating GaAs crystals.

Thus there are close experimental precedents to the modeling results. A more detailed report of these modeling results can be found in [39.110]. With the rapid increase in computing power, and the greater availability of parallel clusters, a more detailed optimization study with better temperature histories – perhaps obtained from experiment or high-fidelity CFD models – has the potential to devise growth programs that completely minimize the dislocation densities in III–V as-grown crystals.

39.11 Summary

An effort has been made to present the state of the art in the theoretical and modeling aspects of dislocation generation in III–V compound semiconductors. Sections 39.1 and 39.2 set up and motivated the discussion. Section 39.3 discussed the need for the reduction of dislocation density, and traces the origin of these, the most important being the nucleation of dislocations from point defects. Section 39.4 described models for dislocation multiplication due to thermal stress; the popular Jordan model and also a more general and computationally intensive viscoplastic framework were introduced. Section 39.5 and beyond dealt with viscoplastic models in the literature, their main features, and their origins from microscopic dislocation dynamics. Section 39.5 gave an introduction to the structural aspects of common dislocations in diamond- and zincblende-type lattices. Section 39.6 discussed the deformation behavior of semiconductors and presents the Alexander–Haasen viscoplastic model. A brief description of the various components of the model is given. Section 39.7 gave a description of existing literature on the application of Alexander–Haasen-type models to the crystal growth problem. As a thorough examination of this class of models, Sect. 39.8 and the sections that follow discussed the development and application to crystal growth of a particular viscoplastic model. Section 39.8 described in great detail the various building blocks of the model, and the many limitations and assumptions that go into such models, as well as insights into future work that can be performed. Section 39.9 described briefly the numerical implementation of the model in a finite element framework and the methodology of its coupling with computational fluid dynamics and heat-transfer-based thermal models for crystal growth processes. Section 39.10 presented the results of one such successful coupling, and various conclusions are drawn.

Finally, some open questions and avenues for research are now presented:

1. A question frequently posed is the reason for the growth of dislocation-free silicon. In view of the above conclusion, a bold hypothesis may be the absence of substantial gas convection in low-pressure Czochralski growth of silicon. It might be useful to make a viscoplastic model for silicon and investigate it further.
2. A substantial improvement can be made in the thermal models. With the advent of high-performance computing, it might be possible to develop a continuous, nonsteady, three-dimensional simulation of the entire growth process, instead of the present two-dimensional axisymmetric quasistatic series of simulations. Such a simulation would resolve instabilities in the gas flow, and the temperature and flow oscillations could be resolved with more certainty.
3. The viscoplastic model suffers from various difficulties. Firstly, the topological aspect of dislocations was neglected. Dislocations are continuous line defects and can either end at the surfaces of the material or form closed loops. Physically, this is reflected in a divergence-free dislocation density tensor. Present formulations ignore the topological aspects of dislocation density and simply model it as scalar variables. Again, it might be possible to formulate alternate models, with dislocation as line objects rather than scalar variables. This might lead to theories with a structure similar to strain gradient theories.

4. Instead of assuming an initially negligible but finite total dislocation density, a suitable alternative would be to model the initial dislocation density from condensation of point defects into Frank loops, and their subsequent transformation into glide dislocations. Such a model should necessarily be coupled with models for point defect dynamics in the crystal.
5. There is scope for a synergetic effort between experimentalists – TEM, constant-strain-rate tests – and theoreticians. There is a great deal of uncertainty in measurements of dislocation velocities and stress–strain curves. Often these experiments are done on materials from different sources; this gives rise to great variations in the results. Furthermore, the virgin state of the material before the deformation, such as the initial dislocation density, is often not recorded. Hence a consistent set of experiments should be designed to measure quantities such as the dislocation density and stress–strain curves for material from the same set of sources. At the same time quantities such as the hardening coefficients and multiplication coefficients can be resolved with a mixture of atomistic quantum and mesoscopic discrete dislocation dynamics calculations. Such an effort will involve modeling on multiple scales along with the experimental data.

References

39.1 J. Czochralski: Ein neues Verfahren zur Messung der Kristallisationsgeschwindigkeit der Metalle, Z. Phys. Chem. **92**, 219–221 (1917), in German

39.2 G.K. Teal, J.B. Little: Growth of germanium single crystals, Phys. Rev. **78**, 647 (1950)

39.3 W.C. Dash: Dislocation free silicon crystals. In: *Growth and Perfection of Crystals*, ed. by R.M. Doremus, B.W. Roberts, D. Turnbull (Wiley, New York 1958)

39.4 V. Swaminathan, A.S. Jordan: Dislocations in III/V compounds, Semicond. Semimet. **38**, 293–341 (1993)

39.5 R.J. Roedel, A.R. Von Neida, R. Caruso, L.R. Dawson: The effect of dislocations in $Ga_{1-x}Al_xAs$:Si light-emitting diodes, J. Electrochem. Soc. **126**, 637–641 (1979)

39.6 J.P. Hirth, J. Lothe: *Theory of Dislocations* (Krieger, Malabar 1992)

39.7 H. Alexander: On dislocation generation in semiconductor crystals, Radiat. Eff. Defects Solids **112**(1/2), 1–12 (1989)

39.8 B.T. Lee, R. Gronsky, E.D. Bourret: Dislocation loops and precipitates associated with excess arsenic in GaAs, J. Appl. Phys. **64**(1), 114–118 (1988)

39.9 J. Lagowski, H.C. Gatos, T. Aoyama, D.G. Lin: Fermi energy control of vacancy coalescence and dislocation density in melt-grown GaAs, Appl. Phys. Lett. **45**(6), 680–682 (1984)

39.10 W. Zulehner: Czochralski growth of silicon, J. Cryst. Growth **65**(1–3), 189–213 (1983)

39.11 M.F. Ashby, L. Johnson: On the generation of dislocations at misfitting particles in a ductile matrix, Philos. Mag. **20**, 1009–1022 (1969)

39.12 A.S. Jordan, R. Caruso, A.R. Von Neida: A thermoelastic analysis of dislocation generation in pulled GaAs crystals, Bell Syst. Technol. J. **59**(4), 593–637 (1980)

39.13 N. Kobayashi, T. Iwaki: A thermoelastic analysis of the thermal stress produced in a semi-infinite cylindrical single crystal during the Czochralski growth, J. Cryst. Growth **73**, 96–110 (1985)

39.14 M. Duseaux: Temperature profile and thermal-stress calculations in GaAs crystals growing from the melt, J. Cryst. Growth **61**(3), 576–590 (1983)

39.15 J.C. Lambropoulos: Stresses near the solid-liquid interface during the growth of a Czochralski crystal, J. Cryst. Growth **80**, 245–256 (1987)

39.16 C.E. Schvezov, I.V. Samarasekera, F. Weinberg: Calculation of the shear stress distribution in LEC gallium arsenide for different growth conditions, J. Cryst. Growth **92**, 479–488 (1988)

39.17 G.O. Meduoye, K.E. Evans, D.J. Bacon: Modelling of the growth of the LEC technique II. Thermal stress distribution and influence of interface shape, J. Cryst. Growth **97**, 709–719 (1989)

39.18 G.O. Meduoye, D.J. Bacon, K.E. Evans: Computer modelling of temperature and stress distributions in LEC-grown GaAs crystals, J. Cryst. Growth **108**, 627–636 (1991)

39.19 S. Motakef, K.W. Kelly, K. Koai: Comparison of calculated and measured dislocation density in LEC-grown GaAs crystals, J. Cryst. Growth **113**, 279–288 (1991)

39.20 F. Dupret, P. Necodeme, Y. Ryckmans: Numerical method for reducing stress level in GaAs crystals, J. Cryst. Growth **97**, 162–172 (1989)

39.21 D.E. Bornside, T.A. Kinney, R.A. Brown: Minimization of thermoelastic stresses in Czochralski grown silicon: Application of the integrated system model, J. Cryst. Growth **108**, 779–805 (1991)

39.22 Y.F. Zou, H. Zhang, V. Prasad: Dynamics of melt-crystal interface and coupled convection-stress predictions for Czochralski crystal growth processes, J. Cryst. Growth **166**, 476–482 (1996)

39.23 I. Yonenaga, K. Sumino: Impurity effects on the generation, velocity, and immobilization of dislocations in GaAs, J. Appl. Phys. **65**, 85–92 (1989)

39.24 J. Lubliner: *Plasticity Theory* (Macmillan, New York 1990)

39.25 K. Sumino: Mechanical behavior of semiconductors. In: *Handbook on Semiconductors*, Vol. 3a, ed. by S. Mahajan, T.S. Moss (Elsevier, Amsterdam 1994) pp. 73–181

39.26 J. Hornstra: Dislocations in the diamond lattice, J. Phys. Chem. Solids **5**, 129–141 (1958)

39.27 H. Alexander: Dislocations in covalent crystals. In: *Dislocations in Solids*, Vol. 7, ed. by F.R.N. Nabarro (North-Holland, Amsterdam 1986) pp. 113–234

39.28 D.J.H. Cockayne, A. Hons: Dislocations in semiconductors as studied by weak-beam electron-microscopy, J. Phys. **40**(6), 11–18 (1979)

39.29 H. Gottschalk, G. Patzer, H. Alexander: Stacking-fault energy and ionicity of cubic III–V compounds, Phys. Status Solidi (a) **45**(1), 207–217 (1978)

39.30 R. Meingast, H. Alexander: Dissociated dislocations in germanium, Phys. Status Solidi (a) **17**(1), 229–236 (1973)

39.31 A. George, J. Rabier: Dislocations and plasticity in semiconductors. I – Dislocation structures and dynamics, Rev. Phys. Appl. **22**, 941–966 (1987)

39.32 W.G. Johnston, J.J. Gilman: Dislocation velocities, dislocation densities, and plastic flow in lithium fluoride crystals, J. Appl. Phys. **30**, 129–144 (1959)

39.33 H. Alexander, P. Haasen: Dislocations and plastic flow in the diamond structure. In: *Solid State Physics*, Vol. 22, ed. by F. Seitz, D. Turnbull, H. Ehrenreich (Academic, New York 1968) pp. 28–158

39.34 A.R. Chaudhuri, J.R. Patel, L.G. Rubin: Velocities and densities of dislocations in germanium and other semiconductor crystals, J. Appl. Phys. **33**, 2736–2746 (1962)

39.35 G.I. Taylor: The mechanism of plastic deformation of crystals. Part I – Theoretical, Proc. R. Soc. Lond. Ser. A **145**, 362–387 (1934)

39.36 F.R.N. Babarro, Z.S. Basinski, D.B. Holt: The plasticity of pure single crystals, Adv. Phys. **13**, 193–323 (1964)

39.37 E. Peissker, P. Haasen, H. Alexander: Anisotropic plastic deformation of indium antimonide, Philos. Mag. **7**, 1279 (1962)

39.38 I. Yonenaga, K. Sumino: Effects of in impurity on the dynamic behavior of dislocations in GaAs, J. Appl. Phys. **62**(4), 1212–1219 (1987)

39.39 I. Yonenaga, K. Sumino: Mechanical properties and dislocation dynamics of GaP, J. Mater. Res. **4**(2), 355–360 (1989)

39.40 J. Völkl: Stress in the cooling crystal. In: *Handbook of Crystal Growth*, Vol. 2, ed. by D.T.J. Hurle (North Holland, Amsterdam 1994) pp. 823–874

39.41 H. Siethoff, W. Schröter: Work-hardening and dynamical recovery in silicon and germanium at high-temperatures and comparison with FCC metals, Scr. Metall. **17**(3), 393–398 (1983)

39.42 H. Siethoff, R. Behrensmeier: Plasticity of undoped GaAs deformed under liquid encapsulation, J. Appl. Phys. **67**(8), 3673–3680 (1990)

39.43 H. Siethoff, K. Ahlborn, H.G. Brion, J. Völkl: Dynamical recovery and self-diffusion in InP, Philos. Mag. A **57**(2), 235–244 (1988)

39.44 H. Siethoff, W. Schröeter: New phenomena in the plasticity of semiconductors and FCC metals at high temperatures, Z. Metall. **75**(7), 475–491 (1984)

39.45 A.K. Mukherjee, J.E. Bird, J.E. Dorn: Experimental correlations for high temperature creep, ASM Transactions **62**, 155–179 (1969)

39.46 C.R. Barrett, W.D. Nix: A Model for steady state creep based on the motion of jogged screw dislocations, Acta Metall. **13**, 1247–1258 (1965)

39.47 H.G. Brion, H. Siethoff, W. Schröter: New stages in stress–strain curves of germanium at high-temperatures, Philos. Mag. A **43**(6), 1505–1513 (1981)

39.48 H. Siethoff: Cross-slip in the high-temperature deformation of germanium, silicon and indium-antimonide, Philos. Mag. A **47**(5), 657–669 (1983)

39.49 B. Escaig: Cross-slip processes in the fcc structure. In: *Dislocation Dynamics*, ed. by A.R. Rosenfield, R. Alan (McGraw-Hill, London 1968) pp. 655–677

39.50 D. Maroudas, R.A. Brown: On the prediction of dislocation formation in semiconductor crystals grown from the melt – Analysis of the Haasen model for plastic deformation dynamics, J. Cryst. Growth **108**, 399–415 (1991)

39.51 C.T. Tsai: On the finite-element modeling of dislocation dynamics during semiconductor-crystal growth, J. Cryst. Growth **113**, 499–507 (1991)

39.52 C.T. Tsai, A.N. Gulluoglu, C.S. Hertley: A crystallographic methodology for modeling dislocation dynamics in GaAs crystals grown from the melt, J. Appl. Phys. **73**, 1650–1656 (1993)

39.53 J.C. Lambropoulos, C.H. Wu: Mechanics of shaped crystal growth from the melt, J. Mater. Res. **11**, 2163–2176 (1996)

39.54 N. Miyazaki, Y. Kuroda: Dislocation density simulations for bulk single crystal growth process, Met. Mater. Int. **4**(4), 883–890 (1998)

39.55 J.C. Moosbrugger: Continuum slip viscoplasticity with the Haasen constitutive model – application to single-crystal inelasticity, Int. J. Plast. **11**, 799–826 (1995)

39.56 J.C. Moosbrugger, A. Levy: Constitutive modelling for CdTe single-crystals, Metall. Mater. Trans. A **26**(10), 2687–2697 (1995)

39.57 H. Chung, W. Si, M. Dudley, A. Anselmo, D.F. Bliss, A. Maniatty, H. Zhang, V. Prasad: Characterization of structural defects in MLEK grown InP single crystals using synchotron beam x-ray topography, J. Cryst. Growth **174**(1–4), 230–237 (1997)

39.58 H. Chung, W. Si, M. Dudley, D.F. Bliss, R. Kalan, A. Maniatty, H. Zhang, V. Prasad: Characterization

of defect structures in magnetic liquid encapsulated Kyropoulos grown InP single crystals, J. Cryst. Growth **181**(1–2), 17–25 (1997)
39.59 H. Steinhardt, P. Haasen: Creep and dislocation velocities in GaAs, Phys. Status Solidi (a) **49**, 93–101 (1978)
39.60 I. Yonenaga, U. Unose, K. Sumino: Mechanical properties of GaAs crystals, J. Mater. Res. **2**, 252–261 (1987)
39.61 A. George, C. Escaravage, G. Champier, W. Schröter: Velocities of screw and 60°-dislocations in silicon, Phys. Status Solidi (b) **53**, 483–496 (1972)
39.62 M. Imai, K. Sumino: Insitu x-ray topographic study of the dislocation mobility in high-purity and impurity-doped silicon-crystals, Philos. Mag. A **47**(4), 599–621 (1983)
39.63 B.Y. Farber, V.I. Nikitenko: Change of dislocation mobility characteristics in silicon single-crystals at elevated-temperatures, Phys. Status Solidi (a) **73**(1), K141–144 (1982)
39.64 I. Yonenaga, K. Sumino: Dislocation velocity in indium-phospide, Appl. Phys. Lett. **58**(1), 48–50 (1991)
39.65 H. Nagai: Dislocation velocities in indium phospide, Jpn. J. Appl. Phys. **20**(4), 793–794 (1981)
39.66 K. Maeda, S. Takeuchi: Recombination enhanced glide in InP single crystals, Appl. Phys. Lett. **42**(8), 664–666 (1983)
39.67 F. Louchet: On the mobility of dislocations in silicon by insitu straining in a high-voltage electron-microscope, Philos. Mag. **43**(5), 1289–1297 (1981)
39.68 V. Celli, M. Kabler, T. Ninoyama, R. Thomson: Theory of dislocation mobility in semiconductors, Phys. Rev. **131**(1), 58–72 (1963)
39.69 V.V. Rybin, A.N. Orlov: Theory of dislocation motion in low-velocity range, Sov. Phys. Solid State **11**, 2635–2641 (1970)
39.70 S. Öberg, P.K. Sitch, R. Jones, M.I. Heggie: First-principles calculations of the energy barrier to dislocation motion in Si and GaAs, Phys. Rev. B **51**(19), 13138–13145 (1995)
39.71 V.V. Bulatov, S. Yip, A.S. Argon: Atomic modes of dislocation mobility in silicon, Philos. Mag. A **72**(2), 453–496 (1995)
39.72 H.R. Kolar, J.C.H. Spencer, H. Alexander: Observation of moving dislocation kinks and unpinning, Phys. Rev. Lett. **77**(19), 4031–4034 (1996)
39.73 H.J. Möller: The movement of dissociated dislocations in the diamond–cubic structure, Acta Metall. **26**, 963–973 (1977)
39.74 P. Haasen: Kink formation in charged dislocation, Phys. Status Solidi (a) **28**(1), 145–155 (1975)
39.75 P.B. Hirsch: Mechanism for the effect of doping on dislocation mobility, J. Phys. **40**(6), 117–121 (1979)
39.76 K. Sumino, I. Yonenaga: Dislocation dynamics and mechanical behavior of elemental and compound semiconductors, Phys. Status Solidi (a) **138**, 573–581 (1993)
39.77 K. Sumino, H. Harada: In situ x-ray topographic studies of the generation and the multiplication processes of dislocations in silicon crystals at elevated temperature, Philos. Mag. A **44**(6), 1319–1334 (1981)
39.78 P. Franciosi, A. Zaoui: Multislip in fcc. crystals: A theoretical approach compared with experimental data, Acta Metall. **30**, 1627–1637 (1982)
39.79 A. Moulin, M. Condat, L.P. Kubin: Mesoscale modelling of the yield point properties of silicon crystals, Acta Metall. **47**(10), 2879–2888 (1999)
39.80 H. Alexander, J.J. Crawford: Latent hardening of germanium crystals, Phys. Status Solidi (b) **222**, 41–49 (2000)
39.81 A.A. Chernov: *Modern Crystallography III. Crystal Growth* (Springer, Berlin 1984)
39.82 H. Klapper: Generation and propagation of dislocations during crystal growth, Mater. Chem. Phys. **66**, 101–109 (2000)
39.83 G. Dhanaraj, B. Raghothamachar, J. Bai, H. Chung, M. Dudley: Synchrotron x-ray topographic characterization of defects in InP bulk crystals, Proc. Int. Conf. Indium Phosphide Relat. Mater. (2005) pp. 643–648
39.84 G.T. Brown, B. Cockayne, W.R. Macewan: Deformation behavior of single crystals of InP in uniaxial compression, J. Mater. Sci. **15**, 1469–1477 (1980)
39.85 S. Pendurti: Modeling Dislocation Generation in High Pressure Czochralski Growth of InP Single Crystals. Ph.D. Thesis (State University of New York, Stony Brook 2003)
39.86 A.S. Jordan: Some thermal and mechanical properties of InP essential to crystal growth modeling, J. Cryst. Growth **71**, 559–565 (1985)
39.87 H. Siethoff: The plasticity of elemental and compound semiconductors, Semicond. Semimet. **37**, 143–187 (1992)
39.88 J.C. Simo, T.J.R. Hughes: *Computational Inelasticity* (Springer, New York 1998)
39.89 H. Zhang, V. Prasad: A multizone adaptive process model for low and high pressure crystal growth, J. Cryst. Growth **155**, 47–65 (1995)
39.90 P. Rudolph, M. Jurisch: Bulk growth of GaAs – An overview, J. Cryst. Growth **199**(1), 325–335 (1999)
39.91 V.A. Antonov, V.G. Elsakov, T.I. Olkhovikova, V.V. Selin: Dislocations and 90°-twins in LEC-grown InP crystals, J. Cryst. Growth **235**(1–4), 35–39 (2002)
39.92 T.-C. Chen, H.-C. Wu, C.-I. Weng: The effect of interface shape on anisotropic thermal stress of bulk single crystal during Czochralski growth, J. Cryst. Growth **173**, 367–379 (1997)
39.93 J. Matsui: Study of strain variation in LEC-grown GaAs bulk crystals by synchotron radiation x-ray, Appl. Surf. Sci. **50**, 1–8 (1991)
39.94 H.M. Buchheit, A. Khoukh, M. Bejar, S.K. Krawczyk, R.C. Blanchet: Residual strain mapping in III–V materials by spectrally resolving scanning pho-

toluminescence, Microelectron. J. **30**(7), 651–657 (1999)
39.95 S. Pendurti, V. Prasad, H. Zhang: Modelling dislocation generation in high pressure Czochralski growth of InP single crystals: Part I. Construction of a visco-plastic deformation model, Model. Simul. Mater. Sci. Eng. **13**, 249–266 (2005)
39.96 V. Prasad, H. Zhang: Transport phenomena in Czochralski crystal growth processes, Adv. Heat Transf. **30**, 313–435 (1997)
39.97 A.G. Elliot, A. Flat, D.A. Vanderwater: Silicon incorporation in LEC growth of single-crystal gallium-arsenide, J. Cryst. Growth **121**(3), 349–359 (1992)
39.98 M. Neubert, P. Rudolph: Growth of semi-insulating GaAs crystals in low temperature gradients by using the vapour pressure controlled Czochralski method (VCZ), Prog. Cryst. Growth Charact. Mater. **43**(2/3), 119–185 (2001)
39.99 G. Müller, J. Völkl, E. Tomzig: Thermal analysis of LEC InP growth, J. Cryst. Growth **64**(1), 40–47 (1983)
39.100 A.R. Von Neida, A.S. Jordan: Reducing dislocations in GaAs and InP, J. Met. **38**, 35–40 (1986)
39.101 A.G. Elliot, C.L. Wei, R. Farraro, G. Woolhouse, M. Scott, R. Hiskes: Low dislocation density, large diameter, liquid encapsulated Czochralski growth of GaAs, J. Cryst. Growth **70**, 169–178 (1984)
39.102 K. Katagiri, S. Yamazaki, A. Takagi, O. Oda, H. Araki, I. Tsuboya: LEC growth of large diameter InP single crystals doped with Sn and with S, Inst. Phys. Conf. Ser. **79**, 67–72 (1986)
39.103 R. Hirano, M. Uchida: Reduction of dislocation densities in InP single crystals by the LEC method using thermal baffles, J. Electron. Mater. **25**, 347–351 (1996)
39.104 R. Hirano: Growth of low etch pit density homogeneous 2″ InP crystals using a newly developed thermal baffle, Jpn. J. Appl. Phys. **38**(2B), 969–971 (1999)
39.105 K. Terashima, T. Fukuda: A new magnetic–field applied pulling apparatus for LEC GaAs single-crystal growth, J. Cryst. Growth **63**, 423–425 (1983)
39.106 H. Miyairi, T. Inada, M. Eguchi, T. Fukuda: Growth and properties of InP single crystals grown by the magnetic-field applied LEC method, J. Cryst. Growth **79**(1–3), 291–295 (1986)
39.107 J. Osaka, H. Kohda, T. Kobayashi, K. Hoshikawa: Homogeneity of vertical magnetic-field applied LEC GaAs crystal, Jpn. J. Appl. Phys. Part 2 – Lett. **23**(4), L194–197 (1984)
39.108 S. Ozawa, T. Kimura, J. Kobayashi, T. Fukuda: Programmed magnetic-field applied liquid encapsulated Czochralski crystal-growth, Appl. Phys. Lett. **50**(6), 329–331 (1987)
39.109 H. Kohda, K. Yamada, H. Nakanishi, T. Kobayashi, J. Osaka, K. Hoshikawa: Crystal-growth of completely dislocation-free and striation-free GaAs, J. Cryst. Growth **71**(3), 813–816 (1985)
39.110 S. Pendurti, H. Zhang, V. Prasad: Modeling dislocation generation in high pressure Czochralski growth of InP single crystals: Part II, Model. Simul. Mater. Sci. Eng. **13**, 267–297 (2005)

40. Mass and Heat Transport in BS and EFG Systems

Thomas F. George, Stefan Balint, Liliana Braescu

In this chapter several mathematical models describing processes which take place in the Bridgman–Stockbarger (BS) and edge-defined film-fed growth (EFG) systems are presented. Predictions are made concerning the impurity repartition in the crystal in the framework of each of the models. First, a short description of the real processes which are modeled is given, along with the equations, boundary conditions, and initial values defining the mathematical model. After that, numerical results obtained by computations in the framework of the model are provided, making a comparison between the computed results and those obtained in other models, and with the experimental data.

40.1 **Model-Based Prediction of the Impurity Distribution – Vertical BS System** 1380
40.1.1 Burton–Prim–Slichter Uniform-Diffusion-Layer Model (UDLM) 1380
40.1.2 Chang–Brown Quasi-Steady-State Model (QSSM) 1381
40.1.3 Adornato–Brown Pseudo-Steady-State Model (PSSM) 1383
40.1.4 Nonstationary Model (NSM) 1384
40.1.5 Modified Quasi-Steady-State Model (MQSSM) and Modified Nonstationary Model (MNSM) 1386
40.1.6 Larson–Zhang–Zheng Thermal-Diffusion Model (TDM) 1387
40.2 **Model-Based Prediction of the Impurity Distribution – EFG System** 1389
40.2.1 The Uniform-Diffusion-Layer Model (UDLM) 1389
40.2.2 Tatarchenko Steady-State Model (TSSM) 1389
40.2.3 Melt Replenishment Model (MRM) 1390
40.2.4 Melt Without Replenishment Model (MWRM) 1397
References 1400

The quality of the crystal being grown depends to a considerable extent on homogeneity and the distribution of both detrimental and specially added impurities. Variations of the concentration of impurities along the crystal length and cross section lead to variations of the mechanical, electrical, and optical properties in the mass of the crystal. The impurity repartition in the crystal is determined by the processes which take place in the melt during the growth and by mass transfer. Quantitative descriptions of these processes permit the prediction of such repartition in the crystal, which is a reason to build mathematical models in order to describe processes which take place during the growth. Consequently, computer modeling of the crystal growth processes has often been described as an art as well as a science, to acknowledge the seemingly endless difficulties that arise in the application of numerical methods. Three core competencies are required for the effective use of modeling in crystal growth. One is a thorough grasp of the fundamentals of continuum transport phenomena. The second is the building of an appropriate mathematical model, and the third is a general understanding of the numerical methods necessary for solving the governing equations of transport phenomena.

The philosophy behind building the mathematical model is to provide a quantitative description of the mechanism by balancing what is wanted with what can be done. The quantitative description is made in terms of a certain number of variables (called the model variables) such that the mathematical model is a set of equations concerning these variables. The analysis of

a mathematical model leads to results or predictions that can be tested against observations. It is important to realize that all models are idealizations and hence are limited in their applicability; experimental verification of the predictions can establish the authenticity of the model.

40.1 Model-Based Prediction of the Impurity Distribution – Vertical BS System

The prototype vertical BS crystal growth system (discussed here) consists of a crystal and melt contained in a cylindrical ampoule of radius R and length L pulled slowly through a vertically aligned furnace with hot and cold isothermal zones, as shown in Fig. 40.1.

The isothermal zones are separated by an adiabatic region of length L_g, designed to promote a steep axial temperature gradient and to maintain a flat solidification interface. A purpose of different mathematical models conceived to describe the processes which take place in such a system is the prediction of the distribution of detrimental and added impurities in the crystal being grown. The models discussed here describe one or several of the following processes: rejection of impurities at the melt–crystal interface, molecular diffusion, thermodiffusion, thermal convection, solutal convection, heat transfer between the furnace and ampoule walls, decrease of the melt in the ampoule, and melt–crystal interface morphology.

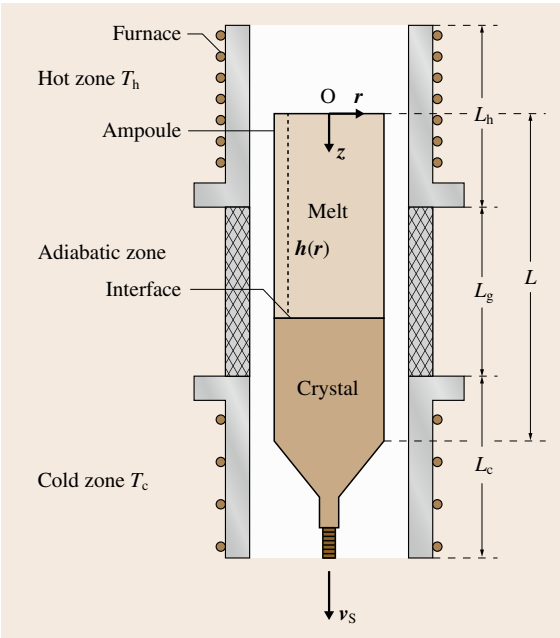


Fig. 40.1 Prototype BS crystal growth system

40.1.1 Burton–Prim–Slichter Uniform-Diffusion-Layer Model (UDLM)

For a large set of binary alloys, the rejection of impurities at the melt–crystal interface causes an unequal partitioning of the impurities between the melt and crystal during solidification. This leads to the accumulation of impurities in the melt and creates in the melt a nonuniform concentration field. A model to describe the above concentration field in the neighborhood of the melt–crystal interface (at the melt side) and the transfer of impurities in the crystal is the so-called uniform-diffusion-layer model (UDLM). The concept of an *axial boundary layer* evidently was introduced by *Burton* et al. [40.1]. Others [40.2–6] have extended this concept by introducing the notion of a *uniform diffusion layer* or *stagnant film*, without any picture for the fluid motion in the melt. Physically, the uniform diffusion layer is a thin layer masking the growing crystal which separates the crystal surface from the bulk melt [40.7].

According to UDLM, the isoconcentration lines of impurities are parallel to the interface, and the concentration decreases exponentially into the melt in this layer. The transport of impurities in the layer is molecular diffusion controlled, and the prediction is that, after an initial transient, the crystal exhibits a uniform (axial and radial) impurity distribution [40.8]. The effective partition coefficient k_{eff}, the equilibrium partition coefficient k, and the dimensionless diffusion-layer thickness $\delta = \tilde{\delta}/L$ are related by Fleming's formula

$$k_{\mathrm{eff}} = \frac{k}{k + (1-k)\exp\left(\frac{\mathrm{Pe\,Sc}}{\mathrm{Pr}}\delta\right)} , \tag{40.1}$$

where Pr is the Prandtl number, Pe is the Péclet number, and Sc is the Schmidt number [40.7]. The predicted impurity concentration in the crystal is given by

$$c_{\mathrm{crys}} = k_{\mathrm{eff}} c_0 , \tag{40.2}$$

where c_0 represents the impurity concentration in the bulk melt.

This model is appropriate for the prediction of impurities in crystals grown from a quiescent melt. Thus, a reduction in the magnitude of the buoyancy forces when processing semiconductors, such as in a low-gravity environment, has been pursued over the past several decades. The effectiveness of space processing for the growth of chemically uniform crystals is supported by experimental and theoretical studies [40.9, 10]. For example, the InSb crystal reported in [40.11] exhibits axial distribution profiles that are characteristic of diffusion-controlled mass transfer growth. In [40.7] it is shown, based on modeling studies, that the low gravity levels achieved in space are sufficient to inhibit thermal convection in small-diameter Ge and Ga-Ge melts. At the same time, it is pointed out that UDLM is an oversimplified model for accurate prediction of the impurity distribution in the crystal [40.12].

40.1.2 Chang–Brown Quasi-Steady-State Model (QSSM)

An important process in a BS system is thermal convection, the effect of which is essentially neglected in UDLM. Thermal convection is driven by buoyancy differences induced by axial and radial temperature gradients, and it may be laminar, periodic or turbulent. When the melt is below the crystal, the axial temperature gradient places the hottest melt at the bottom of the ampoule and generates convection. *Kim* et al. [40.3] used gallium-doped germanium to identify the melt length at which three forms of convection were present. In [40.13–26], the melt was oriented above the crystal in order to produce a stable axial density gradient. Because radial temperature gradients cause lateral density variations [40.27], convection still exists in this configuration. Thermal convection in a BS system contributes to the distribution of the impurities in the crystal.

The QSSM is a simplified model of the thermal convection and of the heat and mass transport in a BS system. The real process is replaced by a quasi-steady-state process viewed from a stationary reference frame and described in the cylindrical polar coordinate system shown in Fig. 40.1. The translation of the ampoule is replaced by supplying melt into the ampoule at $z = 0$ with velocity $\tilde{v}_z = v_L$ and removing the crystal at the other end of the ampoule at the rate $v_S = (\rho_S/\rho_L)\, v_L$, where ρ_L represents the melt density and ρ_S the crystal density. Changes in the length of the melt are neglected in QSSM. The location of the melt–solid interface is represented as $z = h(r)$, $0 \le r \le \Lambda$, $\Lambda = R/L$, and the unit vectors normal $\boldsymbol{n}$ and tangent $\boldsymbol{t}$ everywhere to the interface are

$$\boldsymbol{n} = \frac{\boldsymbol{e}_z - h_r \boldsymbol{e}_r}{\left(1+h_r^2\right)^{1/2}}\,, \quad \boldsymbol{t} = \frac{\boldsymbol{e}_r + h_r \boldsymbol{e}_z}{\left(1+h_r^2\right)^{1/2}}\,, \tag{40.3}$$

where $h_r = \mathrm{d}h/\mathrm{d}r$ and the set $(\boldsymbol{e}_r, \boldsymbol{e}_\theta, \boldsymbol{e}_z)$ are the unit vectors of the cylindrical polar coordinate system. According to [40.7], the equations governing the melt flow and heat and mass transport in the melt which define QSSM, in axisymmetric and dimensionless form, are

$$(\boldsymbol{v}\nabla)\,\boldsymbol{v} = -\nabla p + \mathrm{Pr}\nabla^2 \boldsymbol{v} + \mathrm{Ra_T Pr}\theta \boldsymbol{e}_z\,, \tag{40.4}$$

$$\nabla \boldsymbol{v} = 0\,, \tag{40.5}$$

$$\boldsymbol{v}\nabla\theta = \nabla^2\theta\,, \tag{40.6}$$

$$\left(\frac{\mathrm{Sc}}{\mathrm{Pr}}\right)\boldsymbol{v}\nabla c = \nabla^2 c\,. \tag{40.7}$$

Here, $\nabla = \boldsymbol{e}_r \partial/\partial r + \boldsymbol{e}_z \partial/\partial z$ represents the gradient operator in cylindrical coordinates; Pr is the Prandtl number; $\mathrm{Ra_T}$ is the thermal Rayleigh number; Sc is the Schmidt number; $\boldsymbol{v}$ is the dimensionless melt flow velocity; θ is the dimensionless melt temperature; c is the impurity concentration in the melt; and $0 \le z \le h(r)$, $0 \le r \le \Lambda$, $\Lambda = R/L$.

In the crystal, the velocity field $\boldsymbol{v}$ is uniform at speed v_S

$$\boldsymbol{v} = v_S \boldsymbol{e}_z\,, \quad h(r) \le z \le 1\,, \quad 0 \le r \le \Lambda\,. \tag{40.8}$$

The heat transport in the crystal is described by

$$\mathrm{Pe}\,\boldsymbol{e}_z \nabla\theta = \gamma \nabla^2\theta\,, \quad h(r) \le z \le 1\,, \quad 0 \le r \le \Lambda\,, \tag{40.9}$$

where $\gamma = \alpha_S/\alpha_L$ represents the ratio of thermal diffusivities in crystal and melt, and $\mathrm{Pe} = v_S L/\alpha_L$ is the Péclet number for convective heat transfer. The concentration of impurities in the crystal is given by

$$c_S(r) = kc(r, h(r))\,, \quad 0 \le r \le \Lambda\,. \tag{40.10}$$

Boundary conditions for the melt flow field $\boldsymbol{v}$ at the interface ensure no slip tangential to the crystal and the solidification of the melt at rate v_S. The top surface and sidewall of the ampoule are assumed to be no-slip surfaces, i. e.,

$$v_r = 0\,, \quad v_z = \frac{\rho_S}{\rho_L}\mathrm{Pe}\,. \tag{40.11}$$

This last condition dictates a jump in the velocity of the ampoule wall when $\sigma = \rho_S/\rho_L \neq 1$.

The axisymmetric boundary conditions for $\boldsymbol{v}$ and θ at the axis of the cylinder are

$$v_r = \frac{\partial v_z}{\partial r} = \frac{\partial \theta}{\partial r} = 0\,, \quad r = 0\,, \quad 0 \le z \le 1\,. \tag{40.12}$$

The thermal boundary conditions incorporate the assumption that the ampoule has negligible thermal mass, the adiabatic region is a perfect insulator, and the ampoule has the temperature of the surrounding furnace; the ends of the ampoule are perfectly insulated. At the melt–crystal interface, the temperature satisfies

$$\theta\,(r, h(r)) = \frac{T_\mathrm{m} - T_\mathrm{c}}{T_\mathrm{h} - T_\mathrm{c}}\,, \quad 0 \le r \le \Lambda\,, \tag{40.13}$$

$$(\boldsymbol{n}\nabla\theta)_\mathrm{L} - K\,(\boldsymbol{n}\nabla\theta)_\mathrm{S} = \mathrm{St}\,\mathrm{Pe}\,\boldsymbol{n}\boldsymbol{e}_\mathrm{z}\,; \quad z = h(r)\,, \quad 0 \le r \le \Lambda\,, \tag{40.14}$$

where T_m is the melting temperature, K is the ratio of the thermal conductivities between crystal and melt, and St is the Stefan number. The boundary conditions for the impurity concentrations are

$$\frac{\partial c}{\partial z} = \left(\frac{\mathrm{Pe}\,\mathrm{Sc}}{\mathrm{Pr}}\right)(c-1)\,; \quad z = 0\,, \quad 0 \le r \le \Lambda\,, \tag{40.15}$$

$$\boldsymbol{n}\nabla c = \left(\frac{\mathrm{Pe}\,\mathrm{Sc}}{\mathrm{Pr}}\right)(\boldsymbol{n}\boldsymbol{e}_\mathrm{z})\,(1-k)\,c\,; \quad z = h(r)\,, \quad 0 \le r \le \Lambda\,, \tag{40.16}$$

$$\frac{\partial c}{\partial r} = 0\,; \quad 0 \le z \le 1\,, \quad r = 0\,, \quad r = \Lambda\,. \tag{40.17}$$

These conditions express the following: a supply melt inlet at the top of the BS system; the rejected impurities at the melt–crystal interface; and the no-flux condition which is valid at the center line and at the side wall of the ampoule.

Using finite-element numerical methods for crystals and melts with thermophysical properties similar to gallium-doped germanium, the temperature field and dopant field in the melt and in the crystal were computed [40.7, 28] for several different $\mathrm{Ra_T}$ numbers in QSSM. Results are presented for two Bridgman configurations: vertically stabilized (melt on the top) and vertically destabilized (melt on bottom). Steady axisymmetric flows are classified according to $\mathrm{Ra_T}$ and are nearly equal to the growth velocity. Three distinct types of flow patterns were observed with changing $\mathrm{Ra_T}$. At low Rayleigh numbers, the streamlines were rectilinear and only slightly distorted by buoyancy forces. Increasing the Rayleigh number developed first a cellular flow, and after that a weak secondary cell for stabilized configuration. The flows in these two configurations are in opposite directions. Calculations of the transport of a dilute dopant by these flow fields reveal radial segregation levels as large as 60% of the mean concentration. Radial segregation is most severe at an intermediate value of the Rayleigh number, above which the dopant distribution along the interface levels increases. The complexity of the concentration field coupled with calculations of the effective segregation coefficient show the coarseness of the usual diffusion-layer approximation for describing the dopant distribution adjacent to the crystal.

The results reveal that UDLM is a gross oversimplification of the interactions between the complicated flow patterns and the concentration field of impurities. The diffusion-layer thickness δ considered in UDLM in the best case is an empirical fit to the effective segregation coefficient. Although the radially averaged diffusion-layer thickness computed in QSSM is of the same order of magnitude as those considered in UDLM, the actual concentration gradient adjacent to the crystal is far from being radially uniform. In the computations made in QSSM, the ratio of thermal conductivities K between the melt and the crystal and the length of the gradient zone L_g have emerged as important parameters for determining the interface shape and radial segregation. According to [40.7], the value of K sets the shape of the interface and determines the qualitative behavior of the radial segregation with increased convection, and L_g sets the degree of interaction between the cellular flow and the concentration gradient adjacent to the interface. For long enough gradient zones and moderate convection, the flow causes lower radial segregation. A critical issue not addressed in this chapter is the transition from the steady-state flow patterns predicted here to time-periodic flows observed in some experiments. Mathematically, these transitions occur as Hopf bifurcations from the steady flows and have been predicted theoretically by computer-aided bifurcation analysis for low-Prandtl-number melts in the horizontal boat geometry as studied experimentally by *Hurle* et al. [40.29]. Simulation of the crystal growth in the time-periodic regime requires a time-dependent solution of the Boussinesq equations.

The qualitative understanding of melt flow and impurities segregation offered by QSSM serves as a starting point for more refined and accurate calculations aimed at comparison with experimental data. These would require the inclusion in the model of the thermal interactions between the ampoule and the furnace, the solutal convection, the decrease of the melt in the ampoule, the thermal diffusion (Soret effect), and the effect of the precrystallization zone.

40.1.3 Adornato–Brown Pseudo-Steady-State Model (PSSM)

Thermosolutal convection which can occur in a BS system is driven by buoyancy differences induced by axial and radial temperature gradients and by concentration-dependent density differences. Convection driven by such density differences is initiated in the layer adjacent to the melt–crystal interface and is caused by the accumulation of impurities in this region. In nondilute alloy systems, the interaction of the solutal and thermal driving forces is so closely coupled through convention of the solute field that only a full numerical simulation can give a quantitative description of the flow [40.12]. PSSM is a simplified model of the real thermosolutal convection and of the heat and mass transport in a BS system which takes into account the heat transfer between the furnace and the ampoule. In PSSM, like in QSSM, the translation of the ampoule is replaced by supplying melt at the top of the ampoule with uniform velocity and composition, and removing crystal from the bottom of the ampoule at a certain growth rate. Transients in velocity, pressure, temperature, and concentration caused by the steady decrease of the length of the melt in the real ampoule, and the displacement of the ampoule in the furnace, are neglected in PSSM.

The equations (in axisymmetric and dimensionless form) governing the melt flow, and heat and mass transport in the melt, defining PSSM [40.12] are the following

$$(\boldsymbol{v}\nabla)\,\boldsymbol{v} = -\nabla p + \mathrm{Pr}\nabla^2\boldsymbol{v} + \mathrm{Pr}\,(\mathrm{Ra_T}\theta - \mathrm{Ra_S}c)\,\boldsymbol{e}_z\,, \tag{40.18}$$

$$\nabla\boldsymbol{v} = 0\,, \tag{40.19}$$

$$\boldsymbol{v}\nabla\theta = \nabla^2\theta\,, \tag{40.20}$$

$$\left(\frac{\mathrm{Sc}}{\mathrm{Pr}}\right)\boldsymbol{v}\nabla c = \nabla^2 c\,, \tag{40.21}$$

where ∇, Pr, $\mathrm{Ra_T}$, Sc, $\boldsymbol{v}$, θ, and c have the same meanings as in (40.4–40.7), and $\mathrm{Ra_S}$ represents the solutal Rayleigh number ($0 \le z \le h(r)$, $0 \le r \le \Lambda$). The velocity field $\boldsymbol{v}$ in the crystal is uniform at speed v_S

$$\boldsymbol{v} = v_S\boldsymbol{e}_z\,; \quad h(r) \le z \le 1\,, \quad 0 \le r \le \Lambda\,. \tag{40.22}$$

The heat transport equations in the ampoule and in the crystal are

$$\mathrm{Pe_a}\boldsymbol{e}_z\nabla\theta = \gamma_a\nabla^2\theta\,; \quad h(r) \le z \le 1\,, \quad 0 \le r \le \Lambda \tag{40.23}$$

$$\mathrm{Pe_c}\boldsymbol{e}_z\nabla\theta = \gamma_c\nabla^2\theta\,; \quad h(r) \le z \le 1\,, \quad 0 \le r \le \Lambda\,, \tag{40.24}$$

where $\gamma_a = \alpha_a/\alpha_L$ and $\gamma_c = \alpha_c/\alpha_L$ are the ratios between the thermal diffusivities of each phase and the melt. The Péclet number Pe is the dimensionless transition rate of the ampoule scaled by the characteristic velocity for thermal diffusion. The shape of the melt–crystal interface is set by the shape of liquids curve $\theta_L(c)$ from the binary phase diagram as

$$\theta(r, h(r)) = \theta_L(c) = \theta_L^0 + m\left(c + 1 - \frac{1}{k}\right)\,, \quad 0 \le r \le \Lambda\,, \tag{40.25}$$

where the curve is approximated by a straight line with dimensionless slope m, and θ_L^0 is the melting temperature of the alloy with concentration c_0. It is assumed that m and the partition coefficient k are independent of concentration.

The interfacial temperature and impurities field are related to the corresponding values in the bulk melt by the balances

$$(\boldsymbol{n}\nabla\theta)_L - K\,(\boldsymbol{n}\nabla\theta)_S = \mathrm{St}\,\mathrm{Pe}\,\boldsymbol{n}\boldsymbol{e}_z\,, \quad 0 \le r \le \Lambda \tag{40.26}$$

$$\boldsymbol{n}\nabla c = \frac{\mathrm{Pe\,Sc}}{\mathrm{Pr}}\,(\boldsymbol{n}\boldsymbol{e}_z)\,(1-k)\,c\,; \quad z = h(r)\,, \quad 0 \le r \le \Lambda\,. \tag{40.27}$$

The boundary conditions for the melt flow field are the same as those presented for QSSM. The thermal boundary conditions along the ampoule wall are specified according to the heat transfer condition

$$\left.\frac{\partial\theta}{\partial r}\right|_a = Bi(z)\cdot(\theta - \theta_\infty(z))\,, \tag{40.28}$$

where $Bi(z) = h_a \cdot L/k_a$ is a dimensionless heat transfer coefficient defined to include radiative, conductive, and convective transport between the furnace and the ampoule, and $\theta_\infty(z)$ is the furnace temperature distribution. Perfect thermal contact at the melt–ampoule and crystal–ampoule interfaces are assumed, so that the conductive fluxes at the boundaries are equal

$$\left.\frac{\partial\theta}{\partial r}\right|_L = K_a\left.\frac{\partial\theta}{\partial r}\right|_a\,, \quad K_c\left.\frac{\partial\theta}{\partial r}\right|_c = K_a\left.\frac{\partial\theta}{\partial r}\right|_a\,; \quad r = \Lambda\,. \tag{40.29}$$

The top and bottom of the ampoule are assumed to be at the temperature of the hot and cold zones. The specification of the concentration field of impurities in the melt is completed by setting the diffusive flux of impurities through the inlet to be equal to $\mathrm{Pe_c}$, and zero through

the ampoule wall. The concentration of the impurities in the crystal is given by

$$c_S(r) = kc(r, h(r)) , \quad 0 \le r \le \Lambda . \tag{40.30}$$

The axis of the ampoule is taken as a line of symmetry for all field variables.

Using finite-element numerical methods, computations in PSSM were performed for a binary alloy coupling heat transfer in the melt, crystal, and ampoule [40.12] with convection in the melt and segregation in the growing crystal, taking into account convection driven by both temperature and concentration gradients. Calculations presented for growth of gallium-doped germanium in ampoules of boron nitride, quartz, and graphite, along with a quartz/graphite composite ampoule, show the effects of ampoule conductivity on the flow structure and intensity. The thermal conductivity of the ampoule material and the thickness of the ampoule wall can be used as parameters to change the structure of the flow, as well as to modulate the flow intensity. The computations show that flow pattern and concentration fields have rich structures, depending on the furnace, either that constructed at Massachusetts Institute of Technology (MIT) by Wang and Witt (based on heat pipes for forming isothermal hot and cold regions separated by a nearly adiabatic zone; Fig. 40.1) or at Grenoble by Rouzaud and Favier (using a tapered heating element to establish a nearly linear temperature profile over the length of the ampoule). The vertical BS system used at MIT has a two-cell structure: one cell driven by the radial gradients at the junction of the adiabatic and hot zones of the furnace, and the other adjacent to the interface caused by the differences in the thermal conductivities of the melt, crystal, and ampoule. The constant-gradient furnace designed at Grenoble has only a single cell near the interface.

Impurities are transported mostly by diffusion between adjacent cells because the convective motion is laminar. The radial segregation is set primarily by the flow adjacent to the interface when this cell is long enough to completely contain the solute gradient for diffusion-controlled growth. The upper cell in the MIT furnace plays a small role in determining the solute distribution along the interface. Axial segregation depends on the intensity of the convective mixing through the ampoule, so that the cellular structure of the flow pattern is important. In the heat pipe system, solute transport between flow cells limits the effectiveness of this mixing and leads higher solute concentrations in the cell next to the interface. The degree of axial segregation is set mainly by this cell.

The dependence of the degree of radial segregation on Ra_T studied in both the MIT and Grenoble systems is characterized by the following: (i) segregation is lowest for either diffusion-controlled growth (small Ra_T) or for when the intense laminar mixing is present (large Ra_T); (ii) segregation reaches a maximum for intermediate values of low intensity; (iii) the ampoule material, furnace design, and stabilizing solute all affect the values of segregation. This result is crucial for understanding the effects on solute segregation of microgravity solidification and of applied magnetic fields, and answers two questions which were not clarified before: (i) the way in which the reduced convection level affects the solute segregation, and (ii) the dependence of the segregation level on the ampoule size for the same Ra_T. In both situations the body forces acting on the melt are altered in an attempt to damp convection. Both approaches are clearly effective at eliminating time-periodic and chaotic convection present in large-size systems. If the reduced gravitational field has constant magnitude and is aligned with the crystal axis, then it is modeled exactly by reducing the thermal Rayleigh number. Thus, in [40.12], it is shown that the decrease in gravitational level needed to reach diffusion-controlled growth increases with a value which depend on the cube of the ampoule radius. This proves that the segregation effect is more severe for larger ampoules.

The prediction of the melt–crystal interface shape, radial impurity distribution, and effective segregation coefficient for the growth of dilute gallium-doped germanium obtained in [40.12] agrees reasonably well with the measurements by *Wang* [40.30].

40.1.4 Nonstationary Model (NSM)

NSM is a model of the unsteady thermoconvection, heat and impurity transport in a BS system which takes into account the decrease of the melt in the ampoule. The equations that govern such phenomena can be expressed in a mobile reference frame (Fig. 40.1) in axisymmetric and dimensionless form as [40.31]

$$\frac{\partial \boldsymbol{v}}{\partial t} + (\boldsymbol{v}\nabla)\boldsymbol{v} = -\nabla p + \mathrm{Pr}\nabla^2\boldsymbol{v} + \mathrm{Pr}\,\mathrm{Ra_T}\theta\boldsymbol{e}_z , \tag{40.31}$$

$$\nabla\boldsymbol{v} = 0 , \tag{40.32}$$

$$\frac{\partial \theta}{\partial t} + \boldsymbol{v}\nabla\theta = \nabla^2\theta , \tag{40.33}$$

$$\frac{\partial c}{\partial t} + \boldsymbol{v}\nabla c = \frac{\mathrm{Pr}}{\mathrm{Sc}}\nabla^2 c . \tag{40.34}$$

The meanings of the symbols which appear in the above equations correspond to those in (40.18–40.21). These equations are time dependent since the domain Ω_t occupied by the melt is time dependent. The boundary $\partial\Omega_t$ of Ω_t is decomposed as $\partial\Omega_t = (\partial\Omega_t)_{top} \cup (\partial\Omega_t)_w \cup (\partial\Omega_t)_i$, where $(\partial\Omega_t)_{top}$ and $(\partial\Omega_t)_w$ are the melt boundaries limited by the top and side walls of the ampoule, respectively, and $(\partial\Omega_t)_i$ is the melt boundary limited by the bottom of the ampoule for t satisfying $0 \le t \le L_g/(2v_pL)$ and by the melt–crystal interface for t satisfying $L_g/(2v_pL) \le t \le (1 + L_g/(2L))/v_p$, where v_p is the ampoule translation rate of the ampoule.

The boundary conditions for $\boldsymbol{v}$, θ, and c are

$$\boldsymbol{v} = 0 \quad \text{on} \quad \partial\Omega_t \,, \tag{40.35}$$

$$\begin{aligned} \frac{\partial\theta}{\partial z} &= 0 \quad \text{on} \quad (\partial\Omega_t)_{top} \,; \\ \theta &= \theta_f \quad \text{on} \quad (\partial\Omega_t)_w \,; \\ \theta &= \theta_i \quad \text{on} \quad (\partial\Omega_t)_i \,, \end{aligned} \tag{40.36}$$

$$\begin{aligned} \frac{\partial c}{\partial z} &= 0 \quad \text{on} \quad (\partial\Omega_t)_{top} \,; \\ \frac{\partial c}{\partial r} &= 0 \quad \text{on} \quad (\partial\Omega_t)_w \,; \\ \frac{\partial c}{\partial z} &= \Phi_i \quad \text{on} \quad (\partial\Omega_t)_i \,. \end{aligned} \tag{40.37}$$

Here, θ_f is the temperature of the surrounding furnace, θ_i is the temperature of the bottom of the ampoule for t satisfying $0 \le t \le L_g/(2v_pL)$, and temperature at the melt–crystal interface for t satisfying $L_g/(2v_pL) \le t \le (1 + L_g/(2L))/v_p$, $\Phi_i = 0$ for t satisfying $0 \le t \le L_g/(2v_pL)$; and $\Phi_i = (\mathrm{Pe\,Sc})/(\mathrm{Pr})(1-k)c$ for t satisfying $L_g/(2v_pL) \le t \le (1 + L_g/(2L))/v_p$. The initial conditions for the equations (40.31–40.34) are

$$\begin{aligned} &\boldsymbol{v}(0,\ r,\ z) = 0 \,; \\ &\theta(0,\ r,\ z) = 1 \,; \\ &c(0,\ r,\ z) = 1 \,. \end{aligned} \tag{40.38}$$

The impurity distribution in the crystal is given by

$$\begin{aligned} c_S(\xi, r) &= kc\,(t, r, h(r)) \\ &= kc\left(\frac{1}{v_p}\left(\xi + \frac{L_g}{2L}\right),\ r,\ h(r)\right) \,, \end{aligned} \tag{40.39}$$

where ξ represents the solidified fraction

$$\begin{aligned} &\xi = v_p t - \frac{L_g}{2L} \\ &\text{for} \quad \frac{L_g}{2v_pL} \le t \le \left(1 + \frac{L_g}{2L}\right)/v_p \,. \end{aligned} \tag{40.40}$$

Computations in NSM were made by the finite-element method using the software Cosmos/M [40.31] in a low-gravity environment ($\mathrm{Ra_T} = 10^3$) for gallium-doped germanium. The results reveal that, during the growth process, $|\boldsymbol{v}|_{max} = |v_z|_{max}$ is localized at $r = 0$ at the distance $1.25L_g/L$ from the top of the crystal, and $v_r|_{max}$ is localized at $r = 0.125$ at the distance $L_g/(2L)$ from the top of the crystal. In the first part of the growth process, $|\boldsymbol{v}|_{max} = |v_z|_{max}$ increases from 0.1772 to 0.1817 and $|v_r|_{max}$ increases from 0.0598 to 0.0621. In the second part, $|\boldsymbol{v}|_{max} = v_z|_{max}$ decreases from 0.1817 to 0.0193

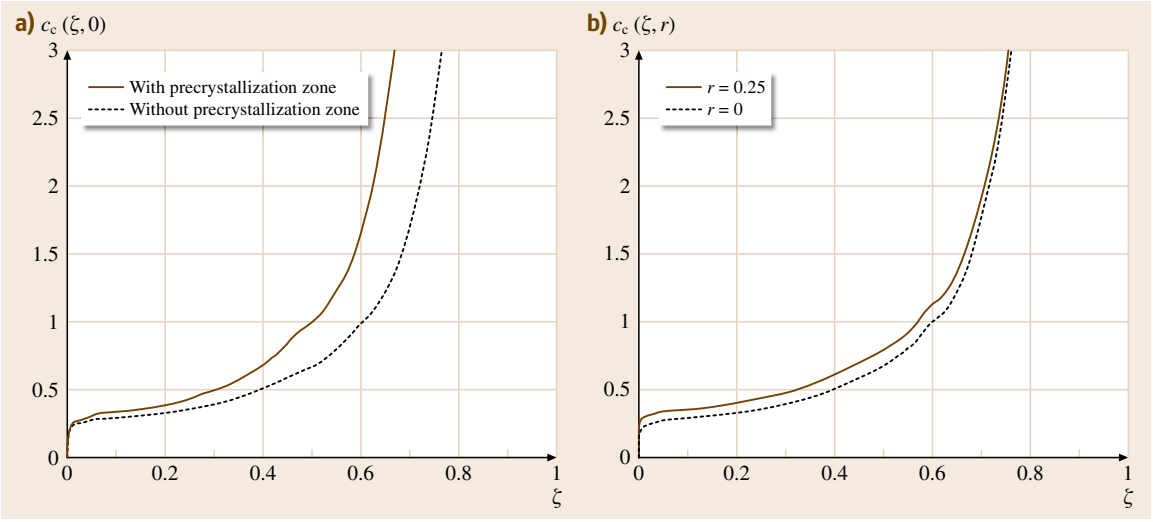


Fig. 40.2a,b Computed Ga concentration in the crystal versus ξ and r

and $|v_r|_{max}$ decreases from 0.0621 to 0.0043. During the growth, at all points of the melt, the impurity concentration c increases but not uniformly. On the axis, parallel to the ampoule axis, c increases from the top towards the interface. In a transversal section of the ampoule, c increases from the axis towards the side wall. The impurity concentration on the axis of the crystal increases from kc_0 to c_0 in the solidified fraction $\xi \in (0;\ 0.6099)$, and the percent axial segregation defined by the formula $\Delta c_{ax} = |c_c(0,\ 0) - c_c(0.6099,\ 0)| \times 100/k$ is $\Delta c_{ax} = 854.5$ (Fig. 40.2a). The percent radial segregation, defined as $\Delta c_{rad} = |c_c(\xi,\ 0) - c_c(\xi,\ \Lambda)| \times 100/k$ for $\xi = 0.6099$, is $\Delta c_{rad} = 134$ (Fig. 40.2b).

Due to the decrease of the melt in the ampoule, the percent axial segregation in the crystal increases with the solidified fraction ξ. The percent radial segregation in the crystal increases for the first three-quarters and decreases thereafter. Finally, it has been shown [40.32] that, if the initial impurity distribution in the melt is not uniform like in [40.31], and if it is given by

$$c_0(r, z) = \frac{1}{\alpha(a)} \left\{ 1 + \frac{1-k}{k} \frac{\Lambda - r}{\Lambda} \times \exp\left[-\frac{v_p \mathrm{Sc}}{\mathrm{Pr}} (1-z) \right] - \frac{a \cdot r}{\Lambda} \right\} , \quad (40.41)$$

then the computed impurities distribution in the crystal (in the NSM) on the solidified fraction $\xi \in (0;\ 0.7)$ is a constant equal to the prescribed concentration $c_c = 1$.

40.1.5 Modified Quasi-Steady-State Model (MQSSM) and Modified Nonstationary Model (MNSM)

In order to evaluate the effect of the precrystallization zone on the impurity distribution in the crystal, two different models – MQSSM and MNSM – have been reported in the literature [40.28, 33–35]. In such models, it is assumed that the melt→solid phase transition does not take place on a flat mathematical surface like in UDLM, QSSM, PSSM, and NSM, but rather in a thin region of width 10^{-9} m masking the crystal, where both phases coexists and there are periodically distributed solid inclusions (clusters) of size 10^{-10} m. This periodic microstructure is created by the crystal which, like a matrix, governs the arrangement of its own atoms from the melt into a specific crystalline lattice and creates periodically distributed clusters (solid inclusions). These clusters do not affect the melt flow but can influence the diffusion of impurities and the heat conductivity. In the isotropic case, for the impurities and heat transport, new diffusion and heat conductivity coefficients appear, and in the anisotropic case, new diffusion and heat conductivity tensors appear. These new coefficients and tensors depend on the spatial coordinates, on the geometry of clusters, and on the volumetric concentration of the clusters.

In MQSSM presented in [40.28], it is assumed that in the precrystallization zone the new diffusion coefficient decreases with respect to the spatial coordinate z (Fig. 40.1). According to [40.36, 37], the clusters do not influence the flow near the interface (i. e., the computed flow fields are identical to those obtained in QSSM model), but have a very small influence (of order 10^{-4}) on the heat transport. For example, changing the Schmidt number in the impurity transport equation from QSSM, the effect of the precrystallization zone on the axial and radial impurity distribution in the neighborhood of the melt–crystal interface was computed for different volumetric concentration of the clusters (solid inclusions). Computations performed with this model show that the influence of the precrystallization zone on the repartition of impurities is relevant. The concentration of impurities in the precrystallization zone reaches values which can be as high as ten times those obtained with QSSM. Consequently, the axial impurity distribution is nonuniform. In MNSM presented in [40.34], it is assumed that the microstructure existing in the precrystallization zone does not influence the flow and the heat transport, but that the transport of impurities is governed by the equation

$$\Pi \frac{\partial c}{\partial t} + \boldsymbol{v} \nabla c = \frac{\mathrm{Pr}}{\mathrm{Sc}} (\boldsymbol{I} + \boldsymbol{\Phi}) \nabla^2 c , \quad (40.42)$$

where Π is porosity, $\boldsymbol{I}$ is the identity tensor, and $\boldsymbol{\Phi}$ is a tensor defined by

$$\boldsymbol{\Phi}_{kl} = -\varphi \left(\delta_{kl} - \frac{1}{|V|} \int_{\partial V} \gamma_k n_l \, \mathrm{d}S \right) . \quad (40.43)$$

In (40.43), φ represents the volumetric concentration of the clusters, $|V|$ represents the volume of a cluster, γ_k is the solution of the boundary-value problem, δ_{kl} is the Kronecker symbol, and $\mathrm{d}S$ represents the elementary surface [40.36]. In particular, if the clusters are spheres, the new diffusion tensor $\boldsymbol{D}(\boldsymbol{I} + \boldsymbol{\Phi})$ is diagonal, and the new diffusion coefficients are given by the Maxwell–Jeffries formula [40.38]

$$\frac{\boldsymbol{D}_{11}}{\boldsymbol{D}} = \frac{\boldsymbol{D}_{22}}{\boldsymbol{D}} = \frac{\boldsymbol{D}_{33}}{\boldsymbol{D}} = 1 - \left(\tfrac{3}{2} \right) \varphi . \quad (40.44)$$

in these conditions, (40.42) becomes

$$\Pi \frac{\partial c}{\partial t} + \boldsymbol{v} \nabla c = \frac{\mathrm{Pr}}{\mathrm{Sc}} \left(1 - \tfrac{3}{2} \varphi \right) \nabla^2 c . \quad (40.45)$$

Therefore, the dispersion of impurities in the precrystallization zone is governed by (40.45), and by (40.34) in the rest of the melt.

Numerical computations of the impurities concentration have been made for the melt and crystal with thermophysical properties similar to gallium-doped germanium [40.34]: $\mathrm{Ra_T} = 10^3$; $\varphi = 0.00985,\ 0.0317,\ 0.0491,\ 0.055$; $\mathrm{Sc} = 10$; $k = 0.1$; and replacing the true unsteady process by the quasi-steady-state process. Computations show that the impurity concentration in the crystal is larger than that obtained with QSSM. The axial impurity distribution is uniform, and the radial segregation of the impurities in the crystal, computed with MQSSM, is less than that computed with QSSM.

In [40.33], it was shown rigorously (from a mathematical point of view) that the assumption of the existence of a periodic structure in the precrystallization zone permits us to deduce MNSM starting from NSM in the framework of continuum mechanics. The obtained MNSM includes the effect of the precrystallization zone. This effect gives a decrease of the molecular diffusivity D of the impurities from D to $D_{\mathrm{eff}} = 0.917D$, and a increase of the heat conductivity K to $K_{\mathrm{eff}} = 1.1K$ in the precrystallization zone. Numerical computations made with MNSM using the quasi-steady-state approximation [40.33, 35] show that, due to the decrease of the molecular diffusivity D, the impurity concentration in the crystal is five times larger than that computed with QSSM. Computations made in [40.31] using MNSM and those presented for a low-gravity environment ($\mathrm{Ra_T} = 10^3$) in [40.33] show that the impurity concentration on the axis and on the walls of the crystal increases faster than that computed with NSM. The ratio of the impurity concentrations computed with MNSM and NSM at $\xi = 0.5$ (on the axis and on the walls of the crystal) is 1.45. According to [40.39], this result is in agreement with the measurements of the samples grown on the Mir space station.

Finally, in [40.40] it is shown that, if the initial impurities distribution in the melt is not uniform like in [40.31], and it is given by the formula

$$c_0(r,\,z) = \frac{1}{\alpha(a)}\left\{1 + \frac{1-k}{k}\,\frac{\Lambda - r}{\Lambda} \times \exp\left[-\frac{v_{\mathrm{p}}\mathrm{Sc}}{\mathrm{Pr}}\,(1-z)\right] - \frac{ar}{\Lambda}\right\}\,, \tag{40.46}$$

then the computed impurity distribution in the crystal (in the framework MNSM) on the solidified fraction $\xi \in (0;\ 0.7)$ is a constant equal to the prescribed concentration $c_{\mathrm{c}} = 1$.

40.1.6 Larson–Zhang–Zheng Thermal-Diffusion Model (TDM)

Doped crystal grown in space in a BS system reported in [40.41, 42] exhibits significant axial compositional nonuniformities. Due to the fact that in these cases the thermosolutal convection can be neglected, the thermodiffusion of impurities is responsible for these nonuniformities. A one-dimensional mathematical model [40.43] has been developed to explain the impurity distribution in the sample presented in [40.41]. According to [40.43], for ampoules with small-diameter cross section and for a small impurity concentration, in a strictly zero-gravity environment, the mathematical model of the heat and impurities dispersion in the melt in a BS system is defined by the following dimensionless equations, boundary conditions, and initial conditions as written in a reference frame with the origin fixed at the interface

$$\frac{\partial\theta}{\partial t} = v_{\mathrm{p}}\frac{\partial\theta}{\partial z} + \frac{1}{\mathrm{Pr}}\frac{\partial^2\theta}{\partial z^2} - \frac{2}{\mathrm{Pr}}\frac{K_{\mathrm{a}}}{K}\frac{L}{\delta}\frac{L}{R}\,(\theta - \theta_{\mathrm{f}})\,,$$
$$\text{for}\quad 0 < t < \frac{1}{v_{\mathrm{p}}}\quad\text{and}\quad 0 < z < 1 - v_{\mathrm{p}}t\,, \tag{40.47}$$

$$\frac{\partial c}{\partial t} = v_{\mathrm{p}}\frac{\partial c}{\partial z} + \frac{1}{\mathrm{Sc}}\frac{\partial^2 c}{\partial z^2} + \frac{1}{\mathrm{Sc}}\frac{\partial}{\partial z}\left(\sigma\Delta T c\frac{\partial\theta}{\partial z}\right)\,,$$
$$\text{for}\quad 0 < t < \frac{1}{v_{\mathrm{p}}}\quad\text{and}\quad 0 < z < 1 - v_{\mathrm{p}}t\,, \tag{40.48}$$

$$\theta(t,\,z) = 0\quad\text{for}\quad 0 < t < \frac{1}{v_{\mathrm{p}}}\quad\text{and}\quad z = 0\,, \tag{40.49}$$

$$\frac{\partial\theta}{\partial z} = 0\quad\text{for}\quad 0 < t < \frac{1}{v_{\mathrm{p}}}\quad\text{and}\quad z = 1 - v_{\mathrm{p}}t\,, \tag{40.50}$$

$$\frac{\partial c}{\partial z} + v_{\mathrm{p}}\mathrm{Sc}\,(1-k)\,c + \sigma c\Delta T\frac{\partial\theta}{\partial z} = 0\,,$$
$$\text{for}\quad 0 < t < \frac{1}{v_{\mathrm{p}}}\quad\text{and}\quad z = 0\,, \tag{40.51}$$

$$\frac{\partial c}{\partial z} = 0\quad\text{for}\quad 0 < t < \frac{1}{v_{\mathrm{p}}}\quad\text{and}\quad z = 1 - v_{\mathrm{p}}t\,, \tag{40.52}$$

$$\theta(0,\,z) = \begin{cases} 1 & \text{for}\ \dfrac{L_{\mathrm{g}}}{2L} \le z \le 1 \\ \dfrac{2L}{L_{\mathrm{g}}}z & \text{for}\ 0 \le z \le \dfrac{L_{\mathrm{g}}}{2L} \end{cases}\,, \tag{40.53}$$

$$c(0,\,z) = c_0(z)\,. \tag{40.54}$$

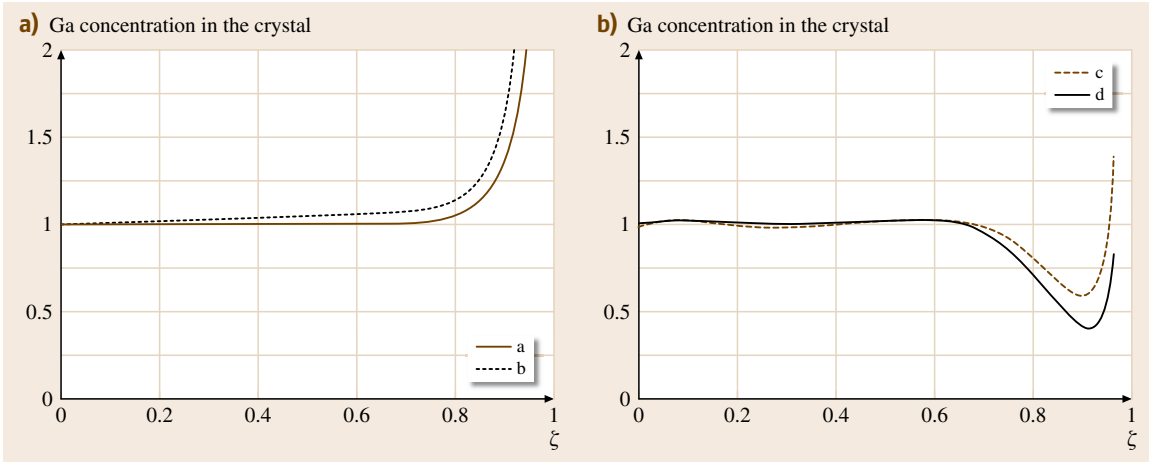


Fig. 40.3a,b Computed Ga concentration in the crystal versus ξ: **(a)** for the nonuniform case and **(b)** for the modified initial dopant repartition (after [40.44])

Equation (40.47) represents the dimensionless heat transport equation in the melt, which takes into account the heat dispersion due to the growth rate v_p, the thermal diffusivity in the melt $\lambda_L = \nu_L/\mathrm{Pr}$, and the heat loss to the environment from the sample surface. Equation (40.48) represents the dimensionless impurities transport equation in the melt which takes into account the impurity dispersion due to the growth rate v_p, the molecular diffusion $D = \nu/\mathrm{Sc}$, and the thermal diffusion σ (Soret effect). The boundary condition (40.49) expresses the melt–crystal interface as fixed at $z = 0$ and the dimensionless temperature at the interface equal to zero. In general, the shape of the melt–crystal interface is set by the conditions for the equilibrium melting temperature and the balance of conductive heat fluxes across the interface with the latent heat released there; for strictly zero gravity ($\mathrm{Ra_T} = 0$) [40.7], computations show that the interface is flat and fixed. The boundary condition (40.50) represents the no-flux condition for the heat at the top of the ampoule, and the boundary condition (40.51) gives the dopant concentration near the melt–crystal interface on the melt side. The boundary condition (40.52) is for no flux of the impurities on the top of the ampoule. The initial condition (40.53) is the temperature distribution in the melt at the time when solidification begins, and the initial condition (40.54) represents the impurity repartition in the melt when solidification begins.

The solidified fraction ξ at time t is given by

$$\xi = v_p t \,, \tag{40.55}$$

and the impurity concentration in the crystal is given by

$$c_c(\xi) = kc\left(\frac{\xi}{v_p}, 0\right) \quad \text{for} \quad 0 \le \xi \le 1 \,. \tag{40.56}$$

Numerical results obtained with TDM using a second-order central difference scheme [40.43] are in agreement with the experimental results reported in [40.41]. The same numerical method with TDM gives numerical results which are qualitatively in agreement with the results reported in [40.42], as shown in [40.45]. Moreover, in [40.45] it is shown that, if the initial impurity distribution in the melt is not uniform (Fig. 40.3a) like in [40.42], but if it is given by the formula

$$c_0(x) = \frac{v_p \mathrm{Sc} L_g}{v_p \mathrm{Sc} L_g + \sigma \Delta T 2L}\left\{1 + a_1 x + a_2 x^2 + \left(\frac{1-k}{k} + \frac{\sigma \Delta T 2L}{k v_p \mathrm{Sc} L_g}\right) \times \exp\left[-\left(v_p \mathrm{Sc} + \frac{\sigma \Delta T 2L}{L_g}\right) z\right]\right\} \,, \tag{40.57}$$

then the computed impurity concentration in the crystal on the solidified fraction $\xi \in (0;\ 0.7)$ is a constant equal to the prescribed concentration $c_c = 1$ (Fig. 40.3b).

40.2 Model-Based Prediction of the Impurity Distribution – EFG System

The prototype EFG system used in this work is presented in Fig. 40.4.

The central component of the system is a die which controls the heat and mass transfer from the crucible to the meniscus, i. e., the liquid bridge retained between the die and the crystal. Processes that take place in the melt (more precisely, in the meniscus melt) near the melt–crystal interface determine the impurity distribution in the crystal. During the past three decades, several mathematical models have been reported for predicting the impurity distribution in a crystal grown in such a system. Some of these models and the predictions obtained are presented below.

40.2.1 The Uniform-Diffusion-Layer Model (UDLM)

The UDLM presented for a BS system was extended for an EFG system considering that the effective partition coefficient k_{eff} is given by Fleming's formula [40.46]. If it is assumed that the uniform diffusion layer thickness is equal to the total height of the meniscus and the die has a long capillary channel, then the effective partition coefficient k_{eff} is equal to unity [40.47–53] for any impurity. Hence, according to UDLM, the impurity distribution in the crystal is uniform (i. e., no variations of the impurity concentration along the crystal length and cross section), and the value of the impurity concentration in the crystal is given by

$$c = k_{\text{eff}} c_\infty = c_\infty \,. \tag{40.58}$$

This prediction implies that the growth process in an EFG system has a low tolerance for rejected melt impurities, which definitively shows that UDLM is an oversimplified model. Impurities are generally trapped in the boundary layer within the meniscus melt. Interface-region convection in EFG can be manipulated through design of die capillaries to redistribute segregated impurities [40.50–53]

40.2.2 Tatarchenko Steady-State Model (TSSM)

TSSM is a model of impurity transport in the meniscus, which relates the values of k_{eff} to the parameters of capillary shaping and feeding, is presented in a book by Tatarchenko [40.46]. The stationary process is considered, and it is assumed that the meniscus and capillary melt have not been stirred, where the conditions of complete stirring are maintained in the crucible. Impurity transport in the meniscus is described by the equation

$$D\left(\frac{\mathrm{d}^2 c}{\mathrm{d}r^2} + \frac{1}{r}\frac{\mathrm{d}c}{\mathrm{d}r}\right) = -v\frac{\mathrm{d}c}{\mathrm{d}r} \tag{40.59}$$

with the following boundary condition at the melt–crystal interface

$$-D\left.\frac{\mathrm{d}c}{\mathrm{d}r}\right|_{r=r_0} = v_0\,(1-k)\,c(r_0)\,. \tag{40.60}$$

Here, c represents the impurity concentration in the melt, D is the diffusion coefficient, k is the partition

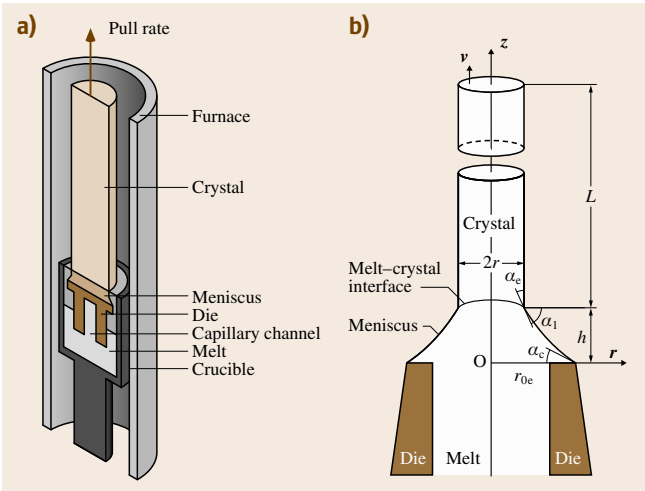


Fig. 40.4a,b Prototype EFG crystal growth system. **(a)** Prototype EFG system, **(b)** schematic EFG used in mathematical model

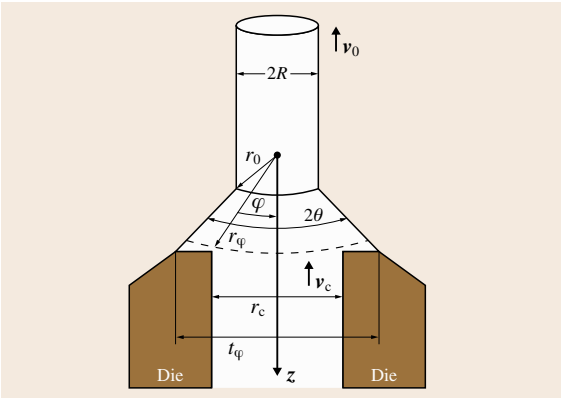


Fig. 40.5 EFG diagram for TSSM

coefficient at equilibrium, and the other notations are given in Fig. 40.5.

The melt flow rate in the meniscus is given by

$$v(r) = \frac{v_0 r_0}{r} \,. \tag{40.61}$$

Impurity transport in the capillary channel is described by the equation

$$D\frac{\mathrm{d}^2 c}{\mathrm{d}z^2} = -v_\mathrm{c}\frac{\mathrm{d}c}{\mathrm{d}z} \tag{40.62}$$

with the following boundary condition

$$c \to c_\infty \quad \text{if} \quad z \to \infty \,. \tag{40.63}$$

Employment of these equations and boundary conditions gives the following expression describing the impurity concentration distribution in the meniscus

$$c(r) = (1-k_\mathrm{eff})\, c\,(r_0)\left(\frac{r_0}{r}\right)^\mathrm{e} + \frac{c_\infty - (1-k)\left(\frac{r_0}{r_\Phi}\right)^\mathrm{e}(\cos\varphi - 1)\, c\,(r_0)}{1+\left[\left(\frac{v_0\theta r_0}{v_\mathrm{c}\varphi r_\Phi}-1\right)\right]\cos\varphi} \,, \tag{40.64}$$

where

$$\cos\varphi = \left(1-\frac{r_\mathrm{c}^2}{4r_\Phi^2}\right)^{1/2} \quad \text{and} \quad e = \frac{v_0 R}{D\sin\theta} \,, \quad \varphi \in [0,\ \Phi] \,. \tag{40.65}$$

Taking into account the equalities $k = c_\mathrm{S}/c(r_0)$, $k_\mathrm{eff} = c_\mathrm{S}/c_\infty$, from (40.64) for $r = r_0$ it follows that

$$k_\mathrm{eff}^{-1} = 1 + \left(\frac{v_0\theta r_0}{v_\mathrm{c}\varphi r_\Phi} - 1\right)\cos\varphi + \frac{(1-k)(1-\cos\varphi)\,(r_0/r_\Phi)^\mathrm{e}}{k} \,. \tag{40.66}$$

Using the relations

$$2v_0 R = v_\mathrm{c} r_\mathrm{c} \,, \quad \frac{r_0}{r_\Phi} = 2Rt_\varphi \,, \quad r_0 = \frac{R}{\sin\theta} \,, \tag{40.67}$$

(40.66) becomes

$$k_\mathrm{eff}^{-1} = 1 + \left(\frac{v_0 r_\mathrm{c}}{\varphi t_\varphi - 1}\right)\cos\varphi + \frac{(1-k)(1-\cos\varphi)\left(2R/t_\varphi\right)^{v_0 R/D\sin\theta}}{k} \,. \tag{40.68}$$

Equation (40.68) relates the value of the effective coefficient of partition k_eff with the process parameters: v_0, $2R$, t_Φ, r_c, θ, D, and k.

When the capillary channel radius its maximum value, $r_\mathrm{c} = t_\Phi$, (40.68) becomes

$$k_\mathrm{eff} = \frac{k}{k + (1-k)(1-\cos\varphi)\left(\frac{2R}{t_\Phi}\right)^{\frac{v_0 R}{D\sin\theta}}} \,. \tag{40.69}$$

According to TSSM, the impurity distribution in the crystal is uniform (i. e., no variations of the impurity concentration along the crystal length and cross section), and the value of the impurity concentration in the crystal is given by

$$c_\mathrm{c} = k_\mathrm{eff} c_\infty \,, \tag{40.70}$$

where k_eff is given by (40.68).

In order to verify the above prediction for Al impurities in Si ($k = 0.002$, $D = 5.3\times10^{-4}\,\mathrm{cm^2/s}$, $2R = 0.03\,\mathrm{cm}$, $t_\Phi = 0.06\,\mathrm{cm}$), the effective partition coefficient k_eff was computed. It was found that $k_\mathrm{eff} \ll 1$, where k_eff increases as both the growth rate and the melt flow in the capillary channel increase [40.46]. Mass-spectrographic analysis of the impurity content in tapes grown from refined metallurgical silicon confirms that $k_\mathrm{eff} \ll 1$. The sulfur distribution along the axis of a silicon tape obtained by laser emission analysis shows also that $k_\mathrm{eff} \ll 1$, in agreement with the value of k_eff given by (40.68). According to [40.46], in order to verify the TSSM-based prediction, a series of 0.01 mass indium-doped silicon tapes was grown, where the predicted values of k_eff are in agreement with those obtained experimentally. The values of k_eff given by Fleming's formula are several orders of magnitude greater than the experimental values. For the variation of the impurity concentration observed experimentally along the crystal cross section in [40.46], the following explanation was given: For the case of a horizontal melt flow in the meniscus from the die edge towards the center, impurities with $k_\mathrm{eff} < 1$ are driven off by the growing crystal and accumulate in the central part of the meniscus. Impurities may be moved away in order to improve the quality of the crystal. For this aim, the rejected impurities were studied using an axisymmetric or *displaced* die shape in an EFG system with melt replenishment [40.54, 55].

40.2.3 Melt Replenishment Model (MRM)

Recently, the influence of the die geometry and various growth conditions on the fluid flow and on the

solute distribution has been studied in an EFG system with melt replenishment (MR) using numerical simulation [40.56]. The nonuniform distribution of the voids (trapped gas bubbles) along the cross section of the sapphire rod was analyzed, assuming that the voids are impurities which are present in the melt and that their transport in the capillary channel and in the meniscus is governed by the stationary convection–diffusion equation. It was also assumed that the mechanism of transfer of the voids from the melt into the crystal is similar to the transfer of impurities. Two types of die designs were investigated: one with a central capillary channel (CCC) and another with an annular capillary channel (ACC). The fluid flow is characterized by a small Reynolds number (due to the small value of the fluid velocity and of the pulling rate) and consequently the fluid flow is laminar. Computations prove that, for both die geometries, the flow field near the meniscus free surface exhibits the same aspect, with only the velocity field values being different: for the ACC die, all the velocity values are higher than for the CCC die. For small Marangoni numbers (Ma), no loop or cell structure appears in the melt flow. As the Ma number increases, the loop position slightly migrates toward the crystallization interface. The isoconcentration field and solute concentration near the solidification interface are different for the two studied shaper designs. The solute fields for the CCC and ACC dies differ in aspect and concentration values: the concentration field for CCC die shows a single maximum close to the free meniscus surface; for the ACC die, a strong maximum that is two times higher appears near the symmetry axis. Moreover, the radial segregation is higher for the ACC die than for the CCC die and increases if the Ma number increases. These numerical results with good agreement with the experiments show that the gas contained in the melt acts as a solute rejected at the interface and nucleates into bubbles at the more concentrated locations. The experimental fact that the outer bubbles are situated at a distance from the external surface of the order of the meniscus height demonstrates that Marangoni convection exists in a shaped sapphire meniscus. Finally, in order to reduce the concentration of voids, a CCC die is recommended as well as growth with low meniscus height and low pulling rate.

For optimization and control of the growth process, in order to concentrate the voids close to the crystal surface which can be removed by polishing, the dependences of the fluid flow and impurity distribution on some process parameters (e.g., the pulling rate, radius of the capillary channel, natural convection, Marangoni convection, and vertical temperature gradient) have been studied [40.44, 57–60]. First, the dependences of the melt flow and impurity distribution on the pulling rate and on the capillary channel radius were studied; the temperature variations and their consequences were completely neglected (no Marangoni convection) [40.57].

The equations which define the model are the incompressible Navier–Stokes equations and the conservative convection–diffusion equation

$$\rho_l \frac{\partial \boldsymbol{u}}{\partial t} - \eta \nabla^2 \boldsymbol{u} + \rho_l (\boldsymbol{u}\nabla)\boldsymbol{u} + \nabla p = \boldsymbol{F}\,, \qquad (40.71)$$

$$\nabla \boldsymbol{u} = 0\,, \qquad (40.72)$$

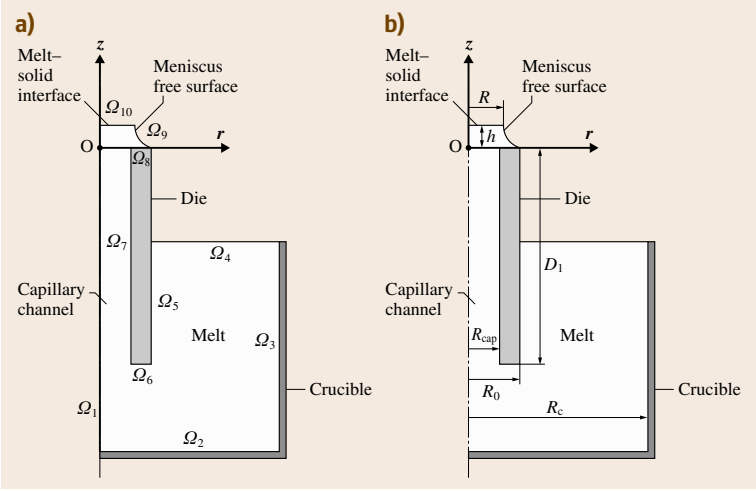


Fig. 40.6a,b EFG diagram in the case of MRM

$$\frac{\partial c}{\partial t} + \nabla(-D\nabla c + c\boldsymbol{u}) = 0\,, \tag{40.73}$$

written in a cylindrical polar coordinate system as shown in Fig. 40.4. In the above equations, $\boldsymbol{u} = (u_\mathrm{r},\ u_\mathrm{z})$ is the velocity vector, c is the impurity concentration; $\boldsymbol{F}$ is the volume force field due to gravity, which for zero gravity is $\boldsymbol{F} = \boldsymbol{0}$ and for terrestrial conditions is $\boldsymbol{F} = \rho_\mathrm{l}\boldsymbol{g}$ when: ρ_l is the melt density; $\boldsymbol{g}$ is the gravitational acceleration; p is the pressure; η is the dynamical viscosity; t is time; and D is the impurity diffusion. Axisymmetric solutions are sought in a two-dimensional domain with boundaries $\Omega_1 - \Omega_{10}$, as shown in Fig. 40.6.

On the boundaries $\Omega_1 - \Omega_{10}$, the following conditions are considered:

1. For the incompressible Navier–Stokes equations:

- Axial symmetry on Ω_1

$$u_\mathrm{r} = 0\,. \tag{40.74}$$

- Given inflow velocity on Ω_4

$$\boldsymbol{u} = -\frac{\rho_\mathrm{S}}{\rho_\mathrm{l}}\frac{R^2}{R_\mathrm{c}^2 - R_0^2}v\boldsymbol{k}\,, \tag{40.75}$$

where $\boldsymbol{k}$ represents the unit vector of the Oz-axis; this condition expresses continuous melt replenishment, i. e., the melt level in the crucible is maintained constant.

- Slip condition on Ω_9

$$\boldsymbol{u}\boldsymbol{n} = 0\,, \tag{40.76}$$

where $\boldsymbol{n}$ is the unit normal vector of Ω_9.

- Given outflow velocity on Ω_{10}

$$\boldsymbol{u} = v\boldsymbol{k}\,. \tag{40.77}$$

- Nonslip condition on $\Omega_2 U \Omega_3 U \Omega_5 U \Omega_6 U \Omega_7 U \Omega_8$

$$\boldsymbol{u} = \boldsymbol{0}\,. \tag{40.78}$$

2. For the conservative convection–diffusion equation

- Axial symmetry on Ω_1

$$\frac{\partial c}{\partial r} = 0\,. \tag{40.79}$$

- Impurity flux rejected into the melt on Ω_{10}

$$\frac{\partial c}{\partial \boldsymbol{n}} = -\frac{v}{D}(1 - K_0)\,c\,. \tag{40.80}$$

- No impurity flux, i. e., insulation on $\Omega_2 U \Omega_3 U \Omega_5\ U \Omega_6 U \Omega_7 U \Omega_8 U \Omega_9$

$$\frac{\partial c}{\partial \boldsymbol{n}} = \boldsymbol{n}\nabla c = 0\,. \tag{40.81}$$

- Given concentration on Ω_4

$$c = C_0\,. \tag{40.82}$$

The system (40.71–40.73) has a unique stationary solution, i. e., $\partial \boldsymbol{u}/\partial t = 0$ and $\partial c/\partial t = 0$, which satisfies the above boundary conditions. This describes the steady-state process.

In order to identify a nonstationary solution of the system (40.71–40.73) which describes a transient process, beside the above boundary conditions, it is necessary to specify an initial melt flow velocity

$$u_\mathrm{r}(t_0) = 0\,, \quad u_\mathrm{z}(t_0) = 0\,, \tag{40.83}$$

and the initial impurity distribution

$$c(t_0) = C_0\,. \tag{40.84}$$

The significance of these quantities and their values for the considered EFG system are given in Table 40.1 for Al-doped Si.

According to MRM, the value of the impurity concentration in the crystal is given by

$$c_\mathrm{c} = kc_\mathrm{interface}\,. \tag{40.85}$$

Using the right-hand side of (40.85) and the stationary solution of the boundary-value problem (40.71–40.82),

Table 40.1 Nomenclature

Nomenclature		Value
c	Impurity concentration (mol/m^3)	
C_0	Initial alloy concentration (mol/m^3)	
D	Impurity diffusion (m^2/s)	5.3×10^{-8}
D_l	Die length (m)	0.04
g	Gravitational acceleration (m/s^2)	9.81
h	Meniscus height (m)	0.5×10^{-3}
K_0	Partition coefficient	0.002
η	Dynamical viscosity (kg/(m s))	7×10^{-4}
p	Pressure (Pa)	0
$\mathrm{Pe}_\mathrm{mass} = \frac{vR_\mathrm{cap}}{D}$	Péclet number for mass transfer	
R	Crystal radius (m)	1.5×10^{-3}
R_cap	Capillary channel radius (m)	
R_c	Inner radius of the crucible (m)	23×10^{-3}
R_0	Die radius (m)	2×10^{-3}
ρ_l	Density of the melt (kg/m^3)	2500
ρ_s	Density of the crystal (kg/m^3)	2300
$\boldsymbol{u}$	Velocity vector	
v	Pulling rate (m/s)	
z	Coordinate in the pulling direction	

the following prediction can be obtained: constant impurity concentration along the crystal length and variation of the impurity concentration along the crystal cross section. The use of the right-hand side of (40.85) for a nonstationary solution of the boundary-value problem (40.71–40.84) leads to the prediction of the variation of the impurity concentration along the crystal length and along the crystal cross section during a transition period. In [40.44, 57], computations were made using COMSOL Multiphysics 3.2 software for zero gravity and for terrestrial conditions in the case of the following EFG system: crucible inner radius of $R_c = 23\times10^{-3}$ m, die radius of $R_0 = 2\times10^{-3}$ m, and die length of 40×10^{-3} m; two-thirds of the die immersed in the crucible melt; constant crucible melt height of 40×10^{-3} m; capillary channel radius of $R_{cap} = (0.5,\ 1,\ 1.5)\times10^{-3}$ m; and pulling rate v in the range 10^{-10} m/s $\le v \le 10^{-6}$ m/s. The meniscus height was $h = 0.5\times10^{-3}$ m, and the single-crystal rod radius (being grown) was $R = 1.5\times10^{-3}$ m.

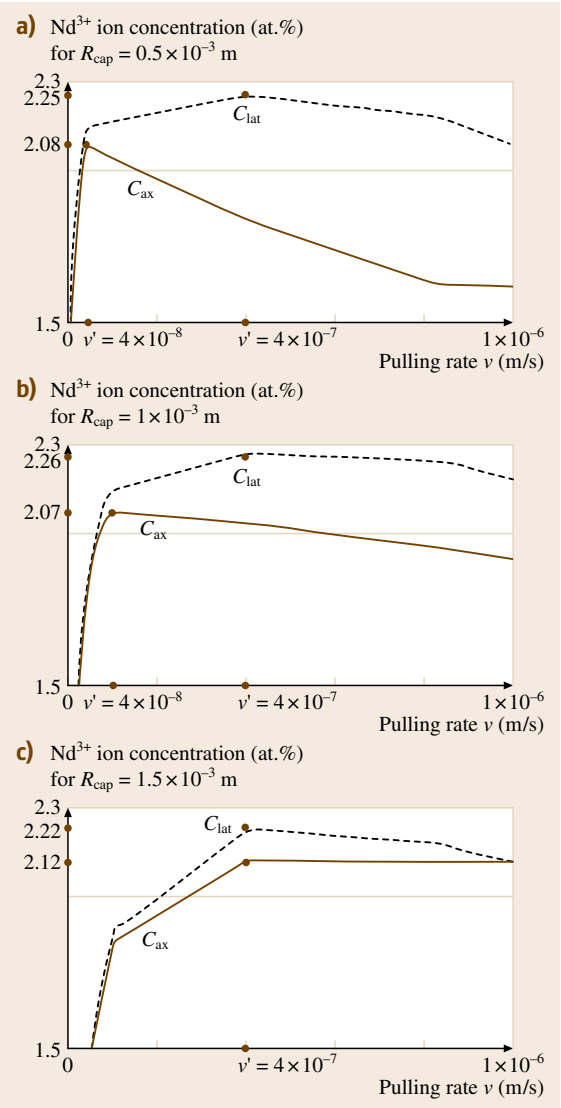


Fig. 40.7a–c Dependence of the Nd^{3+} ion concentration on the pulling rate v for **(a)** $R_{cap} = 0.5\times10^{-3}$ m, **(b)** $R_{cap} = 1\times10^{-3}$ m, and **(c)** $R_{cap} = 1.5\times10^{-3}$ m

The computations reveal that, in the stationary case, for a given R_{cap}, there exist critical pulling rate values v' depending on R_{cap} at which the impurity concentration has a maximum. If v increases in the range $[1\times10^{-10},\ v']$, the impurity concentration increases, and if v increases in $[v',\ 1\times10^{-6}]$, the concentration decreases slowly. Moreover, the impurity concentration on the axis and on the lateral surface of the crystal, and the absolute value of the difference between them – called the radial segregation – depend on the pulling rate and the capillary radius (Fig. 40.7). The optimal value of R_{cap} for which the crystal has the best homogeneity (radial segregation is minimum) is $R_{cap} = R = 1.5\times10^{-3}$ m.

Numerical results concerning a nonsteady state show that the nonstationary melt flow field and the impurity concentration field in the whole melt tend to the stationary melt flow field and to the stationary impurity concentration, respectively. The transition period, defined as that period of time after which the impurity concentration at the interface becomes equal to those obtained in the stationary case, depends on R_{cap} and v.

The impurity concentrations in the meniscus at three different instances of time $0 < t_1 < t_2 < t_3$ are presented in Fig. 40.8 for $R_{cap} = 1.5\times10^{-3}$ m and $v = 10^{-6}$ m/s.

These computed results are in agreement with experimental data reported by *Cao* et al. [40.53]: each impurity has a transient distance or crystal length until the steady-state segregation is established. The localization of the maximum of the dopant concentration on the free meniscus surface shows that the model is incomplete. More precisely, natural and Marangoni convections should be involved in order to *move* the maximum of the dopant concentration near the melt–solid interface, as was observed experimentally [40.56].

In [40.58–60] the combined effect of the buoyancy and Marangoni forces on the fluid flow and on the dopant distribution is analyzed in the stationary case. Thus, the growth process is described by the stationary incompressible Navier–Stokes equations written in the

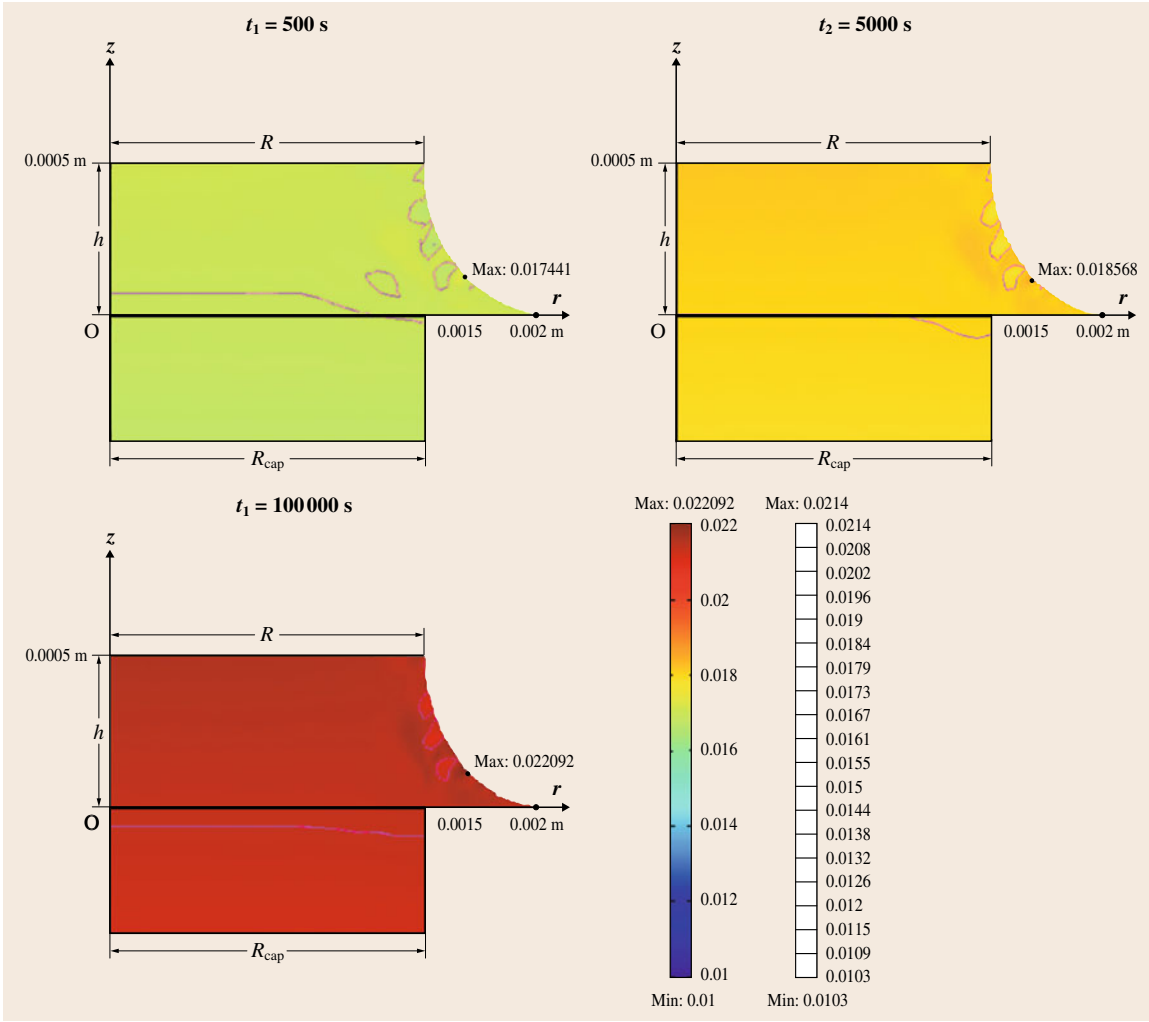


Fig. 40.8 Al concentration at three instances of time $t_1 < t_2 < t_3$, for $v = 10^{-6}$ m/s and $R_{cap} = 1.5 \times 10^{-3}$ m

Boussinesq approximation [i. e., $\boldsymbol{F} = \beta\rho_l \boldsymbol{g}(T - \Delta T)$, where β is heat expansion coefficient and $\Delta T = (T_0 + T_m)/2$ is the reference temperature at the free surface], and the convection–conduction and conservative convection–diffusion equations. For implementation of the Marangoni effect, the weak form of the boundary application mode is employed. The fluid density is assumed to vary with temperature as $\rho(r, z) = \rho_l[1 - \beta(T(r, z) - \Delta T)]$, and the surface tension γ in the meniscus is assumed to vary linearly with temperature as $\gamma(r, z) = \gamma_l + (\mathrm{d}\gamma/\mathrm{d}T)(T(r, z) - \Delta T)$, where γ_l is the surface tension at the temperature ΔT, and $\mathrm{d}\gamma/\mathrm{d}T$ is the rate of change of surface tension with the temperature.

For implementation of the Marangoni effect, the boundary condition

$$\eta \begin{pmatrix} t_r & t_z \end{pmatrix} \begin{pmatrix} 2\dfrac{\partial u}{\partial r} & \dfrac{\partial u}{\partial z} + \dfrac{\partial v}{\partial r} \\ \dfrac{\partial u}{\partial z} + \dfrac{\partial v}{\partial r} & 2\dfrac{\partial v}{\partial r} \end{pmatrix} \begin{pmatrix} n_r \\ n_z \end{pmatrix} = \frac{\mathrm{d}\gamma}{\mathrm{d}T} \left(\frac{\partial T}{\partial r} + \frac{\partial T}{\partial z} \right) , \tag{40.86}$$

is imposed on the free surface (meniscus). This expresses that the gradient velocity field along the meniscus is balanced by the shear stress, where $\boldsymbol{t} = (t_r, t_z)$ and $\boldsymbol{n} = (n_r, n_z)$. The sign of the rate $\mathrm{d}\gamma/\mathrm{d}T$, in general, depends on the material, with leading downward

($d\gamma/dT < 0$) or upward ($d\gamma/dT > 0$) flow on the free liquid surface.

The effect of the buoyancy and Marangoni forces on the fluid flow and on the impurity distribution were investigated numerically including both the cases of downward and upward flows on the liquid free surface (meniscus). Thus, two materials having negative and positive surface tension rates (Nd:$Y_3Al_5O_{12}$ (Nd:YAG) and Nd:$LiNbO_3$) were considered. The computed dopant distribution in the meniscus for different values of the Grashof and Marangoni numbers (Gr = 0.00022 and 0.02231, Ma = 0.049 and 4.913 for Nd:YAG, and Gr = 0.00252 and 0.25222, Ma = 0.671 and 67.072 for Nd:$LiNbO_3$), are illustrated in Figs. 40.9 and 40.10. Computations show that the dopant concentration increases from the initial value C_0 to a maximum value C_{max} situated at the level of the melt–crystal interface $z = h$, marked on Figs. 40.9 and 40.10. The maximum values of the Nd^{3+} ion concentration show that increases of the Marangoni and Grashof numbers lead to a decrease of C_{max} and can move it from the lateral surface of the crystal toward the axis. The effect of the Grashof number is smaller on the C_{max} localization and its values, and shows that the Marangoni convection has a dominant effect on the natural convection.

The buoyancy- and surface-driven flows perturb the forced flow: the arrows presented in Figs. 40.9 and 40.10 denote the flow of the velocity field caused by the surface-tension-driven flow, with downward flow on the meniscus surface for Nd:YAG ($d\gamma\ dT < 0$; Fig. 40.9), and upward flow on the meniscus surface for Nd:$LiNbO_3$ ($d\gamma/dT > 0$; Fig. 40.10). The maximum velocity of the fluid flow in the meniscus presents a linear dependence on the Marangoni number. More precisely, if the Marangoni number increases, then the maximum velocity of the fluid flow increases. Concerning the radial segregation, computations show that an increase of the Grashof or Marangoni number leads to a smaller radial segregation, clearly because of better mixing in these cases. The magnitude of the radial segregation is more sensitive to the Marangoni convection and less sensitive to the natural convection. The magnitude of the radial segregation changes are larger for larger Marangoni numbers.

Figure 40.10 shows a similar location of the maximum of the dopant distribution to those reported experimentally for sapphire [40.56]. The question here

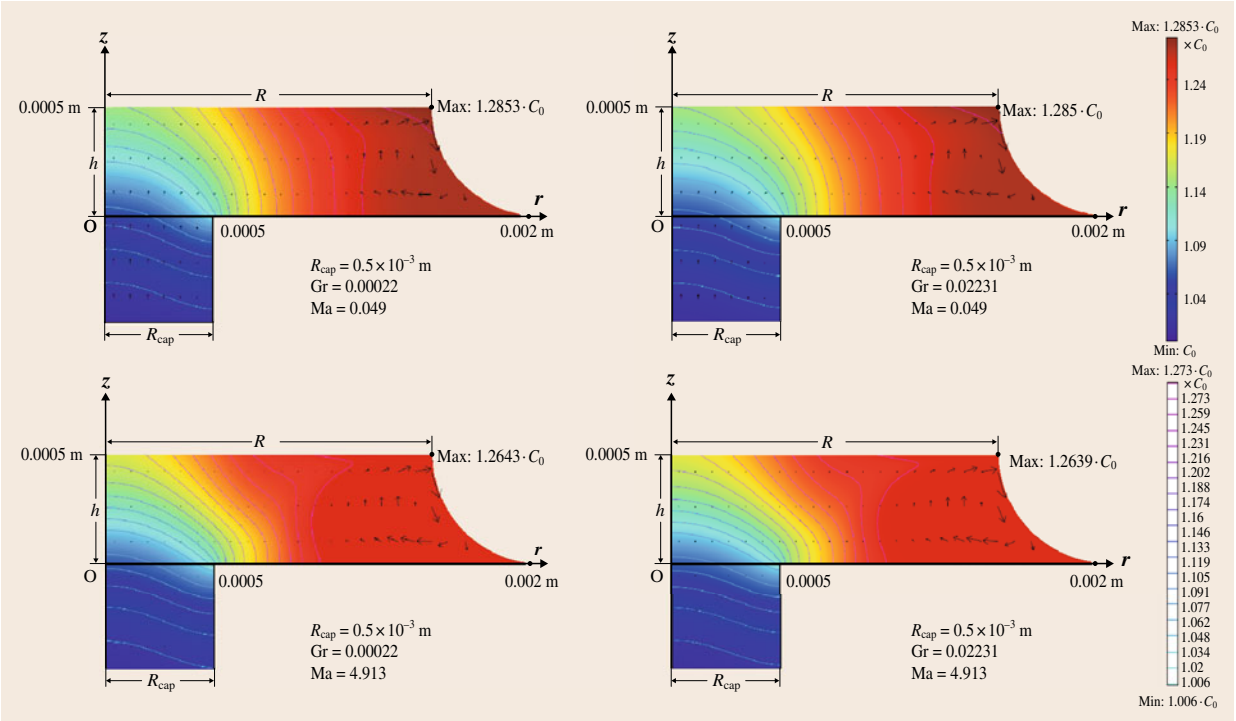


Fig. 40.9 Dependence of the dopant distribution on the Grashof and Marangoni numbers for Nd:YAG

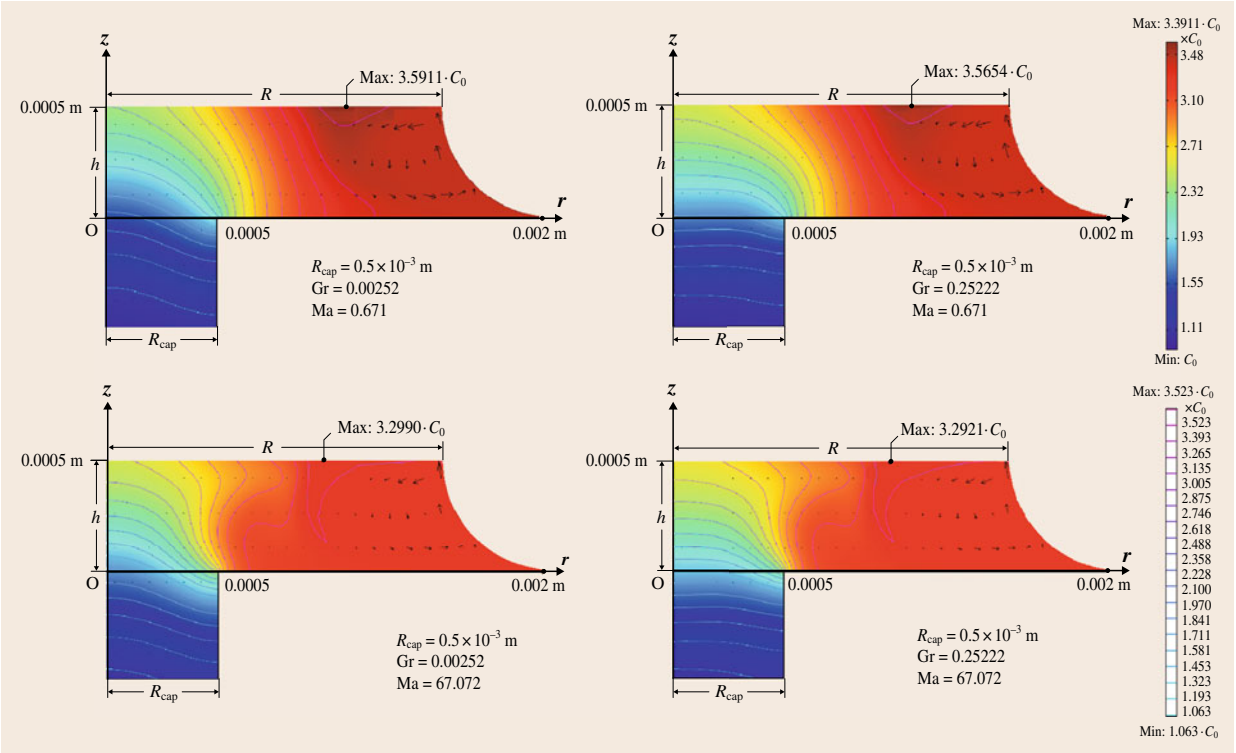


Fig. 40.10 Dependence of the dopant distribution on the Grashof and Marangoni numbers for Nd:$LiNbO_3$

is whether it is possible to determine conditions for which the maximum of the dopant distribution is located at the triple point, and at the distance from the lateral crystal surface equal to the meniscus height, respectively. There is also the question as to which case assures the best homogeneity of the crystal. The answers are found in the paper [40.60] in which the dependences of the Marangoni flow and impurity distribution on the vertical temperature gradient in sapphire (Al_2O_3) fibers are reported. Thus, for three representative values of the vertical temperature gradients k_g and for different Marangoni numbers Ma corresponding to the surface tension gradients $d\gamma/dT$ situated in the range $[-6\times10^{-5}, 0]$ N/(m K), the impurity distributions are computed for two cases: (i) the buoyancy is taken into account (the heat expansion coefficient has the value reported in the literature, i. e., $\beta = 3\times10^{-5}$ 1/K); and (ii) the buoyancy is neglected ($\beta = 0$), revealing the existence of three critical Marangoni numbers – Ma_{c1}, Ma_{c2}, Ma_{c3} – as reported for two-dimensional containers, determined by the behavior of the fluid flow. For Ma in the range [0, Ma_{c1}), the downward flow ($d\gamma/dT < 0$) on the free liquid surface leads to a steady flow, with a decrease of the maximum of the dopant concentration C_{max} located at the triple point if Ma increases. For Ma in the range [Ma_{c1}, Ma_{c2}), turbulence in the fluid flow appears, leading to an increase of C_{max}, still located at the triple point. For Ma in the range [Ma_{c2}, Ma_{c3}), this increased C_{max} is shifted inside the crystal from the melt–crystal interface at a distance on the same order as the meniscus height from the external crystal surface. If Ma is larger than Ma_{c3}, C_{max} is shifted into the center of the crystal.

We thus conclude that the best homogeneity of the crystal is assured in the range [0, Ma_{c1}) in which steady flow takes place (the maximum of the dopant distribution is located at the triple point). In this range, voids are situated at the lateral crystal surface and hence can be removed by polishing with minimum material loss. This range becomes larger if the vertical temperature gradient k_g and heat flux coefficient β decrease, which proves that a smaller vertical temperature gradient and a lower gravity, e.g., zero gravity, assures the best homogeneity of the crystal over a wide range of surface tension gradients. This suggests to practical crystal growers that a possible feedback control for delaying the Marangoni

convection can be obtained by decreasing the vertical temperature gradient in the furnace.

40.2.4 Melt Without Replenishment Model (MWRM)

In [40.61], a MWRM model in which the melt flow and impurity distribution are described in an EFG system without melt replenishment (the melt level in the crucible decreases during the growth process) is performed. It is assumed that temperature variations and their consequences in the system (no Marangoni convection) are negligible. The fluid flow and impurity distribution in the crucible, in the capillary channel, and in the meniscus are considered in a time-dependent domain $\Omega(t)$, $t \in [0, T]$ by the incompressible Navier–Stokes equations in terrestrial conditions (40.71–40.72) and the conservative convection–diffusion (40.73). For these equations, axisymmetric solutions are sought in the cylindrical polar coordinate system (rOz) (Fig. 40.11).

The evolution of $\Omega(t) = \{(r(R, Z, t), z(R, Z, t)): (R, Z) \in \Omega(0)\}$ is described by the system of partial differential equations corresponding to the Laplace smoothing (Poisson equations)

$$\begin{cases} \dfrac{\partial^2}{\partial R^2}\left(\dfrac{\partial r}{\partial t}\right) + \dfrac{\partial^2}{\partial Z^2}\left(\dfrac{\partial r}{\partial t}\right) = 0\,, \\ \dfrac{\partial^2}{\partial R^2}\left(\dfrac{\partial z}{\partial t}\right) + \dfrac{\partial^2}{\partial Z^2}\left(\dfrac{\partial z}{\partial t}\right) = 0\,. \end{cases} \tag{40.87}$$

Here, R, Z represent the reference coordinates in the reference frame (ROZ), i. e., the fixed frame used for the description of $\Omega(t)$ and of the mesh velocity, as depicted in Fig. 40.11a,b. The solution $(r(R, Z, t),\ z(R, Z, t))$ satisfies the condition $(r(R, Z, 0),\ z(R, Z, 0)) = (R, Z)$.

The coupled system (40.71–40.73, 40.87) is considered in the two-dimensional domain $\Omega(t)$ with boundaries $\Omega_1 - \Omega_{12}$, and for solving it, the arbitrary Lagrangian Eulerian (ALE) technique has been used. This technique for mesh movement [40.62, 63] is an intermediate between the Lagrangian and the Eulerian methods, combining the best features of both, i. e., it allows moving boundaries without the need for mesh movement to follow the material. The moving-mesh ALE application mode solves the system of partial differential equations (PDEs) (40.87) for the mesh displacement. This system smoothly deforms the mesh given by constraints on the boundaries. By the Laplace smoothing option (which has been chosen), the software introduces deformed mesh positions as degrees of freedom in the model.

In order to solve the system (40.71–40.73, 40.87), the following boundary conditions on the boundaries $\Omega_1 - \Omega_{12}$ are considered:

1. For the incompressible Navier–Stokes (NS) equations:
   - Axial symmetry on Ω_1

$$u_r = 0\,. \tag{40.88}$$

   - Given outflow velocity on Ω_3

$$\boldsymbol{u} = v\boldsymbol{k}\,. \tag{40.89}$$

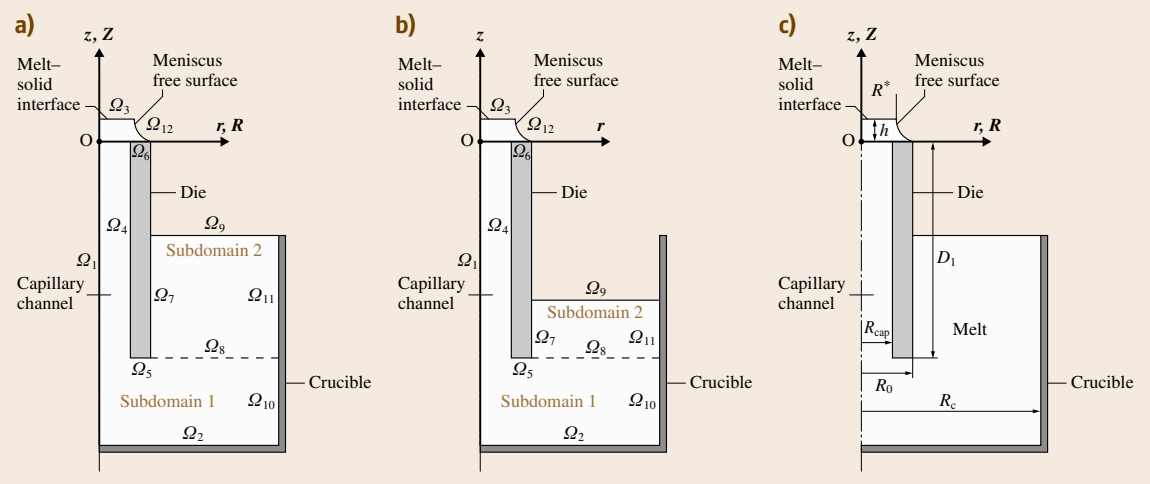


Fig. 40.11a–c EFG system and the boundary regions used in MWRM

- Nonslip condition on $\Omega_2 U \Omega_4 U \Omega_5 U \Omega_6 U \Omega_7 U \Omega_{10} U \Omega_{11} U \Omega_{12}$

$$\boldsymbol{u} = \boldsymbol{0} \,. \tag{40.90}$$

- Neutral condition on $\Omega_8 U \Omega_9$

$$[-p\mathrm{I} + \eta(\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^{\top})]\boldsymbol{n} = 0 \,. \tag{40.91}$$

2. For the moving-mesh ALE:
The domain $\Omega(t)$ is divided in two subdomains, 1 and 2 (Fig. 40.11). Within subdomain 1 there is no displacement (i. e., no motion in this subdomain), whereas within subdomain 2 there are free displacements (i. e., motion is free in this subdomain). Hence, the mesh displacement takes place only in subdomain 2 and it is constrained by the boundary conditions on the surrounding boundaries of Ω_7, Ω_8, Ω_9, and Ω_{11}. The displacement in subdomain 2 is obtained by solving the PDEs (40.87). The boundary conditions involve variables from the NS application mode. For convergence, it is important for the boundary conditions to be consistent. The usual pointwise constraints or ideal constraints for ALE cause unwanted modifications of the boundary conditions for the NS mode coupled with convection diffusion (CD). For this reason, nonideal weak constraints on the boundaries are used in the ALE application mode:
   - The mesh displacements in the r- and z-directions on Ω_9 are the following

$$\mathrm{d}r = 0 \,, \quad \mathrm{d}z = v_n t \,, \tag{40.92}$$

     where $v_n = -(\rho_S/\rho_l)R^{*2}/(R_c^2 - R_0^2)v$, and R^* is the crystal radius. (According to [40.62], the mesh velocity should be equal to the fluid velocity.)
   - The mesh displacement in the r-direction on $\Omega_7 U \Omega_{11}$ is $\mathrm{d}r = 0$; the mesh displacement $\mathrm{d}z$ in the z-direction is not specified, i. e., the mesh follows the fluid movement;
   - The mesh displacements $\mathrm{d}r$ and $\mathrm{d}z$ in the r- and z-direction are not specified on Ω_8 (the mesh follows the fluid flow).
3. For the conservative convection–diffusion equation:
   - Axial symmetry on Ω_1, i. e., $r = 0$.
   - Impurity flux rejected into the melt on Ω_3

$$\frac{\partial c}{\partial \boldsymbol{n}} = -\frac{v}{D}(1 - K_0)\, c \,. \tag{40.93}$$

   - No impurity flux, i. e., insulation on $\Omega_2 U \Omega_4 U \Omega_5 U \Omega_6 U \Omega_7 U \Omega_9 U \Omega_{10} U \Omega_{11} U \Omega_{12}$

$$\frac{\partial c}{\partial \boldsymbol{n}} = \boldsymbol{n} \nabla c = 0 \,. \tag{40.94}$$

   - Continuity on Ω_8

$$\boldsymbol{n}(\boldsymbol{N}_1 - \boldsymbol{N}_2) = 0, \quad \boldsymbol{N}_i = -D_i \nabla c_i + c_i \boldsymbol{u}_i \,, \tag{40.95}$$

     where $i = 1$ for subdomain 1 and $i = 2$ for subdomain 2. Besides the above boundary conditions, the followings initial conditions are also used:
   - For the fluid flow

$$u(t_0) = 0, \quad v(t_0) = 0 \tag{40.96}$$

     in subdomain 1, and

$$u(t_0) = 0, \quad v(t_0) = v_n \tag{40.97}$$

     in subdomain 2 (according to [40.62] the fluid velocity should be equal to the mesh velocity).
   - For the pressure

$$p(t_0) = P_0 = -\rho_l g z \tag{40.98}$$

     in subdomain 1, and

$$p(t_0) = 0 \tag{40.99}$$

     in subdomain 2.
   - For the initial impurity distribution

$$c(t_0) = C_0 \tag{40.100}$$

     in both subdomains.
   - For the mesh displacement

$$r(t_0) = R \,, \quad z(t_0) = Z \,. \tag{40.101}$$

Numerical investigations were carried out for an Al-doped Si rod of radius $R^* = 1.5 \times 10^{-3}$ m, grown in terrestrial conditions with a pulling rate $v = 10^{-6}$ m/s. According to [40.57], for $R_{cap} = R^* = 1.5 \times 10^{-3}$ m, the radial segregation in the Si rod grown in terrestrial conditions is minimal over the range $[10^{-7}, 5 \times 10^{-6}]$ m/s of v. In order to evaluate the way in which the resulting fluid flow and deformed geometry determine the impurity distribution in the melt and in the crystal, COMSOL Multiphysics 3.2 software was used in order to solve the coupled NS-ALE-CD application modes in the ALE frame (rOz). COMSOL Multiphysics does the mathematics necessary to manipulate, move, and deform the mesh simultaneously with the boundary movement as required by the other coupled NS-CD PDEs. The computed impurity distributions at three different instances of time are presented in Fig. 40.12.

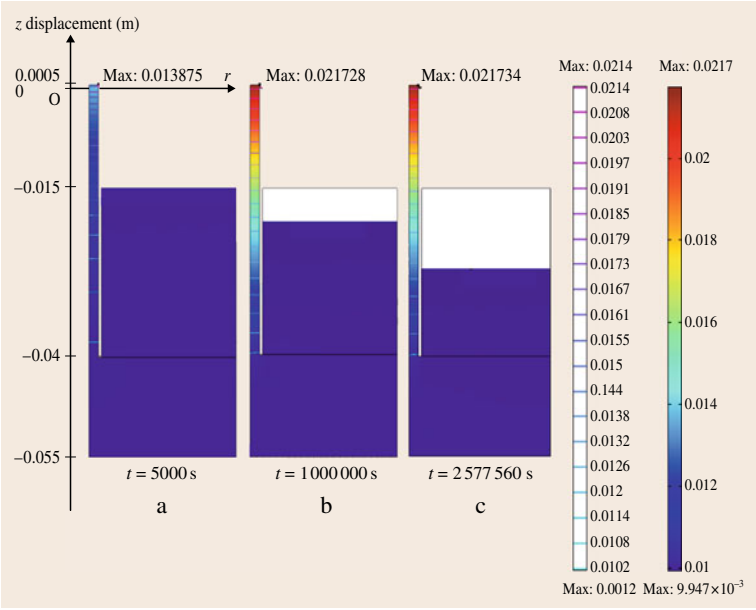


Fig. 40.12 Impurity distribution in the melt at three different instances of time $0 < t_1 < t_2 < t_3 < T$ for $R_{cap} = 1.5 \times 10^{-3}$ m and $v = 10^{-6}$ m/s for the case of no melt replenishment

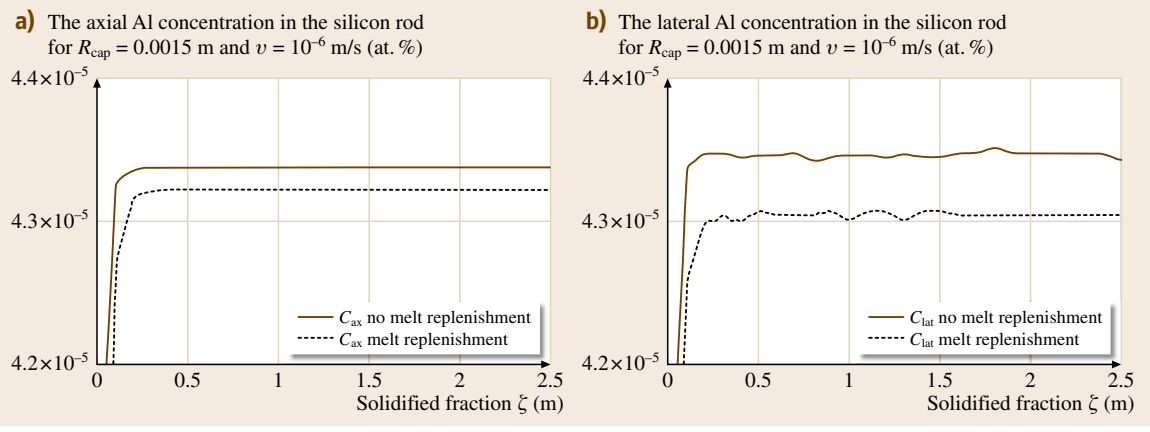


Fig. 40.13a,b Dependence of the impurity distributions on the axis **(a)** and on the lateral surface **(b)** of the crystal versus solidified fraction ξ for $R_{cap} = 1.5 \times 10^{-3}$ m and $v = 10^{-6}$ m/s, with and without melt replenishment

Figure 40.12a–c shows that, at the beginning, the concentration increases and there exists a certain time after which the impurity concentration becomes constant.

The computed impurity distributions versus solidified fraction along the crystal axis (C_{ax}) obtained with and without melt replenishment are presented in Fig. 40.13a. Similarly, the computed impurity distributions versus solidified fraction at the lateral surface (C_{lat}) of the crystal are presented in Fig. 40.13b.

Figure 40.13 shows that, after a transition period, the impurity distributions along the crystal axis (C_{ax}) and on the lateral surface (C_{lat}), computed in MRM and MWRM, become constant. In the MWRM case these constants are larger than those obtained in the MRM case: the difference is on the order of 10^{-7}–10^{-6} in the transition period, and after that, this difference is on the order of 1×10^{-7} on the crystal axis and 3×10^{-7} on the lateral surface. Moreover, Fig. 40.13 shows that the transition periods computed in MWRM are shorter. In MWRM the transition period for C_{ax} and C_{lat} is equal to 200 000 s. In MRM the transition period for C_{ax} is 300 000 s and for C_{lat} is 500 000 s.

References

40.1 J.A. Burton, R.C. Prim, W.P. Slichter: The distribution of solute on crystals grown from the melt, J. Chem. Phys. **21**, 1987–1996 (1953)

40.2 C. Wagner: Theoretical analysis of diffusion of solutes during the solidification of alloys, J. Met. **6**, 154–160 (1954)

40.3 K.M. Kim, A.F. Witt, M. Lichtensteiger, H.C. Gatos: Quantitative analysis of the thermo-hydrodynamic effects on crystal growth and segregation under destabilizing vertical thermal gradients: Ga-doped germanium, J. Electrochem. Soc. **125**, 475–480 (1974)

40.4 J.J. Favier: Macrosegregation-I: Unified analysis during non-steady state. Solidification, Acta Metall. **29**, 197–204 (1981)

40.5 J.J. Favier: Macrosegregation-II: A comparative study of theories, Acta Metall. **29**, 205–214 (1981)

40.6 D.E. Holmes, H.C. Gatos: Convective interference and "effective" diffusion-controlled segregation during directional solidification under stabilizing vertical thermal gradients; Ge, J. Electrochem. Soc. **128**, 429–437 (1981)

40.7 C.J. Chang, R.A. Brown: Radial segregation induced by natural convection and melt/solid interface shape in vertical Bridgman growth, J. Cryst. Growth **63**, 343–364 (1983)

40.8 W.A. Tiller, K.A. Jackson, J.W. Rutter, B. Chalmers: The redistribution of solute atoms during the solidification of metals, Acta Metall. **1**, 428–437 (1953)

40.9 P.R. Griffin, S. Motakef: Influence of non-steady gravity on natural convection during micro-gravity solidification of semiconductors. Part I. Time Scale Analysis, J. Appl. Microgravity II **3**, 121–127 (1989)

40.10 P.R. Griffin, S. Motakef: Influence of non-steady gravity on natural convection during micro-gravity solidification of semiconductors. Part II. Implications for crystal growth experiments, J. Appl. Microgravity II **3**, 128–132 (1989)

40.11 A.F. Witt, H.C. Gatos, M. Lichtensteiger, M.C. Lavine, C.J. Herman: Crystal growth and steady-state segregation under zero gravity, J. Electrochem. Soc. **122**, 276–283 (1975)

40.12 P.M. Adornato, R.A. Brown: Convection and segregation in directional solidification of dilute and non-dilute binary alloys, J. Cryst. Growth **80**, 155–190 (1987)

40.13 P.A. Clark, W.R. Wilcox: Influence of gravity on thermocapillary convection in floating zone melting of silicon, J. Cryst. Growth **50**, 461–469 (1980)

40.14 S.A.I. Nikitin, V. Polezhayev, A.I. Fedyushkin: Mathematical simulation of impurity distribution in crystals prepared under microgravity conditions, J. Cryst. Growth **52**, 471–477 (1981)

40.15 S.R. Coriell, R.F. Sekerka: Lateral solute segregation during unidirectional solidification of a binary alloy with a curved solid-liquid interface, J. Cryst. Growth **46**, 479–482 (1979)

40.16 S.R. Coriell, R.F. Boisvert, R.G. Rehm, R.F. Sekerka: Lateral solute segregation during unidirectional solidification of a binary alloy with a curved solid-liquid interface II. Large departures from planarity, J. Cryst. Growth **54**, 167–175 (1981)

40.17 H.M. Ettouney, R.A. Brown: Effect of heat transfer on melt/solid interface shape and solute segregation in edge-defined film-fed growth: Finite element analysis, J. Cryst. Growth **58**, 313–329 (1982)

40.18 J.P. Kalejs, L.Y. Chin, F.M. Carlson: Interface shape studies for silicon ribbon growth by the EFG technique I. Transport phenomena modeling, J. Cryst. Growth **61**, 473–484 (1983)

40.19 C.E. Chang, W.R. Wilcox: Control of interface shape in the vertical Bridgman-Stockbarger technique, J. Cryst. Growth **21**, 135–140 (1974)

40.20 T.W. Fu, W.R. Wilcox: Influence of insulation on stability of interface shape and position in the vertical Bridgman-Stockbarger technique, J. Cryst. Growth **48**, 416–424 (1980)

40.21 T.W. Clyne: Heat flow in controlled directional solidification of metals I. Experimental investigation, J. Cryst. Growth **50**, 684–690 (1980)

40.22 T.W. Clyne: Heat flow in controlled directional solidification of metals II. Mathematical model, J. Cryst. Growth **50**, 691–700 (1980)

40.23 P.C. Sukanek: Deviation of freezing rate from translation rate in the Bridgman-Stockbarger technique I. Very low translation rates, J. Cryst. Growth **58**, 208–218 (1982)

40.24 P.C. Sukanek: Deviation of freezing rate from translation rate in the Bridgman-Stockbarger technique II. Moderate translation rates, J. Cryst. Growth **58**, 219–228 (1982)

40.25 T. Jasinski, W.M. Rohsenow, A.F. Witt: Heat transfer analysis of the Bridgman-Stockbarger configuration for crystal growth I. Analytical treatment of the axial temperature profile, J. Cryst. Growth **61**, 339–354 (1983)

40.26 L.R. Morris, W.C. Winegard: The development of cells during the solidification dilute Pb-Sb alloy, J. Cryst. Growth **5**, 361–375 (1969)

40.27 A.F. Witt, H.C. Gatos, M. Lichtensteiger, C.J. Herman: Crystal growth and segregation under zero gravity, J. Electrochem. Soc. **125**, 1832–1840 (1978)

40.28 A.M. Balint, D.G. Baltean, T. Levy, M. Mihailovici, A. Neculae, S. Balint: The dopant fields in uniform-diffusion-layer, global-thermal-convection and precrystallization-zone models, Mater. Sci. Semicond. Process. **3**, 115–121 (2000)

40.29 D.T.J. Hurle, E. Jakeman, C.P. Johnson: Convective temperature oscillations in molten gallium, J. Fluid Mech. **64**, 565–576 (1974)

40.30 C.A. Wang, A.F. Witt: *Annual Report Material Processing Center* (Massachusetts Institute of Technology, Massachusetts 1984)

40.31 M.M. Mihailovici, A.M. Balint, S. Balint: The axial and radial segregation due to the thermoconvection, the decrease of the melt in the ampoule and the effect of the precrystallization-zone in the semiconductor crystals grown in a Bridgman-Stockbarger system in a low gravity environment, J. Cryst. Growth **237–239**, 1752–1756 (2002)

40.32 M.M. Mihailovici, A.M. Balint, S. Balint: On the controllability of the level of the dopant concentration and of the compositional uniformity of a doped crystal, grown in a low gravity environment by Bridgman-Stockbarger method, Int. J. Theor. Physics, Group Theory Nonlin. Opt. **10**, 425–436 (2003)

40.33 A.M. Balint, M.M. Mihailovici, D.G. Baltean, S. Balint: A modified Chang-Brown model for the determination of the dopant distribution in a Bridgman-Stockbarger semiconductor crystal growth system, J. Cryst. Growth **230**, 195–201 (2001)

40.34 A.M. Balint, M.M. Mihailovici, D.G. Baltean, S. Balint: Interface structure in the growth of semiconductor crystals using the Bridgman-Stockbarger method, Thin Solid Films **380**, 108–110 (2000)

40.35 M.M. Mihailovici, A.M. Balint, S. Balint: The dopant distribution computed in the modified Chang-Brown model using quasi-steady state approximation, Comput. Mater. Sci. **24**, 262–267 (2002)

40.36 D.G. Baltean, T. Levy, S. Balint: Transport de masse par convection et diffusion dans un milieu multiporeux, C. r. Acad. Sci. Paris, Ser. IIB **326**, 821–826 (1998)

40.37 K. Moutsopoulos, S. Bories: Dispersion en milieux poreux hétérogènes, C. R. Acad. Sci. Paris, Ser. IIB **316**, 1667–1672 (1993), in French

40.38 J.C. Maxwell: *Electricity and Magnetism* (Clarendon, Oxford 1873)

40.39 V.I. Avetisov, Mendeleev Institute Moscow (personal communication)

40.40 M.M. Mihailovici, A.M. Balint, S. Balint: Way to improve the compositional uniformity of doped crystals grown by Bridgman-Stockbarger method in a low gravity environment, Int. J. Theor. Phys. Group Theory Nonlinear Opt. **11**, 109–119 (2004)

40.41 D.J. Larson, J. Bethin, B.S. Dressler: *Shuttle Mission 51-G, Experiment MRS77F055, Flight Sample Characterization* (Grumman Corporate Research Center, Bethpage 1988), NASA Report RE-753

40.42 P.S. Dutta, A.G. Ostrogorsky: Segregation of Ga in Ge and InSb in GaSb, J. Cryst. Growth **217**, 360–365 (2000)

40.43 L.L. Zheng, D.J. Larson Jr., H. Zhang: Role of thermotransport (Sorret effect) in macrosegregation during eutectic/off-eutectic directional solidification, J. Cryst. Growth **191**, 243–251 (1998)

40.44 L. Braescu: The dependence of the dopant distribution on the pulling rate and on the capillary channel radius in the case of a Nd:YVO_4 cylindrical bar grown from the melt by the EFG method, Mater. Sci. Eng. B **146**, 41–44 (2008)

40.45 E. Tulcan-Paulescu, A.M. Balint, S. Balint: The effect of the initial dopant distribution in the melt on the axial compositional uniformity of a thin doped crystal grown in strictly zero-gravity environment by Bridgman-Stockbarger method, J. Cryst. Growth **247**, 313–319 (2003)

40.46 V.A. Tatarchenko: *Shaped Crystal Growth* (Kluwer, Dordecht 1993)

40.47 H.E. LaBelle Jr., A.I. Mlavsky, B. Chalmers: Growth of controlled profile crystals from the melt: Part I – Sapphire filaments, Mater. Res. Bull **6**, 571–579 (1971)

40.48 H.E. LaBelle Jr., A.I. Mlavsky, B. Chalmers: Growth of controlled profile crystals from the melt: Part II – Edge-defined, film-fed growth (EFG), Mater. Res. Bull. **6**, 581 (1971)

40.49 B. Chalmers, H.E. LaBelle Jr., A.I. Mlavsky: Edge-defined, film-fed crystal growth, J. Cryst. Growth **13/14**, 84–87 (1972)

40.50 J.P. Kalejs: Impurity redistribution in EFG, J. Cryst. Growth **44**, 329–344 (1978)

40.51 J.P. Kalejs, G.M. Freedman, F.V. Wald: Aluminium redistribution in EFG silicon ribbon, J. Cryst. Growth **48**, 74–84 (1980)

40.52 B. Chalmers: Transient solute effects in shaped crystal growth of silicon, J. Cryst. Growth **82**, 70–73 (1987)

40.53 J. Cao, M. Prince, J.P. Kalejs: Impurity transients in multiple crystal growth from a single crucible for EFG silicon octagons, J. Cryst. Growth **174**, 170–175 (1997)

40.54 J.P. Kalejs: Interface shape studies for silicon ribbon growth by the EFG technique II. Effect of die asymmetry, J. Cryst. Growth **61**, 485–493 (1983)

40.55 J.P. Kalejs: Modeling contribution in commercialization of silicon ribbon growth from the melt, J. Cryst. Growth **230**, 10–21 (2001)

40.56 O. Bunoiu, I. Nicoara, J.L. Santailler, T. Duffar: Fluid flow and solute segregation in EFG crystal growth process, J. Cryst. Growth **275**, 799–805 (2005)

40.57 L. Braescu, S. Balint, L. Tanasie: Numerical studies concerning the dependence of the impurity distribution on the pulling rate and on the radius of the capillary channel in the case of a thin rod grown from the melt by edge-defined film-fed growth (EFG) method, J. Cryst. Growth **291**, 52–59 (2006)

40.58 L. Braescu, T.F. George, S. Balint: Evaluation and control of the dopant distribution in a Nd:$LiNbO_3$ fiber grown from the melt by the edge-defined

film-fed growth (EFG) method, Photonic Fiber and Crystal Devices: Advances in Materials and Innovations in Device Applications I (Optics and Photonics 2007), SPIE Proc. **6698**, 669803:1–8 (2007)

40.59 L. Braescu, T. Duffar: Effect of buoyancy and Marangoni forces on the dopant distribution in the case of a single crystal fiber grown from the melt by the edge-defined film-fed growth (EFG) method, J. Cryst. Growth **310**, 484–489 (2008)

40.60 T.F. George, L. Braescu: Sapphire fibers grown from the melt by the EFG technique: Dependence of the impurity distribution on temperature and surface tension gradients, Photonic Fiber and Crystal Devices: Advances in Materials and Innovations in Device Applications II (Optics and Photonics 2008), SPIE Proc. **7056**, 705603–1–705603–10 (2008)

40.61 L. Braescu, T.F. George: Arbitrary Lagrangian-Eulerian method for coupled Navier-Stokes and convection-diffusion equations with moving boundaries, Proc. 12th WSEAS Int. Conf. Appl. Math. – Math'07 (2007) pp. 33–36

40.62 F. Duarte, R. Gormaz, S. Natesan: Arbitrary Lagrangian-Eulerian method for Naver-Stokes equations with moving boundaries, Comput. Methods Appl. Mech. Eng. **193**, 4819–4836 (2004)

40.63 M. Fernandez, M. Moubachir: Sensitivity analysis for an incompressible aeroelastic system, Math. Models Methods Appl. Sci. **12**, 1109–1130 (2002)

Part G Defects C

Part G Defects Characterization and Techniques

41 Crystalline Layer Structures with X-Ray Diffractometry
Paul F. Fewster, Brighton, UK

42 X-Ray Topography Techniques for Defect Characterization of Crystals
Balaji Raghothamachar, Stony Brook, USA
Michael Dudley, Stony Brook, USA
Govindhan Dhanaraj, Nashua, USA

43 Defect-Selective Etching of Semiconductors
Jan L. Weyher, Warsaw, Poland
John J. Kelly, Utrecht, The Netherlands

44 Transmission Electron Microscopy Characterization of Crystals
Jie Bai, Hillsboro, USA
Shixin Wang, Boise, USA
Lu-Min Wang, Ann Arbor, USA
Michael Dudley, Stony Brook, USA

45 Electron Paramagnetic Resonance Characterization of Point Defects
Mary E. Zvanut, Birmingham, USA

46 Defect Characterization in Semiconductors with Positron Annihilation Spectroscopy
Filip Tuomisto, Espoo, Finland

41. Crystalline Layer Structures with X-Ray Diffractometry

Paul F. Fewster

X-ray scattering analysis and instrumentation has been evolving rapidly to meet the demands arising from the growth of sophisticated device structures. X-ray scattering is very sensitive to composition, thickness and defects in layered structures of typical present-day electronic device dimensions. Considerable information can be obtained from simple profiles, including an estimate of layer thickness and composition (by measuring peak separations) and a measure of the sample quality (from the peak broadening). The full simulation of the profiles takes this a stage further to interpret very complex structures and obtain more reliable parameter estimates. By obtaining two-dimensional scattering data the information becomes more extensive, including layer strain relaxation and defect analysis, quantum dot composition and shape. Generally the data is averaged over a few mm, however reducing the beam size can break this averaging to reveal inhomogeneities, isolating small regions and in some circumstances isolate individual quantum dots for analysis. This article gives an overview of the status, differentiating those methods that are easily accessible and those that require a collaborative approach because they are still being established.

41.1 **X-Ray Diffractometry** 1406

41.2 **Basic Direct X-Ray Diffraction Analysis from Layered Structures** 1407
41.2.1 Theory 1407
41.2.2 Interpretation of Data Collected from Planes Parallel to the Surface – the ω–2θ Scan – an Example 1409
41.2.3 Interpretation of Data Collected from Several Reflections – The Reciprocal Space Map – An Example 1411

41.3 **Instrumental and Theoretical Considerations** 1412
41.3.1 The Instrument for Collecting X-Ray Diffraction Patterns 1412
41.3.2 Interpreting the Scattering by Simulation 1412

41.4 **Examples of Analysis from Low to High Complexity** 1413
41.4.1 Established Methods 1413
41.4.2 New Methods and New Analyses 1416

41.5 **Rapid Analysis** 1419

41.6 **Wafer Micromapping** 1420

41.7 **The Future** 1421

References 1422

X-ray scattering offers a nondestructive method for determining phase composition in crystalline materials, layer thickness, interface details, shapes of crystallites and quantum dots, etc. The basic principles of the various techniques will be outlined in this chapter. Bulk materials are limited to their intrinsic properties, whereas composite materials of several thin layers extend this range, e.g., semiconductor light-emitting diodes and modern transistors. However these physical properties are influenced by the layer compositions, thickness, defect density, and other structural parameters. X-ray scattering can be used very effectively for measuring all these structural parameters, some to very significant precision. The first part of this chapter discusses the general principles of the analysis, which is applicable to achieving an initial sample characterization by direct interpretation of the diffraction pattern. The limitations of this direct interpretation are also considered and how simulation of the scattering process gives more exact results. The advantages of simulation make the analysis of complex layer structures possible and also extend the applicability of x-ray scattering to

evaluating interface roughness, distortions, etc. Quantum dots are singled out as a particular area to illustrate the developments in x-ray scattering. Any analysis introduces some form of averaging, and approaches to reduce the probed volume are discussed. This has enabled detailed mapping of lateral variations, yielding a wealth of information for detailed characterization and screening, as discussed in the penultimate section.

41.1 X-Ray Diffractometry

The measured x-ray intensity scattered from a sample, in terms of magnitude and distribution, is sensitive to the arrangement and type of atoms in a crystal. If the x-ray probe is very well collimated then the resulting *high-angular-resolution* mode, or more generally phrased *high-resolution x-ray diffraction*, is very sensitive to small changes in the atomic layer spacing and deposited layer thickness. This is a consequence of the angle through which x-rays are scattered, which is inversely related to the separation of the atomic layer planes and the deposited layer thickness. So the sensitivity of x-ray scattering is effectively very high for very thin layers. Fortunately this is the region of interest for many of today's device structures; however, this is only relevant if there is sufficient scattered intensity. This aspect of the compromise between scattered intensity and high resolution will be addressed in Sect. 41.5. To indicate the bounds, it is possible to measure the thickness of layers up to $\approx 4\,\mu\text{m}$ and down to single atomic layers or a single plane of quantum dots using conventional configurations and equipment. For example, a single 1 nm $In_{0.1}Ga_{0.9}As$ layer on GaAs can be observed as an asymmetry on the main substrate profile, and determined by simulation to within ± 0.1 nm; a layer thickness above this value begins to indicate distinctive oscillations, and below this it is possible to observe this asymmetry but in practice it requires considerable care to determine the thickness. If however the layer is buried, then thinner layers of 0.1 nm can be observed by virtue of the interference between the layers above and below. The measurement of the thickness, again by simulation, appears simple and sensitive, however the thickness is correlated with the composition, and so ideally a combination of profiles are required. Any increase in thickness at constant composition increases the amplitude of the interference oscillations. The positions of the main Bragg peaks give a measure of the interplanar spacing in a crystal that can be determined to within ≈ 1 ppm, i. e., considerably less than the classical radius of an electron. However extracting dilute composition information in alloy systems requires a combination of differences in covalent radii and the concentration itself. So, to obtain the alloy composition of Al in GaAs, for example, where the difference in the covalent radius of Al to Ga is ≈ 2865 ppm, we require a method for differentiating the interatomic spacing to 28 ppm to achieve 1% precision in the concentration of Al in AlGaAs. With careful measurement and accounting for sample uncertainties this can be done routinely. For systems with larger differences in covalent radii, e.g., Ge in SiGe or In in GaN, the precision increases for the same amount of care in the measurement.

For obtaining the thickness of a layer, similar rules apply; as the thickness decreases, the width of the associated peak increases and the fringes associated with the interference of the scattering from the top and bottom surface of the layer become further apart. If the fringes are very close together and the central peak is narrow then the thickness measure becomes more problematic and requires higher-resolution instrumentation, up to the limit of the x-ray coherence length (typically $\approx 4\,\mu\text{m}$) or the layer perfection. Rough top and bottom surfaces reduce the intensity of the fringes and so can give a measure of the interface quality. As the thickness is reduced the peak broadens until eventually the scattering is indistinguishable from the inherent noise level of the instrument. If layers are stacked as in most device structures then thin layers modulate the overall pattern and appear more obvious again, e.g., interfacial layers in superlattice structures or single atomic layers separated by thicker layers. It is clear from this that x-ray scattering cannot simply be categorized in terms of what it can and cannot detect in terms of a single layer, for example, but rather the whole structure and the instrument are necessarily intertwined, and the route to extracting useful information generally comes down to simulating the scattering. However, some very useful analyses can be obtained quite directly from the measurements, although awareness of the pitfalls is important.

As far as instrument development is concerned, this can be a large and confusing subject, since as discussed above, the sample, the information to be extracted, and the instrument resolution are all linked. This has led to

instruments becoming more flexible for general materials research, and as the level of understanding or the improvement in control over the crystal growth develops, the resolution often needs to be increased. This allows more detailed information, deviations from perfection, etc., to be probed more easily. Similarly the analysis goes through a cycle of approximate quick interpretation, through to more precise simulation to understand the sample more fully, until eventually the analysis becomes a control feedback method that can be routine yet sophisticated and specific to the crystal properties of interest. When the growth of the material becomes too complex, in that there are many complex parameters to assess with high degrees of correlation, then some statistical significance is required. X-ray scattering can contribute here, since as illustrated above its high sensitivity to structural parameters can give some form of measure for obtaining correlations. A few examples of this will be given in this chapter.

As mentioned above x-ray scattering is very sensitive to structural changes and generally the probe is large, usually mm and for some specialized applications down to μm. However, this does not mean that everything is averaged over that region. In fact the x-ray pattern contains the average and the fluctuations from the average for dimensions over which the x-rays are coherent, so for example, the onset of layer relaxation (the distortion around dislocations) and the existence of precipitates can be observed, both of which modify the scattering in quite characteristic ways. Examining the spatial distribution of the intensity in the scattered beam can also break this averaging effect. Chapter 42 in this volume, describe a very direct approach to obtaining defect information, and when combined with high-resolution reciprocal space mapping, the x-ray topographic contrast can be related to features in the scattering pattern [41.1, 2].

This chapter will cover the rapid direct measures that give an indication of the material's properties. It will also show how far the analysis can be taken and finally the more systematic routine analysis methods for controlled crystal growth.

41.2 Basic Direct X-Ray Diffraction Analysis from Layered Structures

41.2.1 Theory

X-rays scatter strongly from electrons, principally those that are highly localized, i. e., core electrons. The scattering amplitude from an assembly of atoms at positions $\boldsymbol{r}$ with individual scattering factors (or form factors) of f_r is given by

$$F_\mathrm{S} = \int_r f_\mathrm{r} \mathrm{e}^{-2\pi r \boldsymbol{S}\cdot\boldsymbol{r}} \,\mathrm{d}\boldsymbol{r} \,. \tag{41.1}$$

The scattering amplitude and therefore the consequent intensity is distributed but reaches a maximum for atom positions satisfying the condition

$$|\boldsymbol{S}| = \frac{1}{d} \,, \tag{41.2}$$

where d is the separation of the atomic planes. Any position in diffraction space $\boldsymbol{S}$ can be probed by manipulating the incoming ($\boldsymbol{k}_0$) and the outgoing ($\boldsymbol{k}_\mathrm{H}$) x-ray beam paths, where $|\boldsymbol{k}_0| = |\boldsymbol{k}_\mathrm{H}| = 1/\lambda$, such that

$$\boldsymbol{S} = \boldsymbol{k}_\mathrm{H} - \boldsymbol{k}_0 \,, \tag{41.3}$$

which is Bragg's relation in vector form. It can be seen from the geometry of Fig. 41.1 that

$$S = \frac{1}{d} = \frac{2\sin\theta}{\lambda} \,. \tag{41.4}$$

Therefore, to probe different regions of diffraction space, the incident beam direction with respect to the sample and the detector acceptance direction are manip-

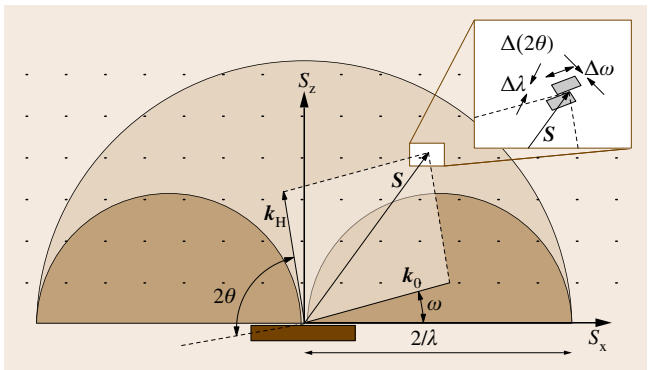


Fig. 41.1 The geometry of reciprocal space, indicating the incident and scattered beam vectors $\boldsymbol{k}_0$ and $\boldsymbol{k}_\mathrm{H}$ and the resultant position probed $\boldsymbol{S}$. The regular array of reciprocal lattice points represents the atomic plane spacings, and the limiting sphere for their observation is defined by the wavelength λ. The two smaller darker grey spheres are inaccessible regions to data collection in conventional reflection mode. The *inset* gives the detail of the size of the reciprocal space probe

ulated to map the distribution of intensity (Fig. 41.1). If the incident and scattered beam directions are very well defined, as with a high-angular-resolution diffractometer, then the length d can be measured very precisely as well as small differences in d. High resolution is very useful for studying closely lattice-matched semiconductor multilayers, for example.

The scattering amplitude F_S is complex, includes aspects of absorption, and is directly related to the refractive index. For most materials the refractive index for x-rays is less than one, and therefore x-rays total-externally reflect, and the actual Bragg scattering peaks are displaced to larger angles. The refractive index is $\approx (1-10^{-5})$, which has advantages in that for most measurements only small uncertainties are introduced, e.g., in thickness measurement from fringe spacings or lattice parameter determination. On the other hand the difference in refractive indices of layers from different materials creates modulations in the specular scattering (total external scattering), which offers another method for determining thickness and composition.

So the simplest information that can be extracted from an x-ray scattering experiment is the location of the peaks; this gives the crystal plane spacing, and their intensity relates to the atomic arrangement. The width of the peak relates to the number of contributing atomic planes, so in general for thin films the peaks become broadened normal to the surface direction, but very narrow parallel to the surface direction, so the term reciprocal space describes it all. This is the general guide relevant to the early stages of analyzing x-ray diffraction profiles.

Consider the early stages of crystal growth of a new material when little is known. Suppose that it consists of some layers and the intention is to grow something epitaxially related to the thick substrate wafer. The structure could appear as in Fig. 41.2a, where some of the likely measurable features are illustrated. A reciprocal space map for a single-layer structure around a reflection close to $s_x = 0$ may appear as in Fig. 41.2b. If the substrate is a reasonable quality single crystal, then the scattering from this will be represented by a narrow peak, and if the layers are not too thick ($< 2\,\mu$m), then this could be the highest peak and can be easily identified. Its position will give a value for the interplanar spacing, and the associated lattice parameter can be obtained if the reflection index H is known, since this relates the length scale d probed to that of the material lattice parameter. The layer peak is broader and wider and is not in line with the substrate peak. If the choice of reflection is from planes parallel to the surface (i. e., $\omega \approx \theta, s_x = 0$; Fig. 41.1), then this difference in angle $\delta\omega$ is indicative of layer planes being tilted with respect to the substrate planes by $\delta\omega$. The width of the layer peak in the direction normal to the surface (s_z) will be related to the layer thickness L_z or some variation of strain in the layer, the microstrain $\varepsilon_\perp$, or a combination of both. This width is obtained by scanning in ω and 2θ, coupled in a 1 : 2 ratio. If this width is purely due to the layer thickness then $L_z \equiv 1/\Delta s_z$, and from the geometry of Fig. 41.1

$$\frac{1}{\Delta s_z} = \frac{\lambda}{[\sin\omega_1 + \sin(2\theta_1 - \omega_1)] - [\sin\omega_2 + \sin(2\theta_2 - \omega_2)]} . \quad (41.5)$$

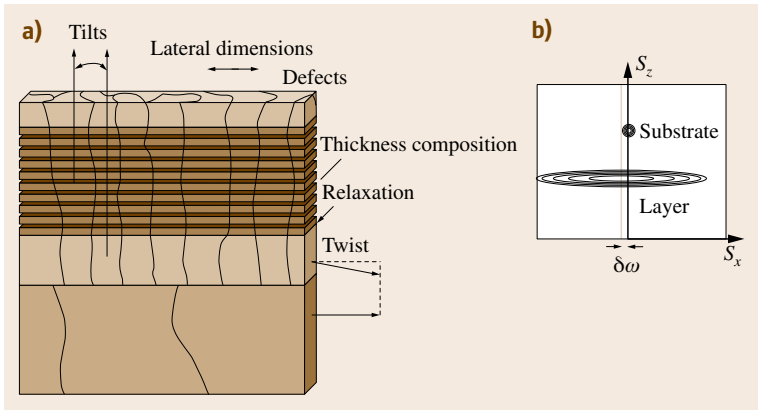


Fig. 41.2a,b The typical information that can be extracted from an x-ray scattering experiment **(a)**, and a schematic of a reciprocal space map of an imperfect layer on a nearly perfect substrate **(b)**

In this case $2\theta_1 \approx 2\omega_1$ and $2\theta_2 \approx 2\omega_2$, $\Delta 2\theta = (2\theta_1 - 2\theta_2)$, and the chosen positions described by the subscripts correspond to the full-width at half-maximum (FWHM), hence

$$L_z \approx \frac{\lambda}{2\Delta\theta\cos\theta}\,, \tag{41.6}$$

which is the Scherrer equation [41.3]. If the peak is related to the microstrain then, from (41.2),

$$\frac{\Delta d}{d} = -\frac{\Delta S}{S}\,. \tag{41.7}$$

Hence for the scattering peak considered here

$$\frac{\Delta d}{d} = -\frac{\Delta s_z}{s_z} = -\frac{\lambda/(2\sin\theta)}{\lambda/(2\Delta\theta\cos\theta)} = -\cot\theta\Delta\theta\,. \tag{41.8}$$

Separating these two influences on the peak width can either be judged by the acceptable ranges for the microstrain or layer thickness, or by combining several reflections (preferably along the same zone axis, i. e., 001, 002, 003, etc.), to separate the different dependency on θ [41.4].

The broadening normal to this also has two components. These two components are composed of local tilts and some lateral dimension: for mosaic layers these can be related to the Burgers vectors of the dislocations forming the mosaic blocks. If the layer is dominated by threading dislocations then the component of tilt is small and the width will give a good estimate of the mosaic block dimension, $L_x \equiv 1/\Delta s_x$, given by

$$\frac{1}{\Delta s_x} = \frac{\lambda}{[\cos\omega_1 - \cos(2\theta_1 - \omega_1)] - [\cos\omega_2 - \cos(2\theta_2 - \omega_2)]}\,, \tag{41.9}$$

where in this case $2\theta_1 \approx 2\theta_2$ and the peak width $\Delta\omega = (\omega_1 - \omega_2)$, thus

$$L_x \approx \frac{\lambda}{2\Delta\omega\sin\omega}\,. \tag{41.10}$$

If there is a high density of misfit dislocations that have Burgers vectors out of the plane of the interface, e.g., 60° dislocations in Si, then these will have a tilt component. The distribution of tilted regions and the size of these regions start to produce a more complex picture. However if the contribution is purely due to tilted regions of large dimensions then the spread of microscopic tilts is simply given by $\Delta\omega$.

41.2.2 Interpretation of Data Collected from Planes Parallel to the Surface – the ω–2θ Scan – an Example

Consider the example of a GaN structure given in Fig. 41.3a and the reciprocal space map around the 0002 reflection (Fig. 41.3b). The reciprocal space map has many satellites associated with this periodic layer structure and two strong peaks having two very different widths along s_z and similar widths along s_x (Fig. 41.3c,d). A considerable amount of information can be extracted from this data. Firstly, the separation of the two peaks (Fig. 41.3c) relates to the difference in lattice parameter along s_z. The individual lattice parameter value along this direction is given by

$$a_z = d_z H_z = \frac{H_z}{s_z}\,, \tag{41.11}$$

where H_z is the index of the reflection along the surface normal direction, simply equal to 2 here. Therefore the most precise measure of the lattice parameter is obtained for larger values of s_z, as is also clear from (41.8). Referring to Fig. 41.3c, we could in the first instance assume that the underlying substrate, or a thick GaN buffer in this case, has a known lattice parameter (some database value) and therefore the layer lattice parameter ${}_{\mathrm{L}}a_z$ is simply given by combining (41.11) and (41.8) as

$$\begin{aligned} {}_{\mathrm{L}}a_z &= \frac{{}_{\mathrm{S}}a_z}{{}_{\mathrm{S}}H_z}\left(1 + \frac{\Delta d}{d}\right){}_{\mathrm{L}}H_z \\ &\approx \frac{{}_{\mathrm{S}}a_z}{{}_{\mathrm{S}}H_z}\left(1 - \cot\theta\Delta\theta\right){}_{\mathrm{L}}H_z\,, \end{aligned} \tag{41.12}$$

where $\Delta\theta$ is the difference in Bragg angle and ${}_{\mathrm{L}}H_z$ is the reflection index in the direction normal to the surface for the layer and ${}_{\mathrm{S}}H_z$ for the substrate. For the example ${}_{\mathrm{L}}H_z = {}_{\mathrm{S}}H_z = 2$, ${}_{\mathrm{S}}a_z = 0.51851\,\mathrm{nm}$ for the 0002 reflection from GaN, $\theta_{\mathrm{S}} = 17.284°$ for an x-ray wavelength of $\lambda = 0.1540593\,\mathrm{nm}$, using (41.4). Since the measurement gives $\Delta\theta = 0.0777°$, and if we assume that the layer has a good epitaxial relationship to the substrate or layer below, the lateral lattice parameter of the layer ${}_{\mathrm{L}}a_x$ is identical to ${}_{\mathrm{S}}a_x$ (0.31893 nm), so ${}_{\mathrm{L}}a_z = 0.52076\,\mathrm{nm}$.

Now the layer material would have a characteristic and usually unique set of lattice parameters, but in the undistorted state. The unstrained lattice parameters can be determined from taking the distorted unit cell (defined by the lattice parameters) and allowing it to relax to the expected shape. In this case InN and GaN and the phase mixture are hexagonal, such that certain unit cell parameters (a, b, c, α, β, γ) are defined, i. e., $a = b$,

$\alpha = \beta = 90°$, and $\gamma = 120°$. The layer structure in the unstrained state is then just defined by the ratio c/a, which is 1.640 for InN and 1.626 for GaN. The way in which the distorted unit cell is transformed into the unstrained unit cell depends on the elastic coefficients that relate the stress and strains. The strain normal to the surface ε_{zz} is related to strains parallel to the surface, ε_{xx} and ε_{yy}, such that they can be related to a Poisson ratio ν

$$\left(\frac{\Delta d_z}{d_z}\right)_{\mathrm{L}} = \frac{{}_{\mathrm{L}}d_z - {}_{\mathrm{L}}d_{z0}}{{}_{\mathrm{L}}d_{z0}} = \varepsilon_{zz} = \frac{-\nu}{1-\nu}\left(\varepsilon_{xx}+\varepsilon_{yy}\right) = \frac{-\nu}{1-\nu}\left(\frac{{}_{\mathrm{L}}d_x - {}_{\mathrm{L}}d_{x0}}{{}_{\mathrm{L}}d_{x0}} + \frac{{}_{\mathrm{L}}d_y - {}_{\mathrm{L}}d_{y0}}{{}_{\mathrm{L}}d_{y0}}\right), \tag{41.13}$$

where ${}_{\mathrm{L}}d_{x0}$, ${}_{\mathrm{L}}d_{y0}$ and ${}_{\mathrm{L}}d_{z0}$ are the lattice plane spacing for the layer parallel and normal to the surface in the unstrained state; these are parameters needed to identify the alloy composition. The Poisson ratio used in this way is dependent on the orientation, however this can be determined from the elastic coefficients or from tabulated values. Many typical semiconductors have values $\nu \approx 0.3$, whereas maintaining constant volume $\nu \approx 0.5$. In our example in the first instance we are assuming $\varepsilon_{\parallel} = \varepsilon_{xx} = \varepsilon_{yy}$, i. e., that the strain is isotropic in the interface plane. Combining (41.13) with the c/a ratio ($= K$) of the expected unstrained state for the layer, we can isolate c_{L} ($= H_z {}_{\mathrm{L}}d_{z0}$) in terms of the measurable (${}_{\mathrm{L}}d_z$ or Δd_z) and known or tabulated values (H_z, K, ν, c_{S}, a_{S}) as

$$c_{\mathrm{L}} = \left(\frac{2\nu}{1-\nu}\right)\left(Ka_{\mathrm{S}} - {}_{\mathrm{L}}d_z H_z\right) + {}_{\mathrm{L}}d_z H_z\,, \tag{41.14}$$

where ${}_{\mathrm{L}}d_z H_z = c_{\mathrm{S}} + \Delta d_z H_z$, c_{S} is the lattice parameter of the underlayer or substrate normal to the surface, and a_{S} is the lattice parameter in the plane of the interface. Within the approximations in this example, the values of K and ν can be estimated, which can be iterated since they are composition dependent. It is assumed that a_{S} is a known or database value, then c_{L} can be derived and the phase composition can be obtained by the interpolation [41.5]

$$x = \frac{c_{\mathrm{L}} - c_{\mathrm{GaN}}}{c_{\mathrm{InN}} - c_{\mathrm{GaN}}}\,. \tag{41.15}$$

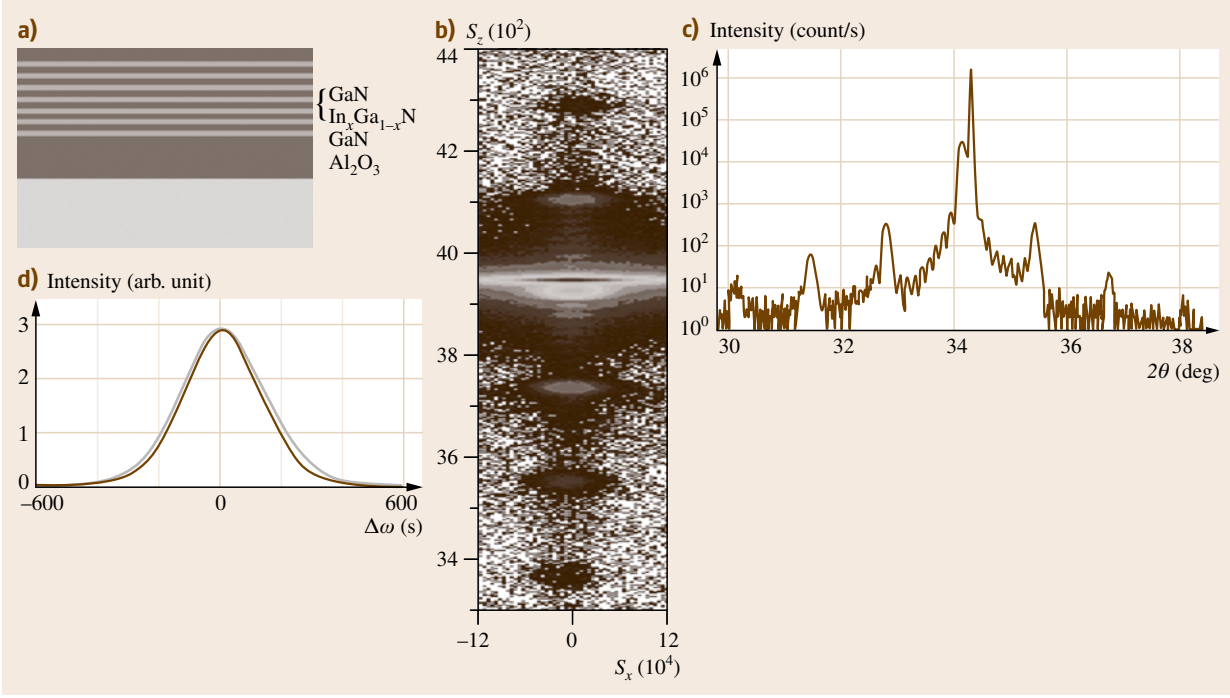


Fig. 41.3a–d The structure of an InGaN/GaN multilayer on a thick GaN buffer layer is given in **(a)**, along with the reciprocal space map **(b)** and some extracted profiles along s_z **(c)** and along s_x **(d)**. The intensities in **(d)** correspond to the GaN buffer (*grey*) and the superlattice (*brown*), indicating that they have similar microstructure, although the superlattice may be of better quality; the intensities are normalized for comparison

In the example discussed, $x = 2.2\%$. The purpose of the above derivations is to indicate some of the approximations required for a direct interpretation of the intensity distribution in reciprocal space. Before discussing a more rigorous approach, the influence of layer strain relaxation should be considered, since similar approaches are applicable.

41.2.3 Interpretation of Data Collected from Several Reflections – The Reciprocal Space Map – An Example

If the atom positions do not align from layer to substrate across the interface, then ${}_{\mathrm{L}}d_x \neq {}_{\mathrm{S}}d_x$ and $a_{\mathrm{L}} \neq a_{\mathrm{S}}$, and hence a_{L} should be substituted for a_{S} in (41.14). This dimension can be obtained from a reciprocal space map of a reflection having a component of this in-plane spacing (Fig. 41.4, compare with Fig. 41.2b). The distance in s_x between the layer peak of Fig. 41.4 and that of Fig. 41.2 gives the lateral plane spacings in the layer ${}_{\mathrm{L}}d_x$ by applying (41.9) and similarly for the substrate reflection ${}_{\mathrm{S}}d_x$, so that $a_{\mathrm{L}} = {}_{\mathrm{L}}d_x \Delta H_x$. For example, when combining the *hkil* reflections 0002 and $2\bar{1}\bar{1}4$, $\Delta H_x = \sqrt{\{4[h^2 + k^2 + hk]/3\}} = \sqrt{\{4[2^2 + 1^2 - 2]/3\}}$ for the hexagonal system with a (0001) surface plane. For the cubic system with a (001) surface, $\Delta H_x = \sqrt{\{[h^2 + k^2]\}}$. Alternatively if the substrate is unstrained and has known lattice parameters $a_{\mathrm{L}} = a_{\mathrm{S}} + \Delta d_x H_x$, where $\Delta d_x \equiv 1/\Delta s_x$, (41.9) the layer lattice parameter can be measured directly from the reciprocal space map of Fig. 41.4 (after accounting for any tilt). The tilt observed in the map of Fig. 41.2b changes $\omega_1 \rightarrow \omega_1 + \delta\omega$ for the layer, and ω_2 for the substrate is unchanged. This approach has been extended to the more general case, i. e., any space group, anisotropic relaxation or orientation [41.6].

From the discussion above a series of reciprocal space maps can yield a considerable wealth of information about the sample: the unit cell dimensions, the distortion, and some characteristic dimensions (of mosaic blocks and thickness, for example) and their relationship to other layers or the substrate. The combination of several reflections or the shape of the peaks in reciprocal space can be used to isolate the contribution of the tilts, sizes, and strains [41.4, 6], using data as illustrated in Fig. 41.3b,d for the 0002 and 0004 reflections of GaN, for example. These direct approaches are very suitable for feedback during the early stages of crystal growth or for monitoring established growth procedures. However when there are several layers, and the material exhibits good crystal quality, the assumptions concerning the peak positions and the peak shapes become very approximate. This is because the simple interpretation assumes that each layer scatters independently, i. e., the scattering theory is represented very simply and it is assumed that the instrument and method of data collection create no artefacts.

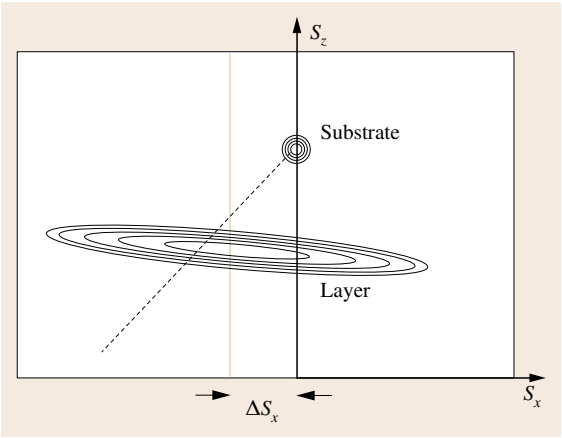


Fig. 41.4 A schematic of a reciprocal space map for a reflection from planes inclined to the surface plane. The lateral misalignment of the layer and substrate peak is indicative of differing lateral lattice parameters (layer strain relaxation) and/or relative lattice plane tilts

41.3 Instrumental and Theoretical Considerations

The basics of the instrument for data collection and the appropriate scattering theories to use are discussed in this section. Many of these aspects will become clear in the examples.

41.3.1 The Instrument for Collecting X-Ray Diffraction Patterns

The reciprocal space maps of Figs. 41.2, 41.3, 41.4, etc., are a very small subset of Fig. 41.1. Referring to the expanded detail in the inset of Fig. 41.1, the representation of the angular divergence of the incident beam $\Delta\omega$, the angular acceptance of the scattered beam $\Delta 2\theta$, and the spread in wavelength of the source $\Delta\lambda$ all add to smearing of the x-ray probe [41.6]. The divergence $\Delta\omega$, the acceptance $\Delta 2\theta$, and the wavelength dispersion $\Delta\lambda$ vary throughout reciprocal space. If this capture volume is reduced too far, the intensity will be reduced significantly; however with recent advances in x-ray mirrors and lenses, and more significantly the availability of synchrotron sources, the expanding scope of x-ray scattering has not been diminished by insufficient intensity. The equipment is not significantly different from laboratory sources to synchrotron sources, but in terms of intensity differences the more exotic experiments only become feasible for synchrotron sources, especially if combinations of wavelengths are needed. However the long-term stability and convenience of laboratory sources make them the best choice for detailed analyses for most problems, and certainly for screening experiments prior to using a synchrotron source.

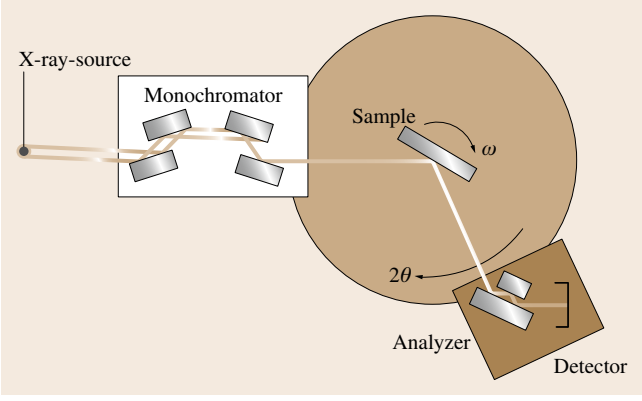


Fig. 41.5 The schematic of the high-resolution multiple-crystal diffractometer, which gives a small and well-defined reciprocal space probe over much of reciprocal space

A good basic *workhorse* for many crystal growth studies is the instrument shown in Fig. 41.5 [41.7]. The monochromator and analyzer combination gives a very small reciprocal space probe over most of the accessible region available (defined by the wavelength used: the smaller the wavelength, the greater the accessible sphere). The intensity for laboratory sources is kept high by maintaining a reasonably large beam, $\approx 1 \times 10\,\text{mm}$. The angular spread in the beam normal to Fig. 41.1 is the axial divergence, which is only restricted to $\approx 0.5°$, compared with $\approx 0.001°$ for $\Delta\omega$ and $\Delta 2\theta$, thus the measured reciprocal space map is a projection. Restricting the axial divergence creates a three-dimensional (3-D) reciprocal space map [41.8], yielding further information, but this approach will not be discussed further in this chapter. The resolution is a complex interplay between the instrument, the sample, the choice of reflection, the wavelength, and the alignment of the instrument and sample. Generally though, the sample is the determining factor and the probe is small compared with the structural features that influence the pattern, although the influence of the instrument should also be modeled.

These are only general comments on the resolution; methods of maintaining resolution and enhancing intensity will be addressed in Sect. 41.5.

41.3.2 Interpreting the Scattering by Simulation

The kinematical theory permits the basic interpretation summarized in Sect. 41.2 but is only valid for structures that are *ideally imperfect*. However, the majority of structures grown by highly controlled growth methods, e.g., molecular-beam epitaxy and metalorganic vapor-phase deposition, will result in crystal structures that are highly perfect. Also the substrate material is often Si, sapphire, MgO, GaAs, InP or $SrTiO_3$, which are nearly perfect. This level of perfection necessitates the application of dynamical theory [41.9], which includes extinction effects, the loss of incident intensity through scattering, interference with the multiply scattered wave, etc. [41.10, 11]. This can have a much more dramatic effect on the penetration depth than normal photoelectric absorption [41.12]. Also the peak positions, even for very simple structures, vary with layer thickness, creating uncertainties in the composition, thickness, and simple interpretation of periodic structures [41.13]. Perhaps the most important reason

for using dynamical theory is that, due to the increasing complexity of today's layer structures, the pattern is not a simple superposition of all the contributing layers, but a complex interference with peak shifts, beat frequencies, etc. Simulation also removes a few myths about direct interpretation, especially concerning superlattices, e.g., interfacial layers (roughening) that might appear as a loss of perfection can under some circumstances increase the satellite intensities [41.14]. Simulation has significant advantages in that the whole structure, thickness, and composition of each layer can now be analyzed simultaneously, and also all the iterative assumptions concerning the changing c/a ratio and Poisson ratio with composition, etc., are all included. The simulation of the profiles, or the reciprocal space maps, rely on the scattering amplitude (41.1), the unit cell parameters, and the elastic parameters, many of which are established database values, and of course some estimate of the structure to be simulated. The sensitivity to structural details is very high, so the need to have a good procedure for fitting the calculated to measured profiles by adjusting the model is evident. These procedures are becoming very sophisticated.

The simulation of more imperfect structures, for example, quantum dots within a perfect crystal matrix or rough interfaces in a period structure, is best achieved with the distorted-wave Born approximation (DWBA) [41.15]. This considers multiple scattering and is therefore dynamical and takes into account effects associated with the transmission of the beams. It is based on solving the scattering from a perfect structure by the incident wave. This average scattered wave defines the local electric field of the wave, which is used to calculate the disturbed wave that results from quantum dots, precipitates, interfaces, and other disturbances. The combination of the undisturbed and disturbed waves gives the resultant scattering pattern. If the disturbance becomes too large then this perturbation method is not applicable and kinematical theory becomes more appropriate and can then be considered as a single scattering theory, i. e., the Born approximation (BA) [41.16].

The analytical approaches available and discussed in the following sections therefore have theoretical models that cover: perfect to nearly perfect structures (dynamical theory), imperfections giving rise to diffuse scattering (distorted-wave Born approximation), and highly imperfect or very small perfect regions (kinematical theory). If however an experiment is carried out when $S \approx 0$, i. e., with very small incident and scattering angles (specular scattering), then we see from (41.1) that the scattering amplitude is purely the sum of the scattering strength of individual atoms, independent of their position $\boldsymbol{r}$ in the unit cell. It is then possible to consider the layers in a structure as consisting of a constant electron density, and therefore the interaction is analogous to light scattering from regions of constant refractive index; hence a much simpler, *optical* theory is valid [41.17]. Consequently specular reflectometry (close to the total external reflection condition) within a few degrees of the 000 reflection, is usually simulated with this optical theory.

41.4 Examples of Analysis from Low to High Complexity

The advances in x-ray scattering have been very rapid in recent years so this section will give a snapshot of the state of the art as well as the more established methods. Eventually the useful state-of-the-art analyses that are proven to be successful in revealing structural parameters, within a broad range of applications, will become established methods that are more accessible to more scientists. However, in the interim, these state-of-the-art methods are generally only accessible through collaboration with scientists in these fields, or through significant self-endeavor.

41.4.1 Established Methods

The interpretation of scattering profiles through simulation is now widely available for perfect and nearly perfect structures, based on dynamical theory. A typical analysis is illustrated in Fig. 41.6a, for a GaAs-based structure. The approximate thickness and composition of each layer was known from the growth and this gave the starting model for its refinement through comparison of measured and simulated profiles. This is a two-layer sample so from the start we have to determine the composition and thickness for the InGaAs and the AlGaAs layers. The peak positions give some rough starting values, however with the limited number of parameters the fitting process can cope easily with the estimated values, which were within $\approx 20\%$ of the final fit values. The fit to the data was good with just this information included, i. e., two layers; however, the intensity of the fringes between -5000 and -4000 arcsec was significantly overestimated. The inclusion of some

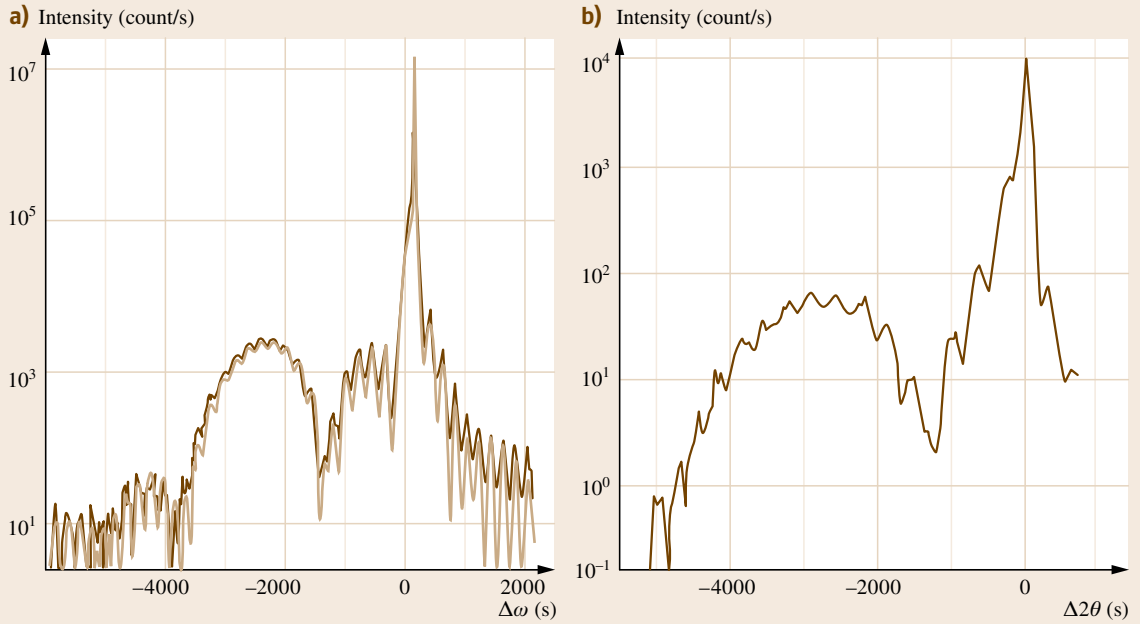


Fig. 41.6a,b The diffraction profile from an AlGaAs/ InGaAs/GaAs structure obtained with the beam-selection diffractometer (*brown*) and simulated best fit profile (*light brown*) **(a)**, and the static diffractometer, discussed in Sect. 41.5 **(b)**. The data collected in **(a)** and **(b)** were collected in 7.3 min and 1 s, respectively

grading at the InGaAs to AlGaAs interface had a strong influence on this broad InGaAs layer fringe (with the AlGaAs thickness fringes superimposed). This grading reduced the intensity to achieve an improved fit (Fig. 41.6a). This additional graded layer is only observable with very high intensities or longer data collection times. This data was obtained with the beam-selection diffractometer, which is composed of an x-ray mirror, scattering from the 004 reflection of the sample to a 004 Ge analyzer in front of the detector [41.18]. This gives a substrate peak intensity of ≈ 30 Mcps, so the ω–2θ scan data was collected in just over 7 min with a 0.005° step size. The final parameters were 0.0130 μm of $In_{0.145}Ga_{0.855}As$, a 0.0059 μm graded layer of InAlGaAs, and 0.0773 μm of $Al_{0.258}Ga_{0.742}As$. These values compare with the best fit of 0.0157 μm of $In_{0.145}Ga_{0.855}As$ and 0.0808 μm of $Al_{0.258}Ga_{0.742}As$ when no interfacial grading was included, which would be the case if the broad InGaAs layer fringes were not observed.

For very complex samples the fitting process requires more persistence and time, mainly because of the high degree of correlation between parameters and the complicated interference in the scattering from individual layers. Typical structures that fall into this category are vertical-cavity surface-emitting lasers (VCSEL). This example is composed of quaternary AlGaInP layers flanking a ternary GaInP quantum well, which are surrounded by Bragg stacks of AlGaAs/AlGaAs periodic layers; the laser wavelength of the radiation generated in the quantum well is defined by its thickness and composition, and the efficiency of the cavity relies on the Bragg-stack composition modulation and thickness. There will be some interfacial grading and further layers making a total in excess of 450. This is quite a challenge, but possible to solve [41.19]. To add to the difficulty the surface is offset from the (001) crystal plane by 10°. Careful control of the fitting process and appropriate choice of reflections gives a very good indication of the sensitivity of certain parameters and some clear bounds that fit the x-ray data. The final fit compared with the experimental profile for the 006 reflection is shown in Fig. 41.7.

The sensitivity to composition and thickness is a clear advantage of x-ray scattering, and this is enhanced at increasing scattering angles (41.8). When the scattering angle is very small, the composition sensitivity from strain is negligible, but the scattering is still

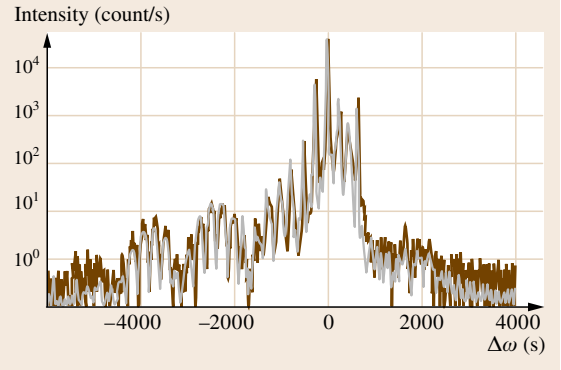


Fig. 41.7 The complex scattering from a VCSEL structure (*brown*) is shown along with the final fit to this profile (*grey*) (data courtesy of P. Kidd). The result gave bounds for the layer composition and thickness throughout this structure consisting of > 450 layers

sensitive to the composition because of the differences in scattering strength (41.1). This can be very useful for isolating the influence of compensating strain in structures such as $Si_{1-x-y}Ge_xC_y$ [41.20]. C has a very small covalent radius, but scatters very weakly, whereas Ge has a covalent radius closer to Si and scatters strongly. Since in this case Ge scatters strongly, the composition can be obtained from the reflectometry profile to within $\approx 0.1\%$ in x, provided the C concentration is estimated to within $\approx 1\%$ in y. Since the typical C concentration is $< 1\%$, the Ge concentration can be determined to good precision; the reflectometry curve is given in Fig. 41.8a. The C concentration is then extracted from the 004 profile (Fig. 41.8b); however, this relies heavily on the assumptions concerning the C covalent radius. For the result to be consistent with secondary-ion mass spectrometry (SIMS), the covalent radius probably matches that in cubic SiC or the relationship between the lattice parameter and composition is far from linear.

From the above analyses it is clear that composition, thickness, and grading due to interdiffusion can be obtained quite directly. The interface roughness in the plane can also be estimated, although the theory does become a little more problematic. As with all the examples above, the structural model is derived via an iterative procedure by comparing simulated with measured profiles. Thickness, composition, and even interfacial grading are reasonably predictable, but lateral roughness is less clearly described. There are several models of interfaces, ranging from a truncated fractal [41.21], staircase for vicinal surfaces [41.22], and castellated surfaces, Holý within [41.6]. The resulting profiles are quite different and require estimates of the extent of the roughness replication layer to layer, i. e., none, partial, and full replication. The truncated fractal

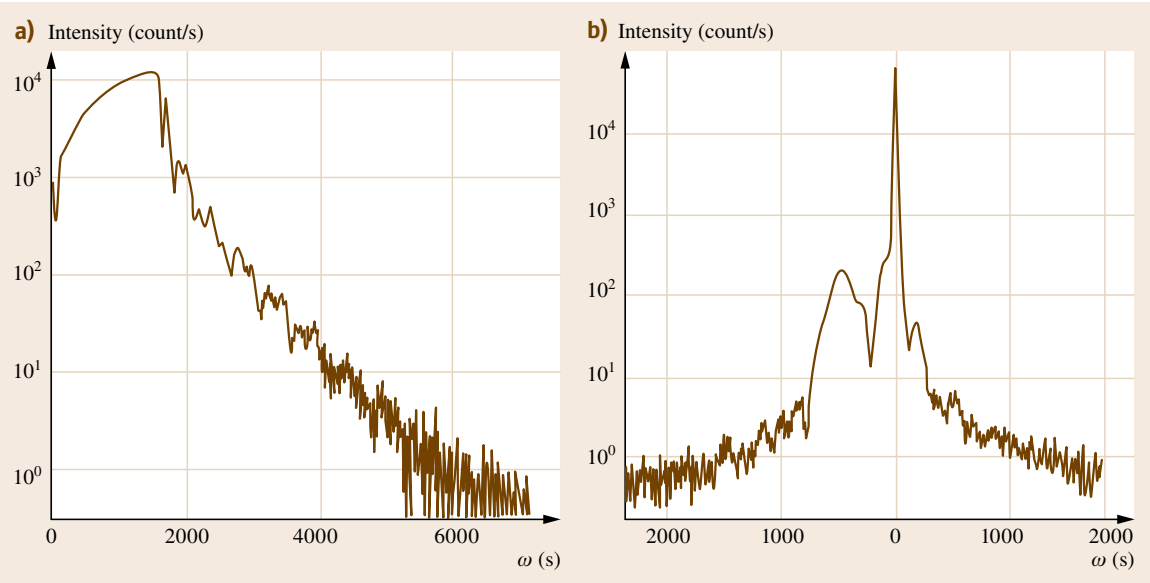


Fig. 41.8a,b The two profiles, close to the 000 **(a)** and the 004 **(b)** reflections, for a SiGeC structure used to break the correlation between parameters. By fitting the two profiles the composition of the Ge and C could be determined, because of their different influences on the two profiles

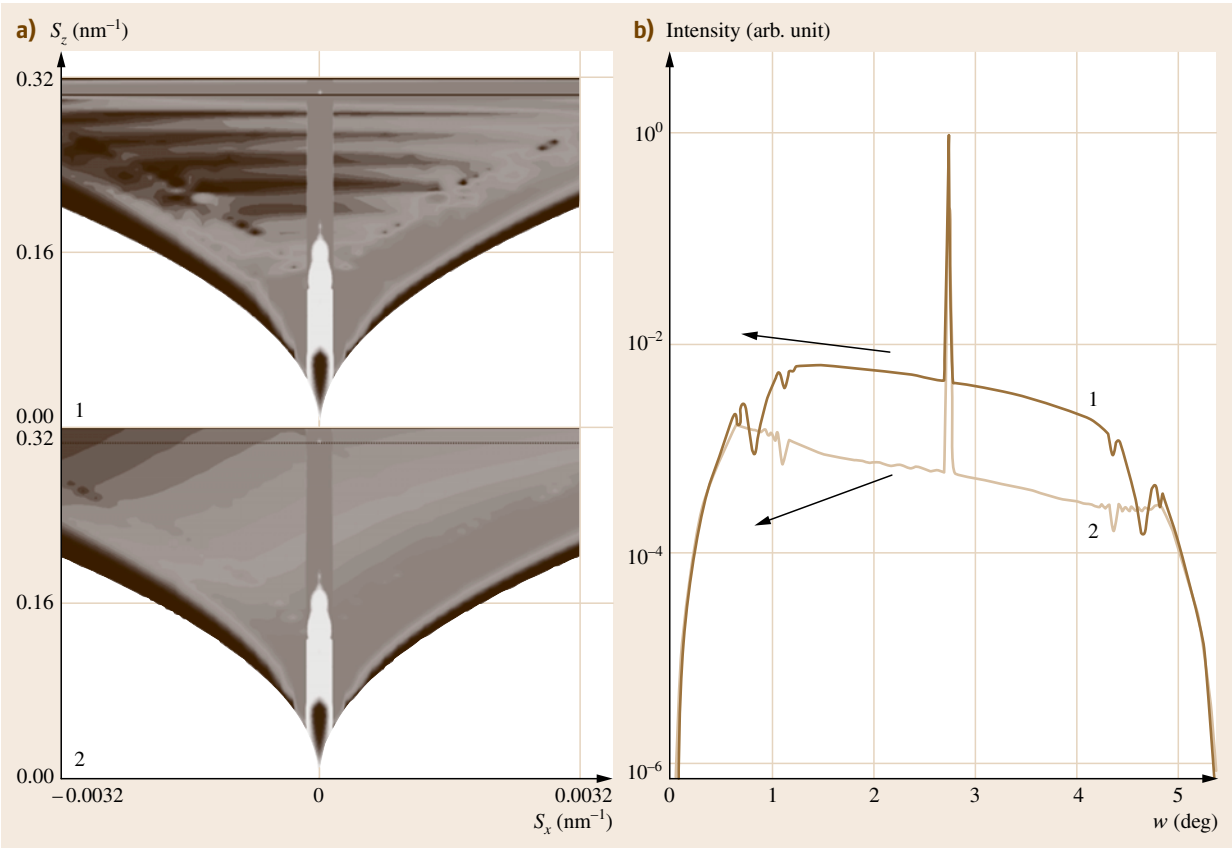


Fig. 41.9a,b The change in diffuse scattering close to the specular profile for various interface types: (1) castellated fully replicated and (2) unreplicated layer to layer for a periodic superlattice. The structure simulated is composed of (AlAs 50 nm + GaAs 50 nm) × 10 on GaAs, all with raised regions of width 50 ± 50 nm and height 1 ± 0.2 nm

description has a maximum length that is often termed the lateral correlation length; however, this is not an easy parameter to visualize. The nature of the jaggedness is also described by a single parameter and the scattering is derived from the correlation between different heights across the interface. The staircase and castellated surfaces require a distribution of heights and widths, with the former having an inclination angle. The castellated surface will almost equate to a fractal surface at one extreme. It should be clear from this that a good physical model for the likely interface shape is required to obtain some significance in the results. If the roughness is not replicated from layer to layer then the distributed intensity around the specular profile is rather uniform and weak, whereas replicated roughness brings the diffuse scattering into bands (Fig. 41.9). The most appropriate theory for modeling the scattering is DWBA, but this is only applicable for small roughness and a limited s_z range. If the roughness becomes significant or the simulation is taken to large s_z, then the kinematic approach (BA) is more applicable. Making a decision on the cross-over of DWBA to BA depends on several aspects [41.23], but certainly interfaces that are comparable in width to the layer thickness itself should use the BA, although this will not reproduce the region close to the critical angle or Yoneda wings [41.24] associated with surface roughness and transmission factors.

41.4.2 New Methods and New Analyses

All the methods and analyses described above are generally available. However, because the science in this field is progressing rapidly, some of the most exciting

work requires considerable effort to reproduce. In many instances though, the equipment to undertake the work is either available commercially or for highly specialized experiments a synchrotron may be the only route. This section will consider some methods that extend the analysis beyond those discussed above; it does not pretend to be comprehensive, but rather picks out a few advances.

Dislocations are one of the significant defects that can destroy devices, and they influence the scattering, as will be clear from Chap. 42 on x-ray topography. Their influence is clear in reciprocal space maps, even at the very early stages of layer relaxation [41.25]. These misfit dislocations have a clear influence since they cause lattice plane rotation and strain that can be modeled [41.26]. The influence of threading dislocations is less clear since the Burgers vectors are not necessarily additive, but do contribute to a twist that is often interpreted purely in terms of peak broadening, although the influence of strain must also be considered [41.27]. Generally misfit dislocations, in GaN for example, are accompanied with threading dislocations, and obtaining the proportions of each requires analysis of the diffraction peak tails as well as the peak widths. A crucial aspect in this analysis is the correlation between dislocations, as this can have a profound influence on the FWHM [41.27, 28]. *Holý* et al. [41.29] are now taking this to the next stage, by reducing the probe volume to 1 μm or less, such that it is closer to the dislocation separation. It may then be possible to determine the threading dislocation densities in GaN epitaxial-lateral-overgrown structures, by averaging the diffraction profiles over several positions. However the simulation of these structures is not yet fully resolved.

The development in the analysis of quantum dots gives a good indication of what can be achieved by x-ray scattering experiments. The interpretation of scattering of buried dots can be based on DWBA. This is most appropriate in semiconductor systems where the dots form a region with a different lattice parameter and scattering strength. The distortion will then extend to the surface where the stress is relieved. The distribution of the dots can be measured by the lateral satellites associated with the Bragg peaks. For ordered arrays of dots these produce very strong effects [41.30]. However, for the more typical self-assembled dots, there may be only one broad satellite. This lateral spacing (41.9) between the satellite and the layer peak will give some measure of the dot spacing and its variation [41.31]. Combining spacings from planes inclined to the surface at several azimuths will give the average arrangement in the plane. For dots on the surface, the best approach is to use grazing-incidence small-angle scattering [41.32] to give satellites either side of the specular scattered beam. To extract details of the shape of buried dots, in-plane scattering appears to be the most unambiguous route [41.33], by modeling the strain, which must be anisotropic [41.34]. Generally though, these approaches have been entirely concentrated at synchrotron sources, although the analysis of buried dots [41.34] was carried out with a standard laboratory instrument with a sealed x-ray source (Fig. 41.5) and has even been extended to the analysis of a single layer of quantum dots buried 100 nm below the surface [41.35].

As discussed in Sect. 41.4.1, the shape (thickness) and not the strain is obtained from scattering close to the 000 *reflection* condition, and therefore the grazing-incidence small-angle scattering configuration will only give the shape of the dots. However to obtain the strain state of surface dots one uses in-plane scattering, i. e., grazing-incidence and grazing-exit scattering close to Bragg peaks, as in the analysis of buried dots. If the dots are known to have a regular shape or the shape can be deduced, some estimate of the strain state (composition or strain relaxation) can be obtained [41.36], although this can lead to complications with uncertainties about the shape [41.37]. Extracting the dot shape is clearly an iterative method based on some initial guess, and this averaging of the diffraction process will result in a weighted average of the information. If the x-ray source is coherent and an individual dot can be isolated, then the resulting diffraction pattern can be directly inverted to reconstruct the dot shape [41.38]. In practice isolating one dot is difficult, as is having perfect coherence, however incorporating a few dots with a partially coherent beam gives adequate data to reconstruct the dot shapes [41.39]. The procedure uses the measured amplitudes and iteratively solves for the phase information using methods developed in optics. The scattering volume is so small that the kinematical (BA) scattering theory is sufficient. The typical data collected from an experiment with a sealed laboratory source, averaging over $\approx 10^7$ dots [41.35], compared with that collected from a coherent synchrotron source for a line of $\approx 10^1$ dots [41.39], is presented in Fig. 41.10. The former obtains the dot shape and composition indirectly from the strain distribution created by the dots, whereas the coherent diffraction experiment gives the scattered amplitudes, i. e., the Fourier coefficients for the shape of surface dots.

The arrangements of quantum dots can be either regular or self-assembled, and both can be studied

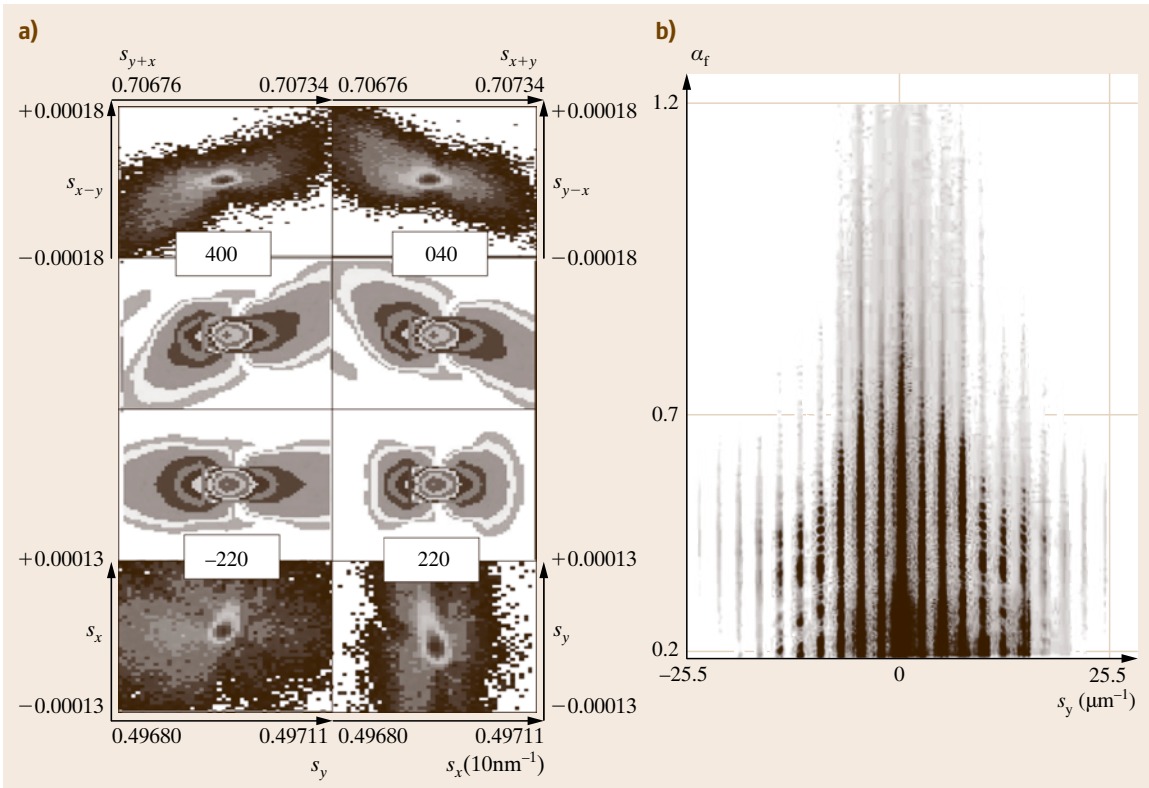


Fig. 41.10a,b Different experiments revealing different levels of information about dots: **(a)** buried dots using a sealed laboratory source, in-plane reciprocal space maps obtained just above the critical angle, for various reflections with their best-fit simulations, **(b)** the forward scattering, close to 000, from surface dots obtained with a coherent synchrotron source where the incident beam is below the critical angle for the substrate and the resultant scattering is purely from the dots (data courtesy of I. Vartanyants)

with x-ray scattering to yield useful information. Quantum wires on the other hand are nearly always regular since their growth is governed by vicinal wafers or patterning. These structures give impressive scattering patterns (Fig. 41.11) that are easily observable with laboratory sources [41.40]. They also change depending on whether the quantum wires or Bragg gratings are parallel or perpendicular to the incident x-ray beam. A rather pragmatic approach to the analysis can be applied. Effectively the Fourier transform based on the BA (kinematical) theory of a trial model shape and period can be used to obtain some very useful shape parameters. This will not give reliable intensities or account for

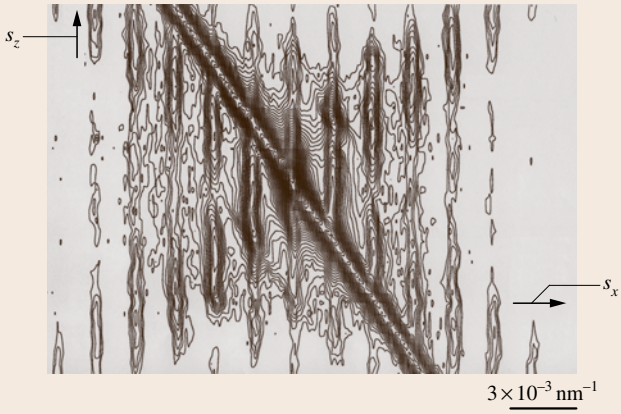


Fig. 41.11 The scattering close to the 004 reflection for a laterally periodic grating on a GaAs substrate. The data was obtained with a high-resolution diffractometer (data courtesy of M. Gailhanou). The *long streak* comes from the analyzer that had insufficient channel bounces before reaching the detector ◀

the true shapes of the scattering peaks for detailed interpretation. All combinations of the beam paths need to be considered [those scattered directly from the grating, those transmitted (refracted) through the grating, etc.] and are coherently brought together to account for all the cross-hatching of intensity, etc. based on the DWBA [41.41]. Surface features, dots or wires, do have an added complication to the analysis in that the strain influence is unlikely to be homogeneous and requires finite-element analysis to predict the strain distribution. However obtaining a good strain model can result in a high level of detail. Surface quantum wires studied with a detailed strain model suggest that SiGe/Si superlattice wires have an oxide coating [41.42]. For buried gratings the strains dominate the scattering [41.43], as with buried quantum dots.

41.5 Rapid Analysis

The methods discussed so far in this chapter have concentrated on analysis, both direct and simulation, and some specialized applications, with some discussion on the instrumentation. However it is largely the development in instrumentation that has made x-ray scattering so much more accessible. The instrumentation at synchrotron x-ray sources is very similar to how x-ray laboratories used to be, i. e., components are built specifically for the type of analysis required, so the set up time can be a significant factor. This does have the advantage in that some of the latest technology can be incorporated very rapidly, but the instruments do require highly trained specialists to set up the experiments, and it all becomes a significant team effort. Commercial laboratory machines are, on the other hand, rather versatile, with x-ray mirrors, lenses, and a plethora of monochromators, analyzers, and detectors. State-of-the-art laboratory equipment has exchangeable components that are precision-aligned so that several experimental configurations can be used without removing the sample. So in many ways the components, through significant cross-fertilization of ideas, are rather similar, although the intensity and wavelength choice at synchrotrons outstrips the laboratory source. However, with the use of mirrors, lenses, etc., the laboratory source gives sufficient intensity for the majority of applications. Some of the most interesting advances have been made in detector technology, which until recently were saturated by the high count rates. New developments in solid-state detectors, which are pixel-array single-photon counting detectors, will be a fascinating area to watch to create new opportunities.

Since the push is for increased speed and feedback in crystal growth there have been some interesting developments. In the quest for higher speed the intensity needs to be increased, or the data can be collected in parallel. Higher intensity usually comes at a cost to the resolution, although this can be overcome by making the experiment much more specific to the analysis. The first example of this was given in Sect. 41.4.1, where the more specific beam-selection diffractometer gives a 30× increase in intensity compared with conventional instrumentation with an x-ray mirror (Fig. 41.5). This is still a scanning method, which cannot compete with parallel data collection. The discussion (Sect. 41.3.1) on the influence of the instrument considered only resolution in diffraction space. However, when this is combined with the *real* space resolution, it is possible to achieve high-resolution profiles instantaneously [41.44]. Consider the profile in Fig. 41.6a, which already has 30× the intensity of a conventional instrument (Fig. 41.5); it took ≈ 7 min to collect. However the profile given in Fig. 41.6b was collected in 1 s. This static diffractometer has rather simple components: an x-ray source, slit, and a linear position-sensitive detector. The simulation of the scattering profile requires modeling of the beam paths to account accurately for all the features. This method can be applied to all wafer-based samples, by finding a suitable reflection, or if that is not possible by choosing an alternative wavelength. The highest resolution relies on a high-quality sample; however, useful information can still be extracted from imperfect materials, as will be shown in Sect. 41.6.

There are several growth chambers at synchrotrons for in situ characterization. Some of the early work concentrated on analyzing the surface reconstruction [41.45], and the investigation of dislocations by topography [41.46]. Some of the methods discussed in Sect. 41.4.2 have been applied in situ at synchrotrons to study the evolving crystal growth, and it is possible to see features even when 0.5 monolayer has been deposited [41.47]. A laboratory source, with the advantage of bringing the x-ray experiment to the metalorganic chemical vapor deposition (MOCVD) growth

chamber, has been used for studying the growth of InGaN on GaN [41.48]. The source is used in combination with a Johansson monochromator to focus the beam onto the rotating sample, and despite the complications associated with rotation and wobbling, the evolution of the InGaAs on a GaN buffer layer is very clear. The integration time is $\approx 17\%$ of the rotation period. This is not high-resolution data, but perfectly sufficient to monitor the growing layer. This instrument is again relatively simple with a source and focusing Johansson crystal (which captures the divergent beam from the source and focuses it onto the sample); the scattered wave is collected by a position-sensitive detector.

These rapid analyses could be very useful for either quick feedback during the initial stages of crystal growth or controlling an established process. However, there are some interesting possibilities associated with rapid data collection and high intensity that will be addressed in the next section. Although these examples are specific for the material of interest, they are easily adapted; the concentration at this stage is on composition and thickness measurement and some estimate of *quality*.

41.6 Wafer Micromapping

Lateral changes in the composition, thickness, and *quality* across a wafer can be very revealing to the crystal grower. Generally the x-ray probing beam is of the order of mm, and therefore can suffer from significant averaging, adding uncertainty to the interpretation. As an example, the influence of peak widths and peak shifts contribute to the measured peak widths. The benefit of rapid data collection is that the probing beam can be reduced and still achieve reasonable measurement times. The typical sizes of the beam of the beam-selection diffractometer and the static diffractometer are $100\,\mu\text{m} \times 10\,\text{mm}$ to achieve these high data collection rates, so reducing this to $100\,\mu\text{m}$ double-pinholes reduces the intensity considerably ($> 100\times$). The intensity achieved is still very usable, although requiring longer count times or a compromise on counting statistics. As an example consider this far-from-perfect GaN sample, which is first probed with the beam-selection diffractometer using the $100\,\mu\text{m}$ double-pinholes, to obtain an estimate of quality; this is achieved by centering on the 0002 GaN peak and scanning in ω (Fig. 41.12a). The width of this peak is related to small tilted regions or some finite lateral dimension, e.g., the distance between defects or mosaic block size. The resulting map over the $2.5 \times 5\,\text{mm}$ sample at $50\,\mu\text{m}$ intervals is given in Figs. 41.12b,c, where the peak positions and peak widths across the wafer are shown. Clearly the peak width varies on this $100\,\mu\text{m}$ scale (Fig. 41.12b) and is most pronounced in one region that is between two areas that have quite different peak positions (Fig. 41.12c). These may be large mosaic blocks or regions of large curvature. If the measurement of the peak width was obtained on the mm scale then all these contributions would be combined and may have led to a different interpretation.

A similar analysis has been performed with the static diffractometer, but for speed obtained with a $200\,\mu\text{m}$ probe and stepping at $100\,\mu\text{m}$ intervals. This analysis gave the variation in composition and period across the wafer. The satellites are very weak (Fig. 41.12d) from this structure (the In concentration was rather low), but still an indication of the period variations could be obtained and did not indicate any significant fluctuations. The composition, though, does vary significantly over the wafer, with the center-right region being 10% below that of the lower region of the wafer.

The variation in the composition and the period can probably be linked quite clearly to the growth of the structure. However, the material *quality* can help in screening wafers. Reducing the probe clearly begins to separate the peak width and the peak position, and since the width, rather than the position, is generally related to distorted regions this may correlate with the local defect density and hence optical emission in GaN lasers and light-emitting diodes (LEDs), for example. The reduction in the probe size has two effects, not only the isolation of small regions but also the variation in peak broadening effects is likely to become larger as the averaging effects are reduced [41.49]. In the extreme case when the probe is close to a micron, the effects of individual dislocations will be observed, as discussed in Sect. 41.4.2. *Baumbach* et al. [41.50] have extended micromapping towards this extreme, using an extension of their *rocking-curve imaging* technique [41.51], which requires a well-collimated monochromatic beam from a synchrotron source; the spatial resolution is

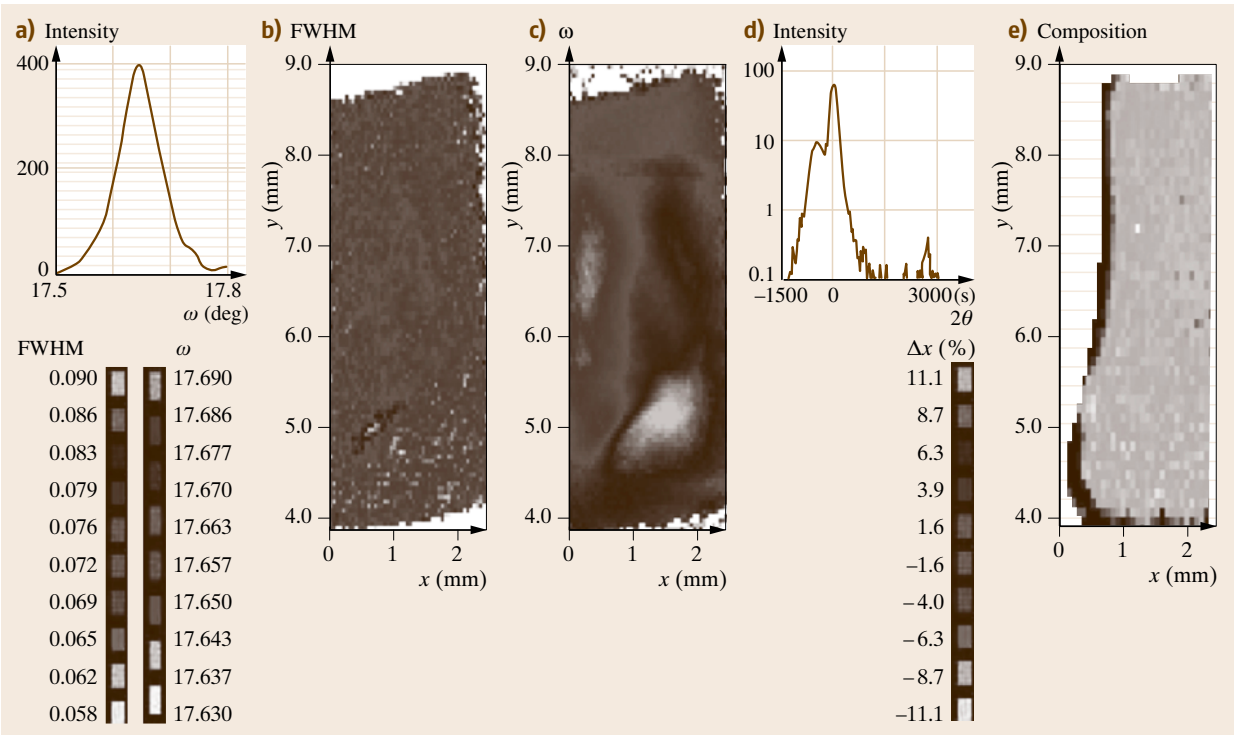


Fig. 41.12a–e Examples of wafer mapping on the 100 μm scale on a GaN wafer with a InGaN/GaN superlattice. Data collected with the beam-selection diffractometer: **(a)** a typical profile, **(b)** the distribution of the FWHM $\Delta\omega$, **(c)** the distribution of the peak position ω. The data collected with the static diffractometer; **(d)** a typical profile; and **(e)** the variation in composition across the wafer

defined by the pixel size of an area detector. By rocking the sample an array of profiles is achieved. An interesting aspect with this is the cross-over between topography and conventional high-resolution scattering, where the latter will not only break the averaging effect of larger probes but could provide ways of interpreting the contrast in topography that can arise from several complicating aspects. This combination of microdiffraction or topography (real space) and reciprocal space has many advantages in unraveling the structural information that is not revealed on the mm scale.

41.7 The Future

There is little doubt that x-ray scattering is very sensitive to structural details, and these methods are pushing the technological, theoretical, and instrumental developments. The intensity and quality of x-ray sources is no longer a restriction, with many synchrotrons existing around the world. The possibilities from coherent x-ray sources will become clearer as these facilities become more widely available, e.g., not only at synchrotrons but at x-ray free-electron laser facilities. The improvements in detectors will also create many opportunities, and are very necessary to cater for the large intensities now available.

Instrument development is benefiting from the cross-fertilization of laboratory and synchrotron-based work, with sophisticated optics being of use for both types of x-ray sources. Theoretical developments will also be a challenge as more scientists struggle to interpret the scattering from ever more exotic materials and structures. Enhanced computing power will benefit this field, for example, when modeling the diffuse scat-

tering or optimization for extracting parameters of very complex materials.

It is very difficult to pinpoint specific areas where the most significant method developments will occur, since there is a large range of possibilities: coherent scattering, microtomography, micromapping, etc., but the most significant will be determined by the materials requiring analysis. The main aspect though, is that this whole field is moving very rapidly, creating an impressive toolset for materials characterization.

References

41.1 P.F. Fewster, N.L. Andrew: Interpretation of the diffuse scattering close to Bragg peaks by x-ray topography, J. Appl. Cryst. **26**, 812–819 (1993)

41.2 P.F. Fewster: Multicrystal x-ray diffraction of heteroepitaxial structures, Appl. Surf. Sci. **50**, 9–18 (1991)

41.3 P. Scherrer: Bestimmung der Grösse und der inneren Struktur von Kolloidteilchen mittels Röntgenstrahlen, Nachr. Ges. Wiss. Göttingen **2**, 98–100 (1918), in German

41.4 G.K. Williamson, W.H. Hall: X-ray line broadening from filed aluminium and wolfram, Acta Metall. **1**, 22–31 (1953)

41.5 L. Vegard: Die Konstitution der Mischkristalle und die Raumfüllung der Atome, Z. Phys. **5**, 17–26 (1921), in German

41.6 P.F. Fewster: *X-ray Scattering from Semiconductors*, 2nd edn. (Imperial College Press, London 2003)

41.7 P.F. Fewster: A high-resolution multiple-crystal multiple-reflection diffractometer, J. Appl. Cryst. **22**, 64–69 (1989)

41.8 P.F. Fewster: Reciprocal space mapping, Crit. Rev. Solid State Mater. Sci. **22**, 69–110 (1997)

41.9 P.P. Ewald: Zur Theorie der Interferenzen der Röntgenstrahlen in Kristallen, Phys. Z. **14**, 465–472 (1913), in German

41.10 C.G. Darwin: The theory of x-ray reflextion, Philos. Mag. **27**, 315–333 (1914)

41.11 C.G. Darwin: The theory of x-ray reflextion, Philos. Mag. **27**, 675–690 (1914)

41.12 J.A. Prins: Die Reflexion von Röntgenstrahlen an absorbierenden idealen Kristallen, Z. Phys. **63**, 477–493 (1930), in German

41.13 P.F. Fewster: X-ray diffraction from low dimensional structures, Semicond. Sci. Technol. **8**, 1915–1934 (1993), review article

41.14 J.M. Vandenberg, M.B. Panish, H. Temkin, R.A. Hamm: Intrinsic strain at lattice-matched $Ga_{0.47}In_{0.53}As/InP$ interfaces as studied with high-resolution x-ray diffraction, Appl. Phys. Lett. **53**, 1920–1922 (1988)

41.15 V. Holý, U. Pietsch, T. Baumbach: *High-Resolution X-ray Scattering from Thin Films and Multilayers* (Springer, Berlin, Heidelberg 1999)

41.16 M. Born, E. Wolf: *Principles of Optics*, 6th edn. (Cambridge Univ. Press, Cambridge 1980)

41.17 L.G. Parrat: Surface studies of solids by total reflection of x-rays, Phys. Rev. **95**, 359–369 (1954)

41.18 P.F. Fewster: A 'beam-selection' high-resolution x-ray diffractometer, J. Appl. Cryst. **37**, 565–574 (2004)

41.19 P. Kidd: Investigation of the precision in x-ray diffraction analysis of VCSEL structures, J. Mater. Sci. Mater. Electron. **14**, 541–550 (2003)

41.20 J. Zhang, J.H. Neave, X.B. Li, P.F. Fewster, H.A.W. El Mubarek, P. Ashburn, I.Z. Mitrovic, O. Buiu, S. Hall: Growth of SiGeC layers by GSMBE and their characterization by x-ray techniques, J. Cryst. Growth **278**, 505–511 (2005)

41.21 S.K. Sinha, E.B. Sirota, S. Garoff, H.B. Stanley: X-ray and neutron scattering from rough surfaces, Phys. Rev. B **38**, 2297–2311 (1988)

41.22 E.A. Kondrashkina, S.A. Stepanov, R. Opitz, M. Schmidbauer, R. Köhler, R. Hey, M. Wassermeier, D.V. Novikov: Grazing-incidence x-ray scattering from stepped interfaces in AlAs/GaAs superlattices, Phys. Rev. B **56**, 10469–10482 (1997)

41.23 I.D. Feranchuk, S.I. Feranchuk, A.P. Ulyanenkov: Self-consitent description of x-ray reflection and diffuse scattering from rough surfaces, unpublished work presented at XTOP 2006, Baden-Baden (2006)

41.24 Y. Yoneda: Anomalous surface reflection of x-rays, Phys. Rev. **131**, 2010–2013 (1963)

41.25 P. Kidd, P.F. Fewster, N.L. Andrew: Interpretation of the diffraction profile resulting from strain relaxation in epilayers, J. Phys. D Appl. Phys. **28**, A133–A138 (1995)

41.26 V. Kaganer, R. Köhler, M. Schmidbauer, R. Opitz, B. Jenichen: X-ray diffraction peaks due to misfit dislocations in heteroepitaxial structures, Phys. Rev. B **55**, 1793–1810 (1997)

41.27 V. Kaganer, O. Brandt, A. Trampert, K.H. Ploog: X-ray diffraction peak profiles from threading dislocations in GaN epitaxial films, Phys. Rev. B **72**, 045423–045434 (2005)

41.28 S. Daniš, V. Holý: Diffuse x-ray scattering from misfit and threading dislocations in $PbTe/BaF_2/Si(111)$ thin layers, Phys. Rev. B **73**, 014102–014107 (2006)

41.29 V. Holý, T. Baumbach, D. Lübbert, M. Elyyan, P. Mikulík: Diffuse x-ray scattering from dislocations in epitaxial layers – Beyond the ensemble

averaging, unpublished work presented at XTOP 2006, Baden-Baden (2006)
41.30 V. Holý, J. Strangl, G. Springholz, M. Pinczolits, G. Bauer, I. Kegel, T.H. Metzger: Lateral and vertical ordering of self-assembled PbSe quantum dots studied by high-resolution x-ray diffraction, Physica B **283**, 65–68 (2000)
41.31 A.A. Darhuber, P. Schittenhelm, V. Holý, J. Strangl, G. Bauer, G. Abstreiter: High-resolution x-ray diffraction from multilayered self-assembled Ge dots, Phys. Rev. B **55**, 15652–15663 (1997)
41.32 M. Schmidbauer, T. Wiebach, H. Raidt, M. Hanke, R. Köhler, H. Wawre: Self-organised ordering of $Si_{1-x}Ge_x$ nanoscale islands studied by grazing incidence small-angle x-ray scattering, J. Phys. D Appl. Phys. **32**, A230–A233 (1999)
41.33 J. Stangl, V. Holý, T. Roch, A. Daniel, G. Bauer, J. Zhu, K. Brunner, G. Abstreiter: Grazing incidence small-angle x-ray scattering study of buried and free-standing SiGe islands in a SiGe/Si superlattice, Phys. Rev. B **62**, 7229–7236 (2000)
41.34 P.F. Fewster, V. Holý, N.L. Andrew: Detailed structural analysis of semiconductors with x-ray scattering, Mater. Sci. Semicond. Process. **4**, 475–481 (2001)
41.35 P.F. Fewster, V. Holý, D. Zhi: Composition determination in quantum dots with in-plane scattering compared with STEM and EDX analysis, J. Phys. D Appl. Phys. **36**, A217–A221 (2003)
41.36 T.H. Metzger, I. Kegel, R. Paniago, J. Piesl: Grazing incidence x-ray scattering: an ideal tool to study the structure of quantum dots, J. Phys. D Appl. Phys. **32**, A202–A207 (1999)
41.37 D. Grigoriev, M. Hanke, M. Schmidbauer, P. Schäfer, O. Konovalov, R. Köhler: Grazing incidence x-ray diffraction at free-standing nanoscale islands: fine structure of diffuse scattering, J. Phys. D Appl. Phys. **36**, A225–A230 (2003)
41.38 I.A. Vartanyants, I.K. Robinson: Imaging of quantum array structures with coherent and partially coherent diffraction, J. Synchrotron Rad. **10**, 409–415 (2003)
41.39 I.A. Vartanyants, I.K. Robinson, J.D. Onken, M.A. Pfeifer, G.J. Williams, F. Pfeiffer, H. Metzger, Z. Zhong, G. Bauer: Coherent x-ray diffraction from quantum dots, Phys. Rev. B **71**, 245302.1–245302.9 (2005)
41.40 P. van der Sluis, J.J.M. Binsma, T. van Dingen: High resolution x-ray diffraction of periodic surface gratings, Appl. Phys. Lett. **62**, 3186–3188 (1993)
41.41 T. Baumbach, M. Gailhanhou: X-ray diffraction from epitaxial multilayered surface gratings, J. Phys. D Appl. Phys. **28**, 2321–2327 (1995)
41.42 Y. Zhang, V. Holý, J. Strangl, A.A. Darhuber, P. Mikulík, S. Zerlauth, F. Schäffler, G. Bauer, N. Darowski, D. Lübbert, U. Pietsch: Strain relaxation in periodic arrays of Si/SiGe quantum wires determined by coplanar high-resolution x-ray diffraction and grazing incidence diffraction, J. Phys. D Appl. Phys. **32**, A224–A229 (1999)
41.43 A. Ulyanenkov, N. Darowski, J. Grenzer, U. Pietsch, H. Wang, A. Forchel: Evaluation of strain distribution in freestanding and buried lateral nanostructures, Phys. Rev. B **60**, 16701–16714 (1999)
41.44 P.F. Fewster: A 'static' high-resolution x-ray diffractometer, J. Appl. Cryst. **38**, 62–68 (2005)
41.45 M. Sauvage-Simkin, Y. Garreau, R. Pinchaux, M.B. Véron, J.P. Landesman, J. Nagle: Commensurate and incommensurate phases at reconstructed (In,Ga)As(001) surfaces: x-ray diffraction evidence for a composition lock-in, Phys. Rev. Lett. **75**, 3485–3488 (1988)
41.46 S.J. Barnett, C.R. Whitehouse, A.M. Keir, G.F. Clark, B. Usher, B.K. Tanner, M.T. Emeny, A.D. Johnson: X-ray topography of lattice relaxation in strained layer semiconductors: Post growth studies and a new facility fo in-situ topography during MBE growth, J. Phys. D Appl. Phys. **26**, A45–A49 (1993)
41.47 D.K. Satapathy, B. Jenichen, W. Braun, V. Kaganer, L. Däweritz, K.H. Ploog: In situ grazing incidence x-ray diffraction study of strain evolution during growth and post growth annealing of MnAs on GaAs(113)A, J. Phys. D Appl. Phys. **38**, A164–A168 (2005)
41.48 C. Simbrunner, A. Navarro-Quezada, K. Schmidegg, A. Bonanni, A. Kharchenko, J. Bethke, K. Lishka, H. Sitter: In-situ x-ray diffraction during MOCVD of III-nitrides, Phys. Status Solidi (a) **204**, 2798–2803 (2007)
41.49 P.F. Fewster, G.A. Tye, N.L. Andrew, P. Kidd: Using the static diffractometer for micro-high-resolution x-ray scattering, unpublished work presented at XTOP 2006, Baden-Baden (2006)
41.50 T. Baumbach, D. Lübbert, V. Holý, P. Mikulík, L. Helfen, P. Pernot, M. Elyyan, S. Keller, B. Heskell, J. Speck: Epitaxial lateral overgrowth of GaN studied by x-ray micro-diffraction imaging, unpublished work presented at XTOP 2006, Baden-Baden (2006)
41.51 D. Lübbert, C. Ferrari, P. Mikulík, P. Pernot, L. Helfen, N. Verdi, D. Korytár, T. Baumbach: Distribution and Burgers vectors of dislocations in semiconductor wafers investigated by rocking-curve imaging, J. Appl. Cryst. **38**, 91–96 (2005)

42. X-Ray Topography Techniques for Defect Characterization of Crystals

Balaji Raghothamachar, Michael Dudley, Govindhan Dhanaraj

X-ray topography is the general term for a family of x-ray diffraction imaging techniques capable of providing information on the nature and distribution of structural defects such as dislocations, inclusions/precipitates, stacking faults, growth sector boundaries, twins, and low-angle grain boundaries in single-crystal materials. From the first x-ray diffraction image, recorded by Berg in 1931, to the double-crystal technique developed by Bond and Andrus in 1952 and the transmission technique developed by Lang in 1958 through to present-day synchrotron-radiation-based techniques, x-ray topography has evolved into a powerful, nondestructive method for the rapid characterization of large single crystals of a wide range of chemical compositions and physical properties, such as semiconductors, oxides, metals, and organic materials. Different defects are readily identified through interpretation of contrast using well-established kinematical and dynamical theories of x-ray diffraction. This method is capable of imaging extended defects in the entire volume of the crystal and in some cases in wafers with devices fabricated on them. It is well established as an indispensable tool for the development of growth techniques for highly perfect crystals (for, e.g., Czochralski growth of silicon) for semiconductor and electronic applications. The capability of in situ characterization during crystal growth, heat treatment, stress application, device operation, etc. to study the generation, interaction, and propagation of defects makes it a versatile technique to study many materials processes.

42.1 **Basic Principles of X-Ray Topography** 1426
42.1.1 Contrast 1426
42.1.2 Resolution 1427

42.2 **Historical Development of the X-Ray Topography Technique** 1428

42.3 **X-Ray Topography Techniques and Geometry** 1430
42.3.1 Conventional X-Ray Topography Techniques 1430
42.3.2 Synchrotron-Radiation-Based X-Ray Topography Techniques 1431
42.3.3 Recording Geometries 1435

42.4 **Theoretical Background for X-Ray Topography** 1435
42.4.1 Limitation of Kinematical Theory of X-Ray Diffraction 1436
42.4.2 Dynamical Theory of X-Ray Diffraction 1436

42.5 **Mechanisms for Contrast on X-Ray Topographs** 1440
42.5.1 Orientation Contrast from Subgrains and Twins 1440
42.5.2 Extinction Contrast 1441

42.6 **Analysis of Defects on X-Ray Topographs** 1445
42.6.1 Basic Dislocation Analysis 1445
42.6.2 Contrast from Inclusions 1446
42.6.3 Contrast Associated with Cracks 1448

42.7 **Current Application Status and Development** 1449

References 1450

The industrial demand for high-quality single crystals has increased manyfold following the inventions of the transistor and the laser, which led to a wide range of applications [42.1]. In semiconductor technology the increasing density of devices on individual integrated circuits requires a high level of homogeneity in terms of

the chemical composition and structural perfection of crystals [42.2]. The presence of imperfections in crystals used in solid-state lasers, including semiconductor lasers, and nonlinear and electrooptical devices, negatively impacts on their reliable operation [42.3]. The study of such imperfections or defects is also important from the point of view of understanding the influence of imperfections on crystal growth processes, and conversely feedback can be used to develop higher-quality crystals [42.4]. Hence it is essential for process engineers and crystal growers to characterize and have knowledge of the amount, distribution, and nature of defects present in crystals. Even though there are several techniques based on optical, electron, and atomic force microscopy or x-ray imaging to reveal dislocations and other defects, none of these methods are ideal. These characterization methods have different ranges of capabilities and limitations, and hence should normally be used complementarily. However, compared with all other techniques, x-ray topography is powerful enough to image extended defects in the entire volume of the crystal and in some cases in wafers with devices fabricated on them. The capability of this technique to image both defects intersecting the wafer surface as well as those located in the bulk combines features specific to chemical etching and transmission electron microscopy (TEM), respectively. Using synchrotron radiation, the x-ray topography technique can be applied to study large wafers up to 300 mm diameter [42.5], crystal plates, and even as-grown boules weighing several kilograms [42.6]. Presently, x-ray topography has become a very important tool in fundamental research as well as in industrial applications, as described in detail in recent literature [42.7, 8].

X-ray topography is the general term for a family of x-ray diffraction imaging techniques capable of providing information on the nature and distribution of structural defects in single crystal materials. The complete name *x-ray diffraction topography* is slightly more informative, indicating that the technique is concerned with the topography of the internal diffracting planes, i. e., local changes in the spacing and rotations of these planes, rather than with external surface topography. Even though the name *x-ray topography* does not articulate the correct meaning of the technique, it is commonly used by researchers, perhaps for convenience or as a continuation from its first usage in early literature [42.9]. This technique is usually nondestructive and suitable for studying single crystals of large cross-section with thickness ranging from hundreds of micrometers to several millimeters. It is superior to the complementary TEM technique in that it enables imaging of the entire crystal and displaying defects particularly at dislocation densities lower than $10^4\,\mathrm{mm}^{-2}$. In x-ray diffraction topography, a collimated area-filling ribbon of x-rays is incident on the single-crystal sample, at a set Bragg angle, and the corresponding area-filling diffracted beam is projected onto a high-resolution x-ray film or detector. The two-dimensional diffraction spot thus obtained constitutes an x-ray topograph and precisely displays the variation of the diffracted intensity as a function of position depending upon the local diffracting power as well as the prevailing overall diffraction conditions. Local diffracting power is affected by the distorted regions surrounding a defect, leading to differences in intensities between these regions and the surrounding more perfect regions. This intensity variation gives rise to contrast, and different defect types can be characterized from the specific contrast produced by the way they distort the local crystal lattice and thereby the local diffracting power. The absence of magnification enables the correlation of the relative position of the image of a defect with its location inside the crystal. Quantitative information such as the line direction and Burgers vector of a dislocation can be obtained by detailed interpretation of the variations in contrast obtained under different diffraction conditions. Such interpretation requires an understanding of the mechanisms of contrast formation. These mechanisms are sensitive functions of the diffraction conditions and are derived from the kinematical and dynamical theories of x-ray diffraction [42.10].

42.1 Basic Principles of X-Ray Topography

42.1.1 Contrast

In general, the individual spots obtained from Laue diffraction patterns from crystals do not have uniform contrast. The localized variations in the intensity within any individual diffracted spot arise from the deviation in structural uniformity in the lattice planes causing the spot, and this forms the basis for the x-ray topographic technique. The real information available in an x-ray topograph is manifested in the form of contrast within

the projected diffraction spot. This topographic contrast arises from differences in the intensity of the diffracted beam as a function of spatial coordinate inside the crystal. The diffracted intensity is a sensitive function of local crystal perfection. For example, under the correct diffraction conditions, highly mosaic regions of a crystal (i. e., those regions comprising small subgrains which are slightly tilted with respect to the perfect crystal) will diffract kinematically, whereas nearly perfect regions of the crystal will diffract dynamically. Kinematical diffraction is characterized by an absence of the primary extinction effects, which are the essence of dynamical diffraction. These primary extinction effects, created by the interference between diffracted waves and incident waves inside the crystal, lead to an overall reduction in diffracted intensity for a perfect crystal over the mosaic crystal. Conversely, there is an increase in diffracted intensity as the degree of mosaicity increases, eventually reaching the limit of the *ideally imperfect crystal* [42.11]. The difference between the intensities diffracted from one region of the crystal which diffracts kinematically and another which diffracts dynamically is one of the ways in which dislocations can be rendered visible in topography. However, the situation is not as simple as stated here. Even for a crystal which diffracts dynamically, the diffracted intensity is a sensitive function of the local distortion in the crystal. In addition, the type of contrast that arises from a particular type of distortion is also a function of the absorption conditions in the crystal. The effects of such conditions on contrast mechanism will be discussed in Sect. 42.5 with emphasis on mechanisms which give rise to dislocation images.

42.1.2 Resolution

The contrast from individual defects will only be clearly discernible if the spatial resolution is adequate. There is no magnification involved in topography and spatial resolution is controlled solely by geometrical factors. The Bragg law is defined as

$$\lambda = 2d \sin\theta_B \,, \tag{42.1}$$

where λ is the x-ray wavelength, d is the spacing between the diffracting planes, and θ_B is the Bragg angle, which is the angle between the incident rays and the diffracting planes. When the Bragg condition is satisfied, possible incident and diffracted beam directions lie, diametrically opposed, on the surface of a cone with semi-apex angle $90° - \theta_B$, the axis of which is the active reciprocal lattice vector $\boldsymbol{g}$ (i. e., the normal to the diffracting planes). Therefore if an x-ray source has finite size, which it always does, the possibility arises that at a given point in the crystal the diffracting planes will receive radiation, at the correct Bragg angle, from several different points located at different positions on the source (which produce rays lying on the surface of the Bragg cone). The locus of these points on the source is the arc defined by the intersection between the operative Bragg cone and the source surface. This acceptance of rays with a finite vertical divergence imparts a finite vertical divergence to the diffracted beam, which emanates from the point of interest in the crystal, and hence gives rise to a blurring effect in the resultant *image* of the point. In fact the diffracted rays emanating from the point of interest in the crystal will lie on the arc defined by the intersection between the same Bragg cone and the detector plane (the length of this arc will, of course, be defined by the divergence angle of the accepted rays from the arc on the source). This is illustrated in Fig. 42.1, which shows how an image of a point P in the crystal, recorded with rays which emanate from points lying on the arc abc on the source, is spread over the arc a′b′c′ on the detector. Therefore, the angle subtended by the longest dimension of the source in the direction perpendicular to the plane of incidence (defined by the incident and diffracted beam directions, and the active reciprocal lattice vector) is the crucial parameter, which should be minimized. The amount of blurring is proportional to the distance between the crystal and detector, and can be approximately written as

$$R = \frac{\mathrm{SD}}{C} \,, \tag{42.2}$$

where R is the blurring (or the effective resolution), S is the maximum source dimension in the direction perpendicular to the plane of incidence, D is the specimen–film distance, and C is the source–specimen distance. In the plane of incidence, diffraction occurs over a finite range of angles determined by the acceptance angle of the crystal in that plane or the rocking curve width. This finite acceptance angle means that a given point in the crystal receives radiation from a finite width on the source (for the case of a horizontal plane of incidence). This width can be determined by back-projecting the fan of incident rays accepted at a point in the crystal to the source itself. Again a finite divergence, this time in the horizontal plane, can be attributed to the diffracted beam, and again a blurring effect will be evident. However, as can be verified by simple calculation, the limiting factor in determining spatial resolution is

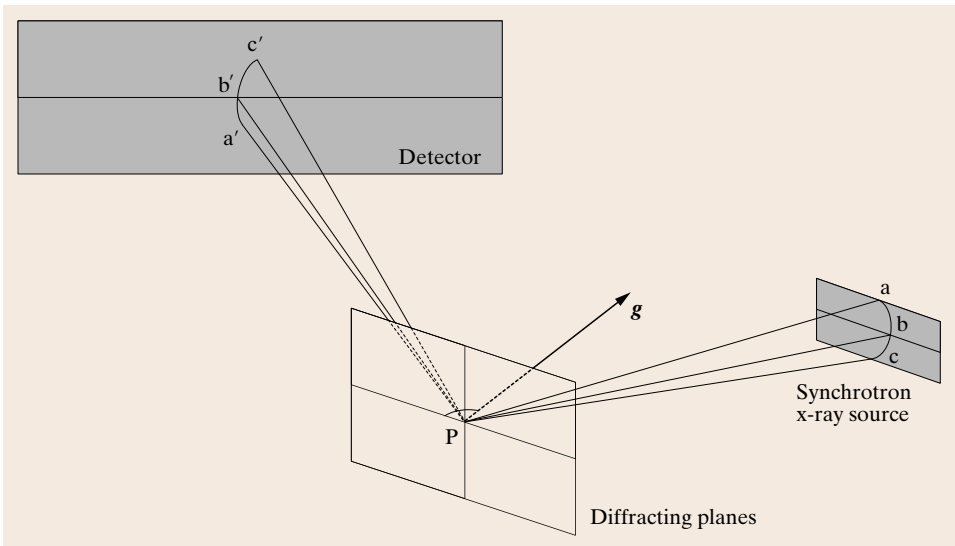


Fig. 42.1 Schematic diagram showing the effect of finite source dimension on resolution; $\boldsymbol{g}$ is the active reciprocal lattice vector. The angle bPg, which is the semi-apex angle of the Bragg cone, is $90° - \theta_B$, where θ_B is the Bragg angle

the finite source height (i. e., the source dimension perpendicular to the incidence plane), rather than the finite effective source width (i. e., the dimension parallel to the incidence plane).

Since the spatial resolution is proportional to the distance between the crystal and detector, this distance should be minimized. However, the optimum spatial resolution of the photographic detectors typically utilized in topography is limited, by photoelectron tracking between adjacent grains in the emulsion of the film, to around 1 μm. The specimen–film distance is usually set to yield a calculated spatial resolution which approximately coincides with this. It should be noted that, while this spatial resolution is greatly inferior to that of electron microscopy, it is more than sufficient for high-quality crystals of many materials currently grown.

42.2 Historical Development of the X-Ray Topography Technique

Even though the first topographic image of a single crystal was recorded as early as 1931 [42.12], the real potential of the technique was understood only in 1958 when *Lang* [42.13] demonstrated the imaging of individual dislocations in a silicon crystal. Different topographic geometries were developed independently during this period, namely, the *Berg–Barrett* reflection technique [42.12,14], the double-crystal technique [42.15], and the Lang technique [42.13] and its variant, the scanning oscillator technique [42.16].

In 1931, *Berg* carried out x-ray diffraction imaging on crystal surfaces using characteristic radiation falling at a very low angle to the crystal surface and obtained point-to-point variation of the reflected intensity resulting in a striated image on a photographic plate placed near the crystal. The striated images produced were attributed to plastic deformation in the crystals. Even though it was realized that x-rays could be used as a powerful tool for studying inhomogeneities in crystals, no further attempt was made to use this tool until *Barrett* [42.14] recorded defect structure images from single crystals of silicon ferrite. *Barrett* improved *Berg's* reflection technique by minimizing the distance between the crystal and the photographic emulsion plate and also by using fine-grain high-resolution x-ray film. This technique is commonly known as Berg–Barrett reflection topography and is used to study large-size crystals. *Wooster* and *Wooster* [42.17] obtained topographic images revealing the defect structure from diamond surfaces using characteristic copper radiation. With further improvement in this technique, *Newkirk* [42.18] showed that individual dislocations could be resolved and their Burgers vectors could be experimentally determined. The *Schulz* technique [42.19] is another reflection topographic technique that uses white radiation from a point-focus x-ray generator.

The divergent x-ray beam diffracted by the crystal is recorded on a film. The white beam ensures diffraction from all the misoriented regions in the crystal. This simple technique can be used to reveal misorientations quickly in the form of separations and overlaps in the image.

Bond and *Andrus* [42.15] developed a high-resolution double-crystal technique for studying structural imperfections on the habit faces of natural quartz crystals. In this technique, x-rays from a line-focus source are Bragg-reflected from a highly perfect monochromator crystal and then diffracted from the specimen crystal in either reflection or transmission mode. High sensitivity is achieved in this technique because the first crystal further narrows the divergence of the slit-collimated beam, which subsequently probes the specimen crystal. This technique is highly sensitive to lattice misorientations: lattice tilts and lattice parameter changes down to 10^{-8} can be detected. The sensitivity can be increased further by introducing another monochromator or beam-conditioner crystal, and rocking curve measurements can also be carried out to assess the crystal quality.

Ramachandran [42.9] studied cleaved plates of diamond crystals using a white x-ray beam from a tungsten target in the transmission Laue geometry. He was the first to use the term *x-ray topograph* to describe the image contrast of full-size diffracted spots. The Laue spots were recognized as a topographic map of the crystal plate exhibiting variations in structure. He felt that the x-ray topographic technique might provide information on lattice perfection and compared the topographs with luminescence images. The term *x-ray topograph* became established in the literature when *Wooster* and *Wooster* [42.17] used the same term in describing diffraction images obtained from diamond surfaces revealing the defect structure. *Guinier* and *Tennevin* [42.20] studied both orientation and extinction contrast effects in aluminum in transmission mode using a polychromatic x-ray beam. *Tuomi* et al. [42.21] used a similar geometry and performed the first x-ray topography experiments using synchrotron radiation on silicon samples.

Important developments in x-ray topography took place in 1957–1958 when *Lang* developed a transmission technique to image the defects in crystals [42.13, 22]. He examined crystal sections using a narrow collimated characteristic radiation beam, a technique referred to as *section topography*. Linear and planar defects could be mapped with this technique by translating the specimen to known positions and taking a series of section topographs. Using this technique, low-angle boundaries in melt-grown metallic crystals were imaged [42.22]. By incorporating a linear traverse mechanism to translate the crystal plate and film cassette synchronously, *Lang* was able to image individual dislocations in an entire silicon crystal sample using projection topography, a milestone in the field of x-ray topography. In the Lang technique, an incident beam of narrow width and low horizontal divergence (about 4–5 arcmin) is obtained using a slit. The secondary slit placed on the other side of the sample blocks the direct beam while allowing the diffracted beam to pass through. It is possible to obtain an image on the film using K_{α_1} only. This technique can permit up to 1 μm resolution under optimized conditions. Coincidentally, during the same year, individual dislocations in silicon as well as germanium crystals were imaged by *Newkirk* [42.18] using the Berg–Barrett technique, and by *Bonse* and *Kappler* [42.23] using double-crystal topography, respectively. The tremendous development in x-ray topography during this period leading to resolving individual dislocations was also due to the fact that better-quality semiconductor crystals such as silicon and germanium were becoming available due to improved crystal growth technology. After *Lang*'s work [42.13, 22] on imaging of individual dislocations, x-ray topography has become an important quality control tool for assessment of semiconductor wafers before and after device fabrication. However, using this technique it is difficult to record topographs of crystals having elastic strains. Using the scanning oscillator technique developed by *Schwuttke* [42.16], it is possible to record transmission topographs of large-size wafers up to 150 mm in diameter, containing appreciable amount of elastic and/or frozen-in strain. This technique is based on the Lang technique with provision to oscillate the crystal and film simultaneously when the crystal is scanned. The oscillation is chosen to cover the whole reflecting range of the crystal to obtain a complete image of the crystal. This technique was widely used to image dislocations in transmission mode in large-size silicon wafers.

While laboratory-based x-ray topography techniques continue to be used, in recent decades the availability of numerous synchrotron radiation facilities providing intense, low-divergence x-ray beams of wide spectral range has allowed the development of synchrotron x-ray topography techniques. This is discussed in detail in Sect. 42.3.2.

42.3 X-Ray Topography Techniques and Geometry

X-ray topographic techniques can be broadly classified into two types: conventional x-ray topographic techniques based on laboratory x-ray sources such as target tubes, rotating anodes, and similar sources; and synchrotron x-ray topographic techniques based on synchrotron radiation.

42.3.1 Conventional X-Ray Topography Techniques

Berg–Barrett Topography

This is one of the oldest x-ray topography technique and is usually based on Bragg (reflection) geometry. The basic experimental setup (Fig. 42.2a) uses an extended x-ray source and the crystal is aligned such that the diffracting conditions are satisfied for the characteristic K_α lines from a set of Bragg planes. The crystal is cut such that the incident beam makes a small angle to the specimen surface and the diffracted beam emerges almost normal to the specimen surface. The photographic film can be placed very close to the specimen surface (as near as 1 mm). This technique is simple and uses low-cost equipment. The use of reflection geometry permits only imaging up to the x-ray penetration depth below the surface of the crystal, as determined by the extinction distance or absorption distance. This allows for the study of high-dislocation-density materials ($\approx 10^6\,\mathrm{cm}^{-2}$). Limitations of this technique include image doubling due to the K_α doublet as well as significant loss of spatial resolution with increasing specimen–film distance. These limitations can be overcome by appropriate adjustments to the recording geometry. This technique is often used for initial assessment of crystals of new materials [42.24]. The transmission Berg–Barrett method [42.12,14] (Fig. 42.2b) (also know as the Barth–Hosemann geometry) is similar to the Lang technique except that it suffers from high background scattered radiation that limits its use for studying dynamical images.

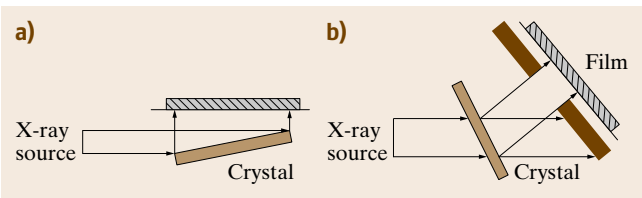


Fig. 42.2a,b Schematic diagrams of (**a**) reflection and (**b**) transmission Berg–Barrett techniques of topography

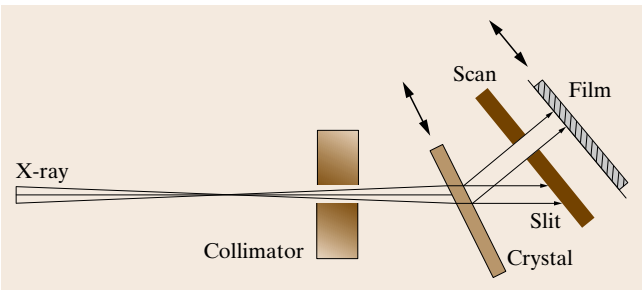


Fig. 42.3 Schematic diagram of the Lang projection technique. The topograph image due to K_{α_1} alone is recorded on the film. The secondary slit blocks the direct beam while allowing the diffracted beam to pass through. The crystal and film are translated synchronously and the whole image of the crystal is recorded

Lang X-Ray Topography

The Lang x-ray topography technique [42.13, 22] is the most widely used laboratory technique and is based on the transmission geometry. Figure 42.3 shows the basic experimental setup, where the x-ray source is collimated to allow diffraction from one K_α line (usually K_{α_1}). The diffracting planes are typically nearly perpendicular to the crystal surface and the diffracted beam passes through a secondary slit that blocks the direct beam before striking the photographic plate. When the incident beam width is small compared with the base of the Borrmann fan formed by the extremes of the diffracted and transmitted beams with the crystal surface (typically about 10 μm), an image of a section through the crystal is obtained. This is the section topograph which allows the study of the three-dimensional defect distribution of defects. By translating both the crystal and film across the stationary beam, an image of the whole crystal can be obtained and this is the projection topograph. However, this results in loss of information available in the section topograph. The section topograph is therefore considered to be more fundamental. Usually, a much wider slit than that used in section topography is used when recording the projection topograph. Commercial Lang cameras consisting of a two-circle goniometer with precision translation stage and adjustable beam slits are available and widely used for laboratory-based x-ray topography [42.24].

42.3.2 Synchrotron-Radiation-Based X-Ray Topography Techniques

The advent of dedicated sources of synchrotron radiation has enabled the development of a new realm of topography known as synchrotron topography. It has numerous advantages over conventional radiation techniques. These advantages derive from the high brightness, tunability, and natural collimation of synchrotron radiation.

White-Beam X-Ray Topography

One of the most important synchrotron topographic techniques developed is white-radiation topography [42.21, 25]. This technique is basically analogous to the Laue transmission technique, except with greatly enhanced capabilities which derive from the aforementioned natural collimation and high intensity of the synchrotron beam. The natural collimation (typically $\approx 2\times10^{-4}$ rad in the vertical plane coupled with an acceptance angle of typically a few milliradians in the horizontal plane) allows the use of very long beamlines (≈ 25 m) to maximize the area of the beam delivered at the sample without incurring significant losses in the total intensity originally available at the tangent point. This situation contrasts with that which would hold for isotropic emission from a conventional source. The large beam area delivered at the sample location allows studies to be carried out on relatively large-scale single crystals, and crystals as large as 150 mm or even 300 mm in diameter can be imaged by using precision translation stages similar to those used in the Lang technique (although exposure times are much shorter). Long beamlines also lead to small angles subtended by the source at points in the specimen, which in turn leads to excellent geometrical resolution capabilities. For example at the Stony Brook Topography Station, on beamline X-19C at the National Synchrotron Light Source (NSLS), the optimum theoretically attainable spatial resolution is $\approx 0.04\,\mu$m/cm of specimen–film distance.

If a single crystal is oriented in the beam, and the diffracted beams are recorded on a photographic detector, each diffraction spot on the resultant Laue pattern will constitute a map of the diffracting power from a particular set of planes as a function of position in the crystal with excellent point-to-point resolution (typically on the order of less than 1 μm). In other words each diffraction spot will be an x-ray topograph (Fig. 42.4a,b). The excellent geometrical resolution capability has another important consequence since it relaxes to a certain extent the requirement of having very small specimen–film distances (< 1 cm) in order to achieve good resolution. Thus, crystals can be surrounded with elaborate environmental chambers, necessitating considerable increases in specimen–detector distances (which would lead to intolerable resolution losses in conventional systems) without significant loss of resolution. The high intensity over the wide spectral range of the radiation emitted has several important advantages. Not only are the exposure times necessary to record a topograph drastically reduced from the order of days on conventional systems to seconds on a synchrotron, but since we have a white beam, a multiplicity of images is recorded simultaneously (Fig. 42.4c). This leads to great enhancement in the rate of data acquisition. The multiplicity of images also enables extensive characterization of strain fields present in the crystal. For example, instantaneous dislocation Burgers vector analysis can become possible by comparison of dislocation images obtained from several different Laue spots. Similarly, lattice rotation in the specimen can be characterized by analysis of the asterism (the spreading of a Laue spot) observed from several different Laue spots. In addition the good signal-to-noise ratios obtained associated with the high intensity also open up the possibility of direct imaging of topographs. Thus it becomes possible to conduct truly dynamic, quasi-real-time studies of crystals subjected to some kind of external stimulus such as applied fields, applied stress, heating, cooling, etc.

Synchrotron topography can also be used to image the surfaces of as-grown boules or large-size crystal plates in reflection geometry. This can reveal the overall distribution of defects and distortion around the cylindrical surface of these crystals. Investigating as-grown boules enables observation of the true microstructures and striations developed during growth and can substantially reduce the time and processing costs in cutting and polishing. Figure 42.5 shows reflection topographic images recorded from an as-grown boule grown by the Bridgman–Stockbarger method in microgravity. Topographs could be recorded covering the entire length of the boule in strips using the synchrotron beam. This boule revealed different defect structures depending on the amount of contact with the crucible wall.

Synchrotron x-ray topography in the reflection geometry can also be used to examine substrate/epilayer systems that have devices fabricated upon them. The features that make up the device topology typically provide contrast on x-ray topographs. The contrast usually

originates from the strain experienced by the crystal at the edges of growth mesas, or metallization layers, although some absorption contrast may also be superimposed on this. Topographs recorded from such structures provide an image of the defect microstructure superimposed on the backdrop of the device topology. Direct comparisons can be drawn between the performance of specific devices and the distribution of defects within their active regions. This has made it possible to determine the influence of threading screw dislocations (closed and hollow core) on device performance [42.26, 27]. The back-reflection geometry is particularly useful here, since it gives a clear image of the distribution of screw dislocations on the background device topology that is imaged with sufficient clarity to unambiguously identify the device. An example of

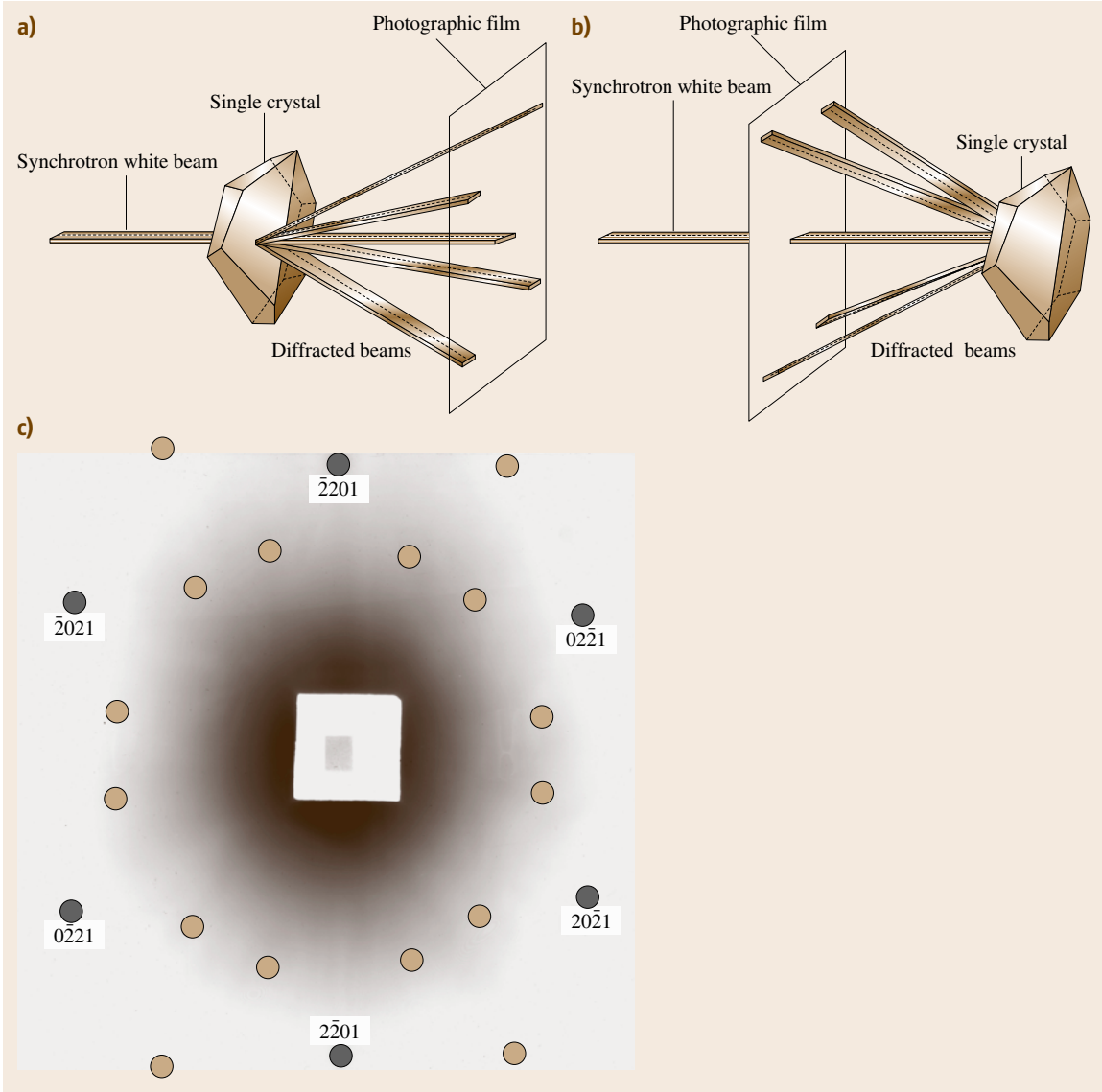


Fig. 42.4a–c Schematic diagram of the white-beam diffraction pattern recorded in **(a)** transmission geometry and **(b)** back-reflection geometry. **(c)** Actual transmission x-ray diffraction pattern recorded from an (0001) AlN single crystal on a SR-45 $8'' \times 10''$ x-ray film

a back-reflection image recorded from a crystal with devices fabricated on it is shown in Fig. 42.6.

Another important advantage of the white-beam topographic technique over conventional topographic techniques is its enhanced tolerance of lattice distortion. The wide spectrum available in the synchrotron beam allows crystals which either exhibit a uniform range of lattice orientation (for example, uniformly bent by a small amount) or which contain several regions of distinctly different orientation (for example, containing subgrains, grains or twinned regions) or regions of different lattice parameter (for example, containing more than one phase or polytype) to be imaged in a single exposure. Analysis of Laue spot shape or Laue spot asterism (the deviation from the shape expected from an undistorted crystal) enables quantitative analysis of lattice rotation. Simultaneous measurement of the variation in diffracted wavelength as a function of position in the Laue spot, for example, using a solid-state detector, also enables determination of any variation in the spacing d of the particular diffracting planes. For those crystals that contain several regions of distinctly different orientations, so-called orientation contrast becomes evident whereby the two neighboring regions of crystal separated by the boundary (a twin boundary or grain boundary, for example) each give rise to diffracted beams which travel in different directions in space, leading to image shifts on the detector. Analysis of the directions of these diffracted beams from the measured image shifts can enable the orientation relationships between the two regions of crystal to be established, leading to, for example, determination of twin laws. An example of a diffraction pattern recorded from a nominally (111) CdZnTe single crystal containing twins is shown in Fig. 42.7. Detailed analysis of the orientation relationships between the segments of the various diffraction spots enables the twin operation to be defined as a 180° rotation about [111].

By comparison, in the conventional topographic case, the maximum tolerable range of misorientation is defined by the convolution of the characteristic line width with the beam divergence. Since both of these quantities are small (typical line widths are of the order of 10^{-3} Å, equivalent to an angular divergence of ≈ 100 arcsec, and typically, beam divergences must be less than ≈ 20 arcsec in order to obtain the necessary angular resolution in the plane of incidence), misorientations greater than a few arcsec lead to the situation

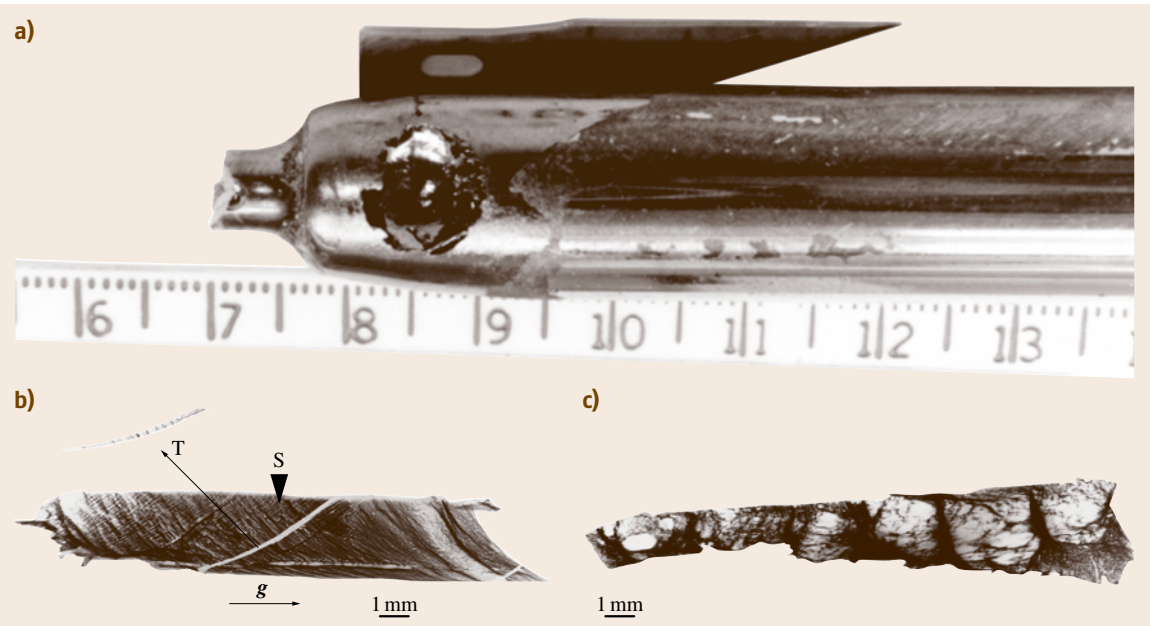


Fig. 42.5 **(a)** Optical picture of an as-grown CdZnTe boule grown in microgravity showing regions with different extent of wall contact. Reflection topographs recorded from the surface of as-grown CdZnTe boule grown in microgravity showing the defect structures. **(b)** Wall-contact region showing dense slip band networks and twins (S: slip bands, T: twin); **(c)** free-surface region showing a lower defect density and no twins

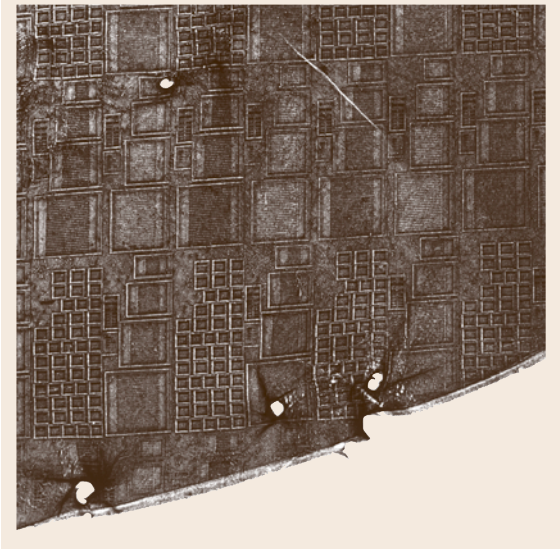


Fig. 42.6 Back-reflection white-beam x-ray topograph recorded from a 6H-SiC single crystal with thyristors fabricated upon it. The *small white spots* distributed over the image are 1*c* and larger screw dislocations. The location of these dislocations with respect to the device topology can be clearly discerned, enabling the influence of the defects on device performance to be determined. The *large white feature* corresponds to damage inflicted by a probe

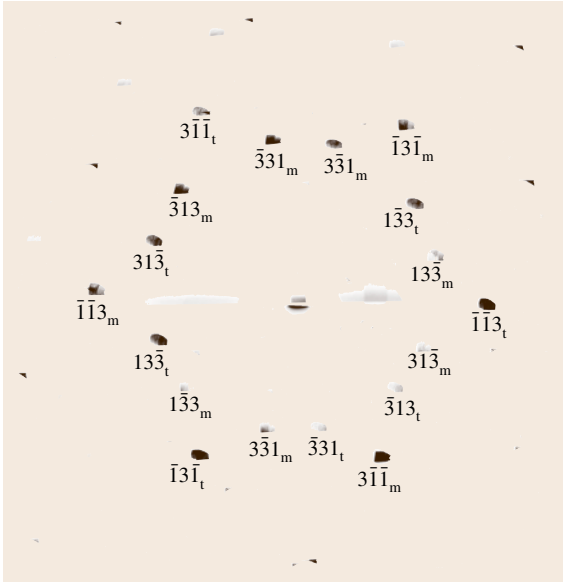


Fig. 42.7 Transmission x-ray diffraction pattern recorded from a CdZnTe crystal showing the presence of a twin. Detailed analysis of the orientation relationships between the segments of the various diffraction spots enables the twin operation to be defined as a 180° rotation about [111]

where only part of the crystal fulfills the diffracting condition at a given time. In other words, Bragg contours are produced. These contours delineate those regions of the crystal which are in the diffracting condition from those which are not. Since their presence means that there are regions not set for diffraction, effectively there are blind spots in the crystal. Clearly this is a very undesirable situation if dynamic-type studies in large single crystals are to be conducted. Similar problems are encountered in monochromatic synchrotron-radiation topography.

Contrast mechanisms operative in white-beam topography are largely similar to those which are operative in Lang topography, with some notable subtle differences [42.28]. However, it has recently been demonstrated that many of these differences are attributable to the large difference in typical specimen–film distances for the two techniques (tens of centimeters for white-beam topography and ≈ 1 cm for Lang topography, see [42.29] for details). Some differences are also attributable to the added complication of harmonic contamination (the presence of several orders of diffraction in a single diffraction spot), which becomes apparent in white-beam topography, potentially leading to a situation where more than one contrast mechanism operates in a single topograph. Harmonic contamination can, however, be avoided by judicious choice of diffraction geometry. In Sect. 42.5, a review of basic contrast mechanisms, pertinent to both Lang and white-beam topography, is presented.

Monochromatic-Beam X-Ray Topography

When a white synchrotron beam is passed through a monochromator, an x-ray topograph is obtained when the crystal is set to the Bragg angle for a specific set of lattice planes for the selected x-ray energy. Images from different atomic planes are acquired by orienting the sample to satisfy the Bragg condition for those planes and orienting the detector to the new scattering angle ($2\theta_B$) to record the image. With monochromatic radiation [42.30] only one topograph is recorded at a time, but the experimenter controls the energy or wavelength of the x-ray beam, the x-ray collimation, the energy or wavelength spread of the x-ray beam, and the size of the incident beam on the sample crystal. The synchrotron beam suffers significant loss of intensity on passing through the monochromator setup. This reduces

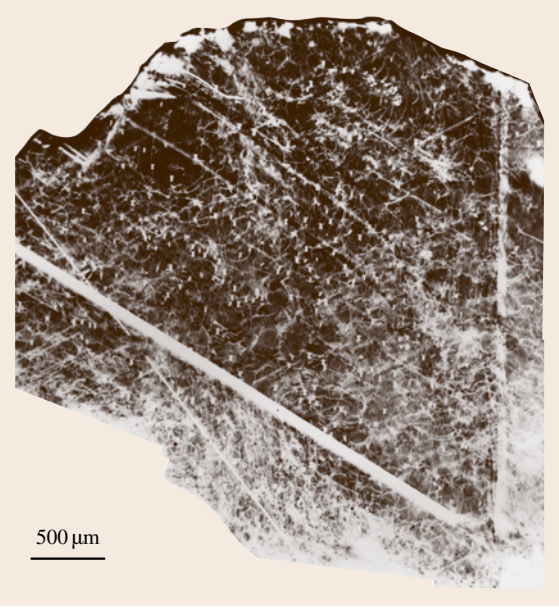


Fig. 42.8 Transmission x-ray topograph recorded from a CdZnTe crystal using a monochromatic synchrotron beam

the probability of radiation damage to the crystal. In addition to wavelength selection, the monochromator also improves angular collimation, and the plane-wave approximation is applicable. Monochromators used at synchrotron-radiation facilities can be either single- or multiple-crystal designs which can condition the x-ray beam to achieve optimal spatial and angular resolutions. An x-ray topograph (Fig. 42.8) recorded in transmission geometry from a CdZnTe single crystal shows dislocations, inclusions, and twinned regions.

42.3.3 Recording Geometries

X-ray topographs are acquired utilizing recording geometries based on either reflection (Bragg) from the surface, or transmission (Laue) through the bulk, of the sample crystal. In general, the reflection topograph geometry is employed for thick crystals or when absorption conditions and/or defect densities are too high to permit the use of transmission geometry. Owing to its surface sensitivity, topography in reflection geometry is also useful for characterization of surface defect structures within semiconductor heterostructures and epitaxial thin films. In all reflection-type geometries, defect information can be obtained from the volume defined by the effective area of the incident beam on the crystal and the penetration depth of the x-ray beam. The penetration depth of x-rays is determined either by the kinematical penetration depth (in imperfect crystals) or by the dynamical penetration depth (in highly perfect single crystals). The kinematical penetration depth (t_p^k) can be determined simply by geometrical relations between the incident and diffracted beams and the sample surface, and is given by

$$t_p^k = \frac{\mu_0(\lambda)}{(\mathrm{cosec}\Phi_0 + \mathrm{cosec}\Phi_H)} , \qquad (42.3)$$

where $\mu_0(\lambda)$ is the linear absorption coefficient and Φ_0 and Φ_H are the angles of the incident and diffracted beams with respect to the surface, respectively. The dynamical penetration depth (t_p^d) is defined to be equal to half the extinction distance ξ_g (42.15). In transmission topography, all the defects within the crystal volume are recorded, provided that the absorption is low enough to permit sufficient transmission through the crystal. Since the x-rays pass through the entire thickness of the sample, this technique is used to characterize the overall bulk defect content of a crystal, such as dislocation networks and inclusions. The high intensities from the synchrotron source enable one to record images from relatively thick specimens using this technique. Laue transmission technique finds limited application in studying crystals of low perfection and high absorption coefficients. The back-reflection technique, commonly used for orienting single crystals, can also be used to record x-ray topographs of crystals containing specific defects such as superscrew dislocations (micropipes) in SiC [42.31]. The grazing Bragg–Laue [42.32] and the grazing-incidence reflection geometries allow precise tuning of the penetration depth for depth profiling studies of epitaxial thin films.

42.4 Theoretical Background for X-Ray Topography

X-ray topography is an imaging technique based on x-ray diffraction. Understanding contrast formation on x-ray topographs requires knowledge of the theory of x-ray diffraction in solids. In this section, the princi-

ples of the kinematical and dynamical theories of x-ray diffraction are briefly discussed with their implications for x-ray topography.

42.4.1 Limitation of Kinematical Theory of X-Ray Diffraction

In the kinematical theory of x-ray diffraction, initially developed to account for the intensities observed in x-ray diffraction studies, the amplitudes of the scattered waves are considered to be at all times small compared with the incident wave amplitude, and scattering from each volume element in the sample is treated as being independent of that from other volume elements. For small crystals, of dimensions less than a micrometer in diameter, and in heavily deformed crystals where the dislocations act to divide the crystal into a mosaic structure of independently diffracting cells, the kinematical theory may be employed satisfactorily to predict diffracted intensities. However, for large single crystals that are also highly perfect, significant discrepancies are found to exist between the measured and theoretically predicted intensities of diffracted beams. The diffracted intensity is predicted to increase continuously with increasing size of crystals and the diffracted intensity is actually predicted to become larger than the incident intensity beyond a certain size (about 1 μm). This is evidently incorrect since it violates the principle of conservation of energy. Under such conditions, the kinematical theory breaks down and the volume elements can no longer be treated as independent of one another. As one can recall from wave theory, x-rays diffracted once from an atomic plane experience a phase change of $\frac{\pi}{2}$. When these waves are scattered again by the backside of the diffracting planes, they propagate in the same direction as the incident beam but are 180° out of phase. This gives rise to an attenuation of the incident intensity due to destructive interference between the primary incident and the secondary scattered beams, which in turn leads to a reduction in the total diffracted beam intensity. This is the so-called primary extinction effect, shown schematically in Fig. 42.9.

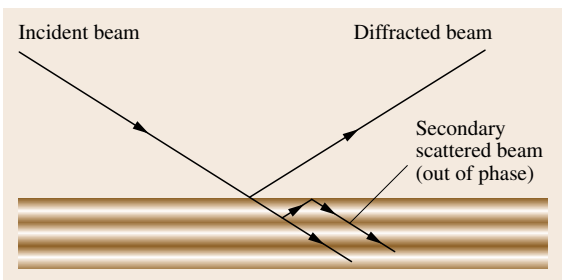


Fig. 42.9 Schematic diagram demonstrating the phenomenon of primary extinction in a perfect crystal. Diffracted beam (90° out of phase) is scattered by the backside of the diffracting planes to produce secondary scattered beam that is 180° out of phase with the incident beam, resulting in attenuation of the intensity of the incident beam, which in turn reduces the total diffracted beam intensity

42.4.2 Dynamical Theory of X-Ray Diffraction

Breakdown of the kinematical theory in large, nearly perfect, single crystals is clearly demonstrated by the phenomenon of primary extinction. For a better understanding, it is necessary to take into account all the wave interactions within the crystal. The dynamical theory of x-ray diffraction considers the total wavefield inside a crystal while diffraction is taking place as a single entity.

The fundamental problem in the dynamical theory is to find solutions to Maxwell's equations in a periodic medium (i. e., the crystal) matched to solutions which are plane waves (the incident $\boldsymbol{k}_0$ and diffracted $\boldsymbol{k}_\mathrm{h}$ x-ray beams). These solutions must reflect the periodicity of the crystal, and such functions are knows as Bloch or lattice functions and may be represented by a Fourier series with appropriate Fourier coefficients.

Maxwell's equations are

$$\nabla\times\boldsymbol{E}=-\frac{1}{c}\frac{\partial\boldsymbol{B}}{\partial t}\quad\text{and}\quad\nabla\times\boldsymbol{H}=-\frac{1}{c}\frac{\partial\boldsymbol{D}}{\partial t}\,, \tag{42.4}$$

where $\boldsymbol{E}$ is the electric field, $\boldsymbol{B}$ the magnetic induction, $\boldsymbol{D}$ the electric displacement, $\boldsymbol{H}$ the magnetic field, and $\boldsymbol{c}$ the velocity of light in vacuum, assuming that the electric conductivity is zero at x-ray frequencies. Properties of the crystal are introduced by representing the crystal as a medium with a periodic, anisotropic, complex dielectric susceptibility $\boldsymbol{\chi}$. Since $\boldsymbol{\chi}$ is a Bloch function, it can be expressed as a Fourier sum over all the reciprocal lattice vectors $\boldsymbol{R}_\mathrm{h}$ as

$$\boldsymbol{\chi}(r)=\sum_\mathrm{h}\boldsymbol{\chi}_\mathrm{h}\exp[-2\pi\mathrm{i}(\boldsymbol{R}_\mathrm{h}\cdot\boldsymbol{r})]\,, \tag{42.5}$$

where the Fourier coefficients are given by

$$\boldsymbol{\chi}_\mathrm{h}=-\frac{r_\mathrm{e}\lambda^2}{\pi V}\boldsymbol{F}_\mathrm{h}\,,$$

where r_e is the classical electron radius, λ is the wavelength of the x-rays, V is the volume of the unit cell,

$\boldsymbol{R}_\mathrm{h}$ is the reciprocal lattice vector, and F_h is the structure factor of the unit cell.

Inside the crystal, the allowed wavevectors should satisfy the Laue equation

$$\boldsymbol{K}_0 + \boldsymbol{R}_\mathrm{h} = \boldsymbol{K}_\mathrm{h} \,, \tag{42.6}$$

where $\boldsymbol{K}_0$ and $\boldsymbol{K}_\mathrm{h}$ are the incident and diffracted wavevectors inside the crystal.

Solutions for the electric displacement $\boldsymbol{D}$ are Bloch or lattice functions of the type

$$\boldsymbol{D} = \sum \boldsymbol{D}_\mathrm{n} \exp[2\pi \mathrm{i}(\nu t - \boldsymbol{K}_\mathrm{h} \cdot \boldsymbol{r})] \,. \tag{42.7}$$

Using the expressions for $\boldsymbol{D}$ (42.7) and $\boldsymbol{\chi}$ (42.5) in Maxwell's equations, we obtain the fundamental equations of the dynamical theory that give the amplitudes of an infinite set of plane waves which together constitute a wavefield that satisfies the wave equation. In x-ray diffraction, only very rarely does more than one reciprocal lattice point provide a diffracted wave of appreciable amplitude. Thus, we need only consider two waves to have appreciable amplitude in the crystal: that associated with the incident wave and that associated with the diffracted wave from a reciprocal lattice vector $\boldsymbol{h}$. The solution of Maxwell's equations inside the crystal is then expressed in terms of the amplitudes $\boldsymbol{D}_0$ and $\boldsymbol{D}_\mathrm{h}$ of these two waves by

$$\left.\begin{aligned} \left[k^2(1+\chi_0) - \boldsymbol{K}_0 \cdot \boldsymbol{K}_0\right] \boldsymbol{D}_0 + k^2 C \chi_{\bar{\mathrm{h}}} \boldsymbol{D}_\mathrm{h} &= 0 \\ k^2 C \chi_\mathrm{h} \boldsymbol{D}_0 + \left[k^2(1+\chi_0) - \boldsymbol{K}_\mathrm{h} \cdot \boldsymbol{K}_\mathrm{h}\right] \boldsymbol{D}_0 &= 0 \end{aligned}\right\} \,, \tag{42.8}$$

where C is the polarization factor and is unity for σ polarization (where vectors $\boldsymbol{D}_0$ and $\boldsymbol{D}_\mathrm{h}$ are perpendicular to the plane containing $\boldsymbol{K}_0$ and $\boldsymbol{K}_\mathrm{h}$, and so are parallel) and $\cos(2\theta_\mathrm{B})$ for π polarization (where vectors $\boldsymbol{D}_0$ and $\boldsymbol{D}_\mathrm{h}$ lie in the plane containing $\boldsymbol{K}_0$ and $\boldsymbol{K}_\mathrm{h}$, perpendicular to $\boldsymbol{K}_0$ and $\boldsymbol{K}_\mathrm{h}$, respectively, but are not parallel to one another). Introducing the deviation parameters α_0 and α_h that express the deviation of the incident and diffracted wavevectors from the kinematic assumption

$$\left.\begin{aligned} \alpha_0 &= 1/2k \left[\boldsymbol{K}_0 \cdot \boldsymbol{K}_0 - k^2(1+\chi_0)\right] \\ \alpha_\mathrm{h} &= 1/2k \left[\boldsymbol{K}_\mathrm{h} \cdot \boldsymbol{K}_\mathrm{h} - k^2(1+\chi_0)\right] \end{aligned}\right\} \,. \tag{42.9}$$

We can write the solution as

$$\alpha_0 \alpha_\mathrm{h} = k^2 C^2 \chi_\mathrm{h} \chi_{\bar{\mathrm{h}}} \,. \tag{42.10}$$

The amplitude ratio R can be written as

$$R = \frac{D_\mathrm{h}}{D_0} = \frac{2\alpha_0}{C\chi_{\bar{\mathrm{h}}} k} = \frac{C\chi_\mathrm{h} k}{2\alpha_\mathrm{h}} = \left(\frac{\alpha_0 \chi_\mathrm{h}}{\alpha_\mathrm{h} \chi_{\bar{\mathrm{h}}}}\right)^{1/2} \,. \tag{42.11}$$

Equations (42.10) and (42.11) are the fundamental equations of the two-beam dynamical theory that allow us to predict the wavefields and their intensities inside (and outside) the crystal.

Dispersion Surface

The wave equations satisfying Maxwell's equations can be represented geometrically by a construction known as the dispersion surface, illustrated in Fig. 42.10. In the kinematical condition, the center of the Ewald sphere is at the Laue point L which is at the intersection of spheres of radius k about the origin O and reciprocal lattice point H in the dispersion plane. In the dynamical condition, the wavevector inside the crystal is corrected for the mean refractive index resulting in a shorter wavevector and therefore the loci of these wavevectors are represented by spheres of diameter $k(1+\chi/2)$ about O and H. This results in a shift of the intersection of these spheres and the dispersion plane from the Laue point L to the Lorentz point L_0 (Fig. 42.10a). Figure 42.10b shows the region around the Laue point at very high magnification ($\approx 10^6$). On the scale of the picture, the spherical sections $\mathrm{O'O''}$ and $\mathrm{H'H''}$ of the projections of the spheres can be approximated as planes. The deviation parameters α_0 and α_h are measured perpendicularly from the planes $\mathrm{O'O''}$ and $\mathrm{H'H''}$, respectively, and denote the *tie point* A at which the tails of wavevectors $\boldsymbol{K}_0$ from O and $\boldsymbol{K}_\mathrm{h}$ from H intersect and diffraction occurs.

Thus, the dispersion equation is the loci of all such tie points and is an equation of a hyperboloid of revolution with OH as its axis. At the diameter points $\alpha_0 = \alpha_\mathrm{h}$ so that the semidiameter of the hyperbola is given by

$$d_\mathrm{h} = \frac{1}{2} kC \left(\chi_\mathrm{h} \chi_{\bar{\mathrm{h}}}\right)^{1/2} \sec\theta \,. \tag{42.12}$$

There are two independent dispersion surfaces for the two polarizations σ and π. The dispersion surface has two branches, the upper one denoted as branch 1 and the lower one as branch 2. Waves from the two branches are in antiphase. The direction of energy flow is described by the Poynting vector, parallel to $\boldsymbol{E} \times \boldsymbol{H}$, and it has been shown that this is perpendicular to the dispersion surface at the tie point [42.33, 34].

The boundary conditions at the crystal surface require that the tangential components of both $\boldsymbol{E}$ and $\boldsymbol{H}$ of the wavevectors should be continuous across the surface. The waves must be matched in amplitude at the crystal surface and in phase velocity parallel to the surface. The wavevectors inside the crystal differ from that

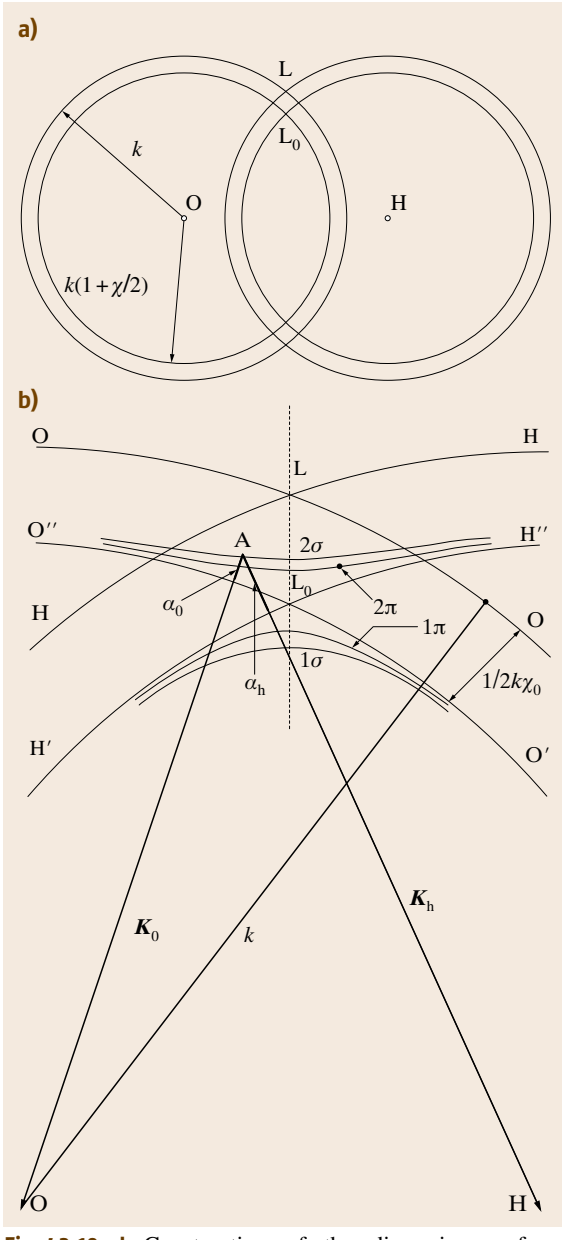


Fig. 42.10a,b Construction of the dispersion surface: **(a)** spheres of radius k and $k(1+\chi/2)$ about the origin O and reciprocal lattice point H in reciprocal space showing the position of the Laue point L and Lorentz point L_0; **(b)** dispersion surface for σ and π polarization states

outside only by a vector normal to the crystal surface

$$\boldsymbol{K}_0 - \boldsymbol{k}_0 = \delta \boldsymbol{n} \,, \qquad (42.13)$$

where $\boldsymbol{n}$ is a unit vector normal to the surface and δ is a scalar variable. A normal drawn line from the tip of the incident wavevector $\boldsymbol{k}_0$ intersects the dispersion surface at the excited tie point and determines the tail of the wavevector $\boldsymbol{K}_0$. In the Laue case (Fig. 42.11a), there are two points excited, one on each branch, labeled A and B. From each tie point, wavevectors directed towards O and H can be generated. There are thus four wavevectors generated in the crystal for each polarization, eight in all. At the exit surface of the crystal, the waves split up into diffracted and forward-diffracted beam and the boundary condition can be similarly determined. In the Bragg case (Fig. 42.11b), the normal from the surface intersects either two tie points on the same branch of the dispersion surface or none at all. The Poynting vectors associated with the two tie points are different; the energy flow from one point is directed into the crystal, but that from the other is directed outwards. The latter therefore does not generate any wavefields inside the crystal and can be ignored. Thus a single wavefield is generated for each polarization. When no tie points are selected, no wavefields are generated inside the crystal and total reflection occurs.

Borrmann Effect

The amplitudes of the wavefields are Bloch functions and modulated with the periodicity corresponding to the Bragg planes. The intensity I of each of this wavefields can be shown to be

$$I = D^2 = D_0^2[1 + R^2 + 2RC\cos(2\pi \boldsymbol{h}\cdot\boldsymbol{r})] \,. \qquad (42.14)$$

The intensity is modulated by the factor $\cos(2\pi \boldsymbol{h}\cdot\boldsymbol{r})\cos(2\pi h\cdot r)$, which has a maxima at $\boldsymbol{h}\cdot\boldsymbol{r} = n$ and minima at $\boldsymbol{h}\cdot\boldsymbol{r} = (2n+1)/2$ for integer n ($\boldsymbol{h}$ is the diffraction vector). That is, maxima and minima of the standing wavefield occur either at, or halfway between, atomic planes (Fig. 42.12). Analysis of (42.14) reveals that the sign of R is opposite for wavefields with tie points on opposite branches of the dispersion surface, and intensity maxima occurs at the atomic plane when R is positive and minima at the atomic plane when R is negative. Maximum modulation occurs when both R and C equal unity, i. e., in centrosymmetric crystal ($\chi_h = \chi_{\bar{h}}$) for the σ polarized wave at the exact Bragg condition; i. e., when tie points excited on the two branches are at the diameter of the dispersion surface. Under such conditions, the wavefield with intensity maxima at the atomic planes (say branch 2) will suffer greater photoelectric absorption because the electron density is maximum at the atomic planes whereas the branch 1 wavefield, with intensity

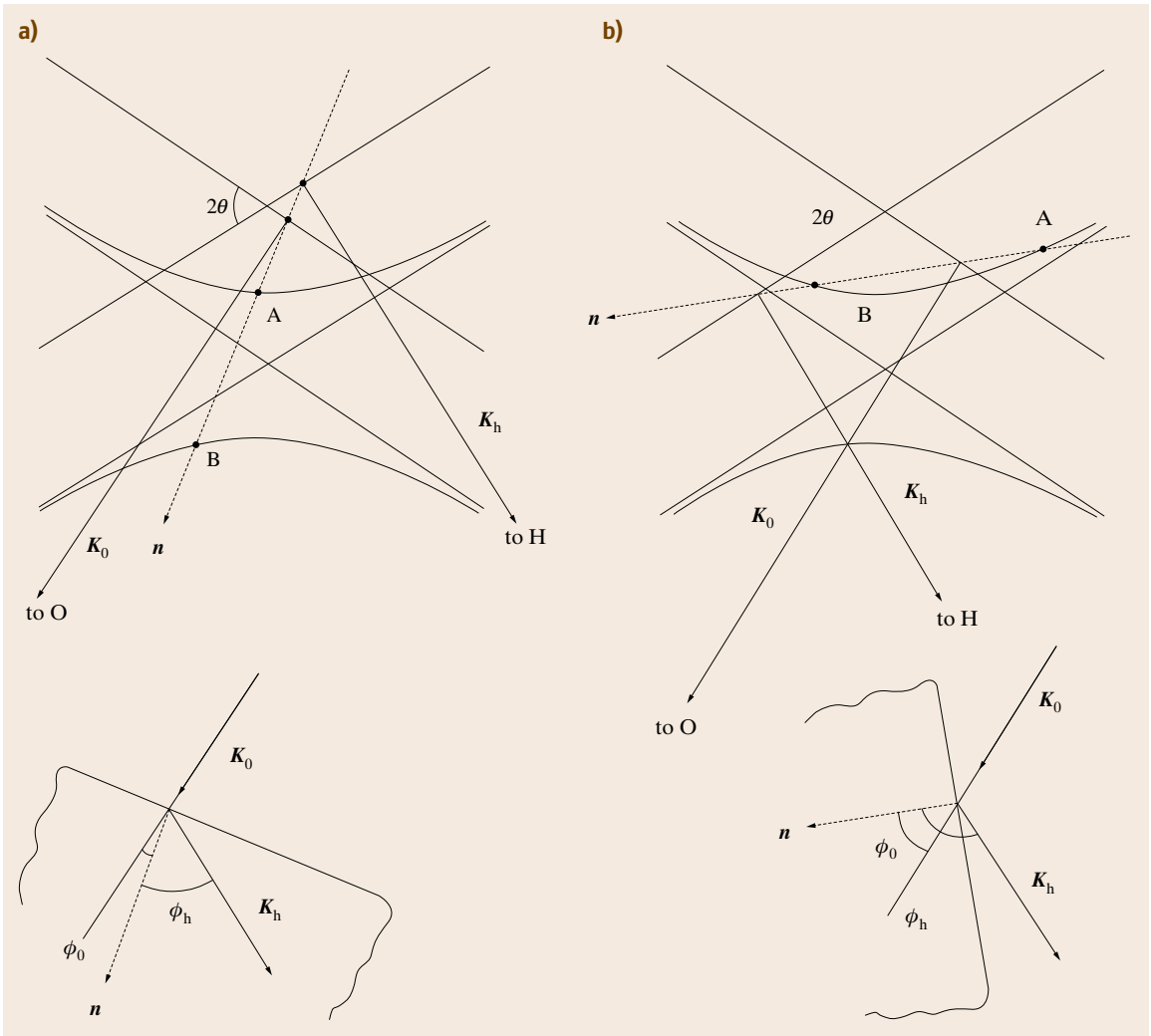


Fig. 42.11a,b Dispersion surface construction showing the tie points excited by an incident wave in **(a)** the Laue geometry: one tie point in each branch (A on branch 1 and B on branch 2) is excited **(b)** the Bragg geometry: two tie points (A and B) on the same branch are excited

maxima between the atomic planes, suffers minimum absorption. This effect is known as anomalous transmission or the Borrmann effect and was discovered by *Borrmann* in calcite [42.35, 36]. Presence of the Borrmann effect is indicative of crystal perfection.

Pendellösung Effect

An incident plane wave excites two tie points on the dispersion surface and generates two Bloch wavefields. The wavevectors associated with these tie points differ. The difference in wavevector leads to a difference in the propagation velocity, and interference effects can occur between the Bloch waves. This gives rise to the production of beats, a phenomenon referred to as Pendellösung. The period of the beats is given by the extinction distance ξ_g, which is the reciprocal of the dispersion surface diameter d_h, for the case of exact fulfillment of the Bragg condition

$$\xi_g = d_h^{-1} = \frac{\cos\theta_B}{Ck(\chi_h\chi_{\bar{h}})^{1/2}} = \frac{\pi V\cos\theta_B}{r_e\lambda C(F_h F_{\bar{h}})^{1/2}}\,, \tag{42.15}$$

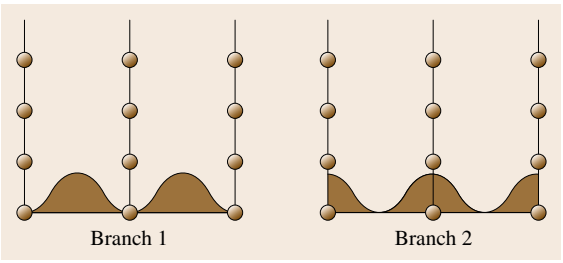


Fig. 42.12 Standing wavefields with a period corresponding to the spacing between the Bragg planes produced at the exact Bragg condition. Branch 1 waves, which have a minimum intensity at the atomic positions, suffer minimal absorption while branch 2 waves, which have a maximum intensity at the atomic positions, are strongly absorbed because of maximum electron density at the atomic planes

where $\boldsymbol{F}_{\mathrm{h}}$ and $\boldsymbol{F}_{\bar{\mathrm{h}}}$ are the structure factors of (hkl) and $(\bar{h}\bar{k}\bar{l})$, respectively. When the Bragg condition is not exactly met, the extinction distance is given by

$$\xi'_{\mathrm{g}} = \frac{\xi_{\mathrm{g}}}{(1+\eta^2)^{1/2}} \,, \tag{42.16}$$

where $\eta = \sin 2\theta_{\mathrm{B}}/(C|\chi_{\mathrm{h}}\chi_{\bar{\mathrm{h}}}|)^{1/2}\Delta\theta$ is the deviation parameter expressing the departure $\Delta\theta$ of the tie points from the exact Bragg condition.

The above results can be extended to cover asymmetric reflections by the introduction of the terms γ_0 and γ_{h}, the cosines of the angles between the surface normal and the incident and diffracted beams, respectively, in the appropriate places.

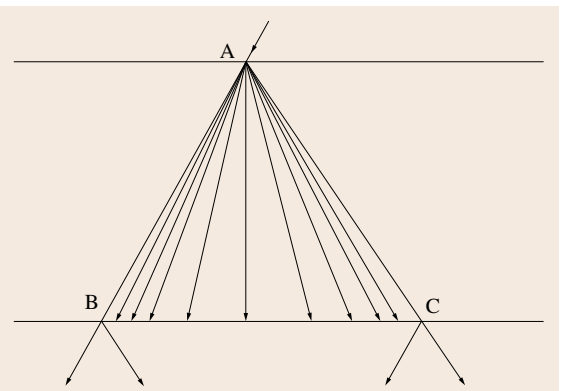


Fig. 42.13 Borrmann fan bounded by the incident (AB) and diffracted (AC) beams showing the distribution of energy for an incident spherical wave that excites all tie points along the dispersion surface

The above analysis applies to an incident plane wave. However, in practice, laboratory sources used have a significant angular divergence because of which the entire dispersion surface is excited simultaneously. This gives rise to energy flow within the Borrmann fan bounded by the incident and diffracted beam directions (Fig. 42.13). Defects at any point within the Borrmann fan may contribute to the change of diffracted intensity at the exit surface of the crystal, and hence to image contrast.

42.5 Mechanisms for Contrast on X-Ray Topographs

Information from an x-ray diffraction topograph is obtained in the form of topographic contrast. As described earlier, local changes in orientation and spacing of crystal lattice planes give rise to local differences in either diffracted beam direction or intensity, which under the appropriate experimental conditions, are manifested as observable contrast on an x-ray topograph. There are two fundamental mechanisms for contrast in x-ray topographs:

1. Orientation contrast
2. Extinction contrast.

Extinction contrast can be further classified into three types:

1. Direct image contrast
2. Dynamical contrast
3. Intermediary contrast.

42.5.1 Orientation Contrast from Subgrains and Twins

Orientation contrast is explicitly defined as resulting from inhomogeneous intensity distributions arising purely from the overlap and/or separation of diffracted x-rays with varying directions. Orientation contrast is observed in crystals containing regions of different orientations such as grains, subgrains, and twins. Misorientations caused by dilations as well as rotations of the lattice can also lead to orientation contrast. Occur-

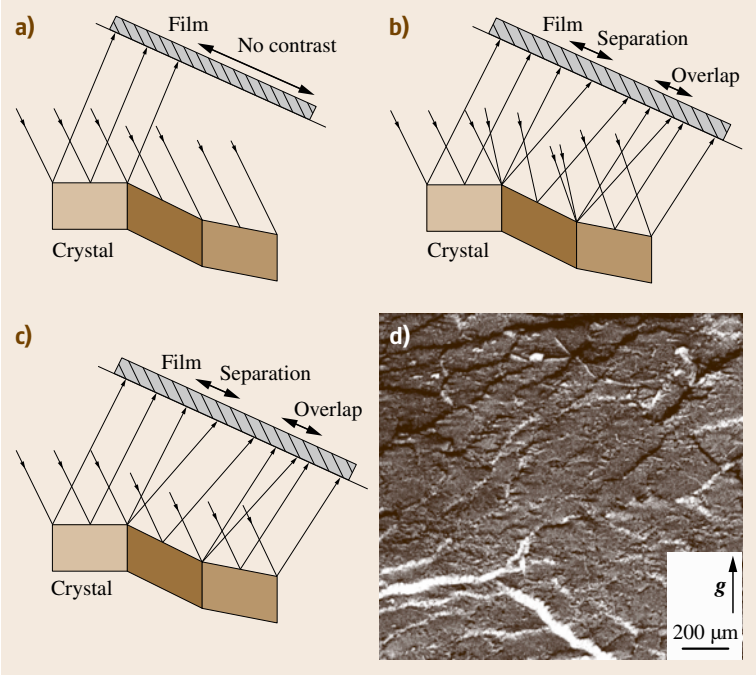


Fig. 42.14a–d Orientation contrast arising from misoriented regions: **(a)** monochromatic radiation (beam divergence < misorientation); **(b)** monochromatic radiation (beam divergence > misorientation); and **(c)** continuous radiation. **(d)** Reflection topograph from an HgCdTe single crystal. The *white bands* correspond to separation between images of adjacent subgrains, while the *dark bands* correspond to image overlap. The tilt angle is estimated at 1–4 arcsec

rence of orientation contrast depends on the nature of x-rays as well as on the nature of the misorientation, as illustrated in Fig. 42.14. If either the incident beam divergence or the range of wavelengths available in the incident x-ray spectrum is smaller than the misorientation between the blocks projected onto the incidence plane (Fig. 42.14a), then only one of the blocks can diffract at a given time. This produces one form of orientation contrast, i. e., the presence or absence of diffracted intensity. Under such conditions, one can adjust the diffraction geometry to bring the other block into the diffraction condition and thus obtain information on the mutual misorientation. However, if either the incident beam divergence or the range of wavelengths available in the incident spectrum allows each block to diffract independently (Fig. 42.14b or c), then the diffracted beams emanating from the individual blocks will travel in slightly different directions and give rise to image overlap if they converge, or separation if they diverge. This produces another form of orientation contrast. The degree of misorientation can be determined by measuring the image shifts (the amount of overlap or separation) as a function of specimen–detector distance on several different reflections. A white-beam x-ray topograph (Fig. 42.14d), recorded from a HgCdTe crystal that shows the overlap and separation of the subgrain images, clearly demonstrates orientation contrast.

42.5.2 Extinction Contrast

Extinction contrast arises when the scattering power around the defects differs from that in the rest of the crystal. Interpretation of this contrast requires understanding of the dynamical diffraction effects as discussed in the previous section (see also [42.37]). In this chapter, extinction contrast formation mechanisms are addressed with respect to images of dislocations on topographs since dislocations are among the most common defects studied by x-ray topography. Topographic contrast of dislocations (as well as other defects) consists of *direct, dynamic, and intermediary images* corresponding to the three different types of extinction contrast. In transmission geometry, which of the three types of image can be observed is determined by the absorption conditions. Absorption conditions are usually defined by the product of the linear absorption coefficient μ and the thickness of the crystal, t, traversed by the x-ray beam, i. e., μt. For topographs recorded under low absorption conditions ($\mu t < 1$), the dislocation image is dominated by the direct image contribution. Under intermediate absorption conditions, i. e., $5 > \mu t > 1$, all three components can contribute, while for high absorption cases ($\mu t > 6$), the dynamical contribution (in this case known as the Borrmann image) dominates.

The Direct Dislocation Image

The direct dislocation image is formed when the angular divergence or wavelength bandwidth of the incident beam is larger than the angular or wavelength acceptance of the perfect crystal [42.38]. Under this condition, only a small proportion of the given incident beam will actually undergo diffraction, with most of the incident beam passing straight through the crystal and simply undergoing normal photoelectric absorption. However, it is possible that the deformed regions around structural defects, such as dislocations and precipitates, present inside the crystal are set at the correct orientation for diffraction provided that their misorientation is larger than the perfect crystal rocking curve width and not greater than the incident beam divergence. The effective misorientation $\delta\theta$ around a defect is the sum of the tilt component in the incidence plane $\delta\varphi$ and the change in the Bragg angle θ_B due to dilation δd and is given by

$$\delta\theta = -\tan\theta_B \frac{\delta d}{d} \pm \delta\varphi \,. \qquad (42.17)$$

Therefore, the distorted region will give rise to a new diffracted beam. Further, if the distorted region is small in size then this region will diffract kinematically and will not suffer the effective enhanced absorption associated with extinction effects to which the diffracted beams from the perfect regions of the crystal are subjected. The enhanced diffracted intensity from the distorted regions compared with the rest of the crystal gives rise to topographic contrast. This is known as direct or kinematical image formation. This form of contrast dominates under low absorption conditions ($\mu t < 1-2$). Generally, the direct dislocation image formed by this extinction contrast model has been used to explain observed contrast features. Although the intensity increase for the direct image was in most cases qualitatively interpreted in a correct way [42.38], detailed measurements of dislocation image width made previously sometimes do not strictly coincide with the predictions of this theory [42.39].

From studies of the direct dislocation images of growth dislocations with large Burgers vectors (superscrew dislocations or micropipes) in x-ray topographs recorded from SiC single crystals, it was recently shown by *Huang* et al. [42.31, 40–42] that the extinction contrast theory alone is incapable of explaining the contrast features associated with superscrew dislocations on synchrotron topographs. By using a simple ray-tracing simulation method, it was demonstrated that the direct images of superscrew dislocations consist mainly of orientation contrast. Moreover, it was shown that this method is also applicable to elementary dislocations, indicating that it is a general phenomenon that orientation contrast makes a significant contribution to the direct dislocation image.

In the orientation contrast model, the mosaic region around the dislocation is divided into a large number of cubic diffraction units with their local misorientations coinciding with the long-range displacement field of the dislocation. These units diffract x-rays kinematically according to their local lattice orientation. Traces of the inhomogeneously diffracted x-rays are projected onto the recording plate to obtain the direct

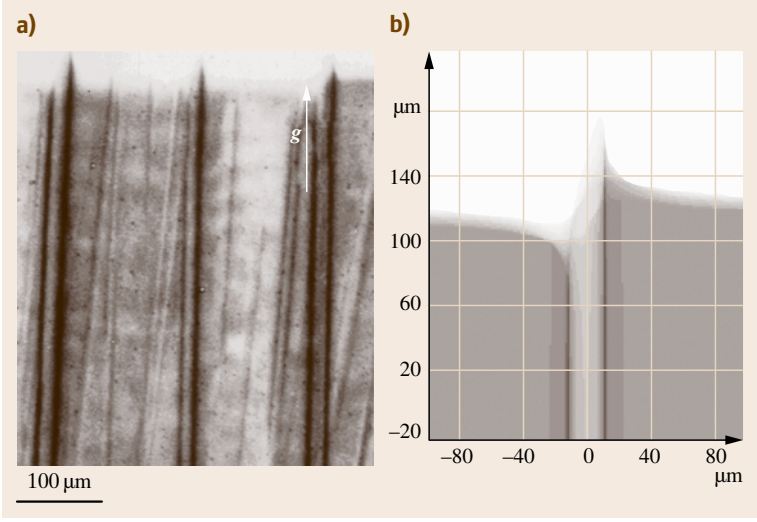


Fig. 42.15 **(a)** Synchrotron transmission topograph of superscrew dislocations in 6H-SiC ($\boldsymbol{g} = 0006$, $\theta_B = 8.5°$, sample–film distance $d = 20$ cm); **(b)** simulation of pure orientation contrast of a $5c$ superscrew dislocation. The simulation parameters are chosen to coincide with the experimental conditions of **(a)**

image. Thus the direct image here is due to orientation contrast arising from the overlap and separation of inhomogeneously diffracted x-rays with continuously varying directions. Figure 42.15a shows several images of giant screw dislocations in SiC recorded in transmission geometry. The Burgers vector of dislocation 1 was independently measured [using optical phase-contrast microscopy, atomic force microscopy (AFM), and other x-ray topography techniques] to be $5c$ ($|\boldsymbol{b}| = 75.85$ Å). The full-width w of the image is around 40 µm, and the separation between the maximum intensity peaks L_0 is around 24 µm. The conventional misorientation contour model predicts an image width of several hundred micrometers, which is clearly not the case. On the other hand, a good correlation is clearly evident between the observed image and the image simulated using the orientation contrast model (Fig. 42.15b). The simulated image consists of pure orientation contrast resulting from the separation or overlap of the inhomogeneously diffracted x-rays. In a similar way, images of superscrew dislocations recorded in back-reflection geometry can be successfully simulated, as shown in Fig. 42.16. It has also been shown that, under low absorption conditions, the above orientation contrast formation mechanism also applies to ordinary dislocations with Burgers vectors smaller than that of micropipes [42.7].

Origins of Dynamical Contrast from Dislocations

As discussed in Sect. 42.4, the dynamical theory of x-ray diffraction obtains solutions to Maxwell's equations in a periodic medium (i. e., the crystal) matched to solutions that are plane waves (the incident and diffracted x-ray beams). The wave equations satisfying Maxwell's equations can be represented geometrically by the dispersion surface [42.43], and a wavefield propagating in the crystal is represented by a tie point on the dispersion surface and comprises two waves corresponding to the incident and diffracted x-ray beams (Figs. 42.9 and 42.10). Dynamical contrast arises from the interaction of this wavefield with the dislocation distortion field (for a review see [42.37]). Under low absorption conditions this dynamical contribution to the dislocation contrast is mostly unobservable due to the fact that the image is dominated by the *direct image* contribution. Dynamical contrast becomes more observable as absorption increases, thereby attenuat-

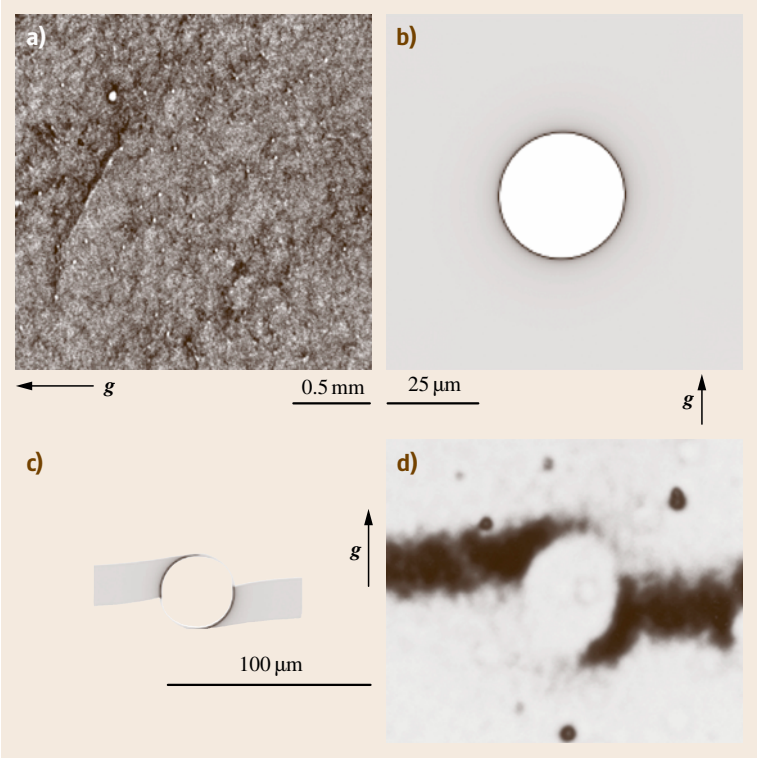


Fig. 42.16 **(a)** Back-reflection synchrotron topograph of a (0001) 4H-SiC wafer showing the circular images of superscrew dislocations ($\boldsymbol{g} = 00016$, $\lambda = 1.25$ Å, sample–film distance $d = 20$ cm); **(b)** enlargement of one of the circular images in **(a)**; **(c)** computer-simulated white-beam back-reflection x-ray section topograph of a screw dislocation (Burgers vector magnitude $b = 3c$) simulated under the diffraction conditions of **(a)**; **(d)** back-reflection x-ray section topograph of a screw dislocation ($b = 3c$)

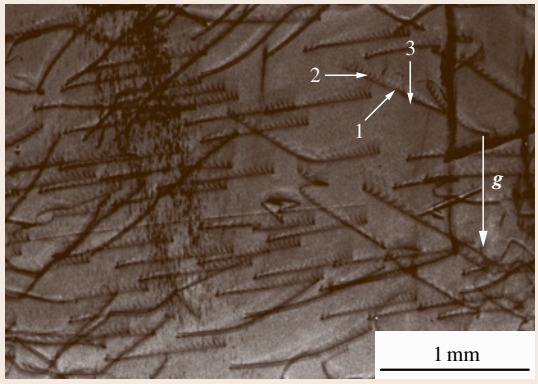


Fig. 42.17 SWBXT transmission topograph ($\boldsymbol{g} = 10\bar{1}0$, $\lambda = 0.75$ Å) recorded from an AlN single crystal under intermediate absorption conditions ($\mu t = 8$) showing the direct (1), dynamical (2), and intermediary (3) images of a dislocation

ing the *direct image* contribution (the kinematically diffracted rays contributing to this image do not survive the absorption). Under these conditions, only wavefields associated with one branch of the dispersion surface (usually branch 1) which are close to the exact Bragg condition survive due to the Borrmann effect [42.35]. These wavefields can survive absorption even when the value of μt is significantly greater than unity.

For a perfect crystal, the incident boundary conditions determine the position of the tie point on the dispersion surface. However for an imperfect crystal, the local lattice distortion can modify the position of the tie point of a wavefield as it passes through the crystal. This can occur by two mechanisms depending on the nature of the distortion field. Tie point migration along the dispersion surface occurs when the wavefield encounters a shallow misorientation gradient upon passing through the crystal (e.g., regions away from the dislocation line). The variation of the misorientation should be less than the rocking curve width over an extinction distance, i. e.,

$$R_c \leq g\xi_g^2 \,, \tag{42.18}$$

where R_c is the radius of the curvature of the reflecting plane and ξ_g is the extinction distance. Under these conditions, both the direction of propagation of the wavefield and the ratio of the amplitudes of its components are changed. This phenomenon can be treated using the so-called eikonal theory [42.44], analogous to its counterpart in geometrical optics, formulated for the case of light traveling through a region of varying refractive index. As a wavefield approaches the long-range distortion field of a dislocation, rays will bend in opposite directions on either side of the core, potentially producing opposite contrast. However, if the dislocation is not located close to the crystal surface, the ray bending experienced above the defect may be compensated for by that experienced below the defect, suggesting that the contrast effects should cancel out. Net contrast is nevertheless observed, since any deviation of the wavefields from the direction of propagation corresponding to the perfect crystal region forces them to experience enhanced absorption, thereby producing a loss of intensity. Under these conditions, the dislocation image will appear white and diffuse. In cases where the dislocation is close to the exit surface of the crystal, the lattice curvature above and below the defect is asymmetric due to the requirements of surface relaxation. This means that the ray bending experienced above the defect is no longer compensated by that experienced below, with the result that opposite contrast is observed from regions either side of the defect.

On the other hand, when the wavefield encounters a sharp misorientation gradient upon passing through the crystal (e.g., regions close to a dislocation core), the strain field would completely destroy the conditions for propagation and force the wavefield to decouple into its component waves. When these component waves reach the perfect crystal on the other side of the defect, they will excite new wavefields (the so-called phenomenon of interbranch scattering). These newly created wavefields will have tie points which are distributed across the dispersion surface, and since only those wavefields with tie points close to the exact Bragg condition survive, they will be heavily attenuated, leading to a loss of intensity from the region surrounding the dislocation. Such images, known as Borrmann images, appear white on a dark background. An example is shown in Fig. 42.17, which is a detail from a white-beam x-ray topograph recorded from an AlN single crystal.

The Intermediary Image

The intermediary image arises from interference effects at the exit surface between the new wavefields created below the defect (as described in Sect. 42.4.2) and the undeviated original wavefield propagating in the perfect regions of crystal. Usually, these images often appear as a bead-like contrast along *direct* dislocation images on projection topographs. Under moderate absorption conditions when the defect (e.g., dislocation line) is inclined to the surface, the intermediary image forms a fan

within the intersections of the exit and entrance surface of the dislocation and has an oscillatory contrast with depth periodicity over an extinction distance ξ_g. Again, this is illustrated in Fig. 42.17.

42.6 Analysis of Defects on X-Ray Topographs

Topography, both synchrotron and conventional, is well suited for analysis of low densities ($< 10^6\,\mathrm{cm}^{-2}$) of dislocations in crystals. The restriction to low densities arises from the fact that topographic dislocation images can be anywhere from around 5 to around 15 µm wide, so that greater densities would lead to image overlap and therefore loss of information.

42.6.1 Basic Dislocation Analysis

Determination of Line Direction

For dislocations created by slip processes, knowledge of the line direction as well as detailed information on the Burgers vector of the dislocations is required to fully assign the active slip system. Knowledge of both line direction and Burgers vector of crystal-growth-induced dislocations is also very important in understanding their origin and for the development of strategies to reduce the density of such dislocations. The projected directions of direct images of growth dislocations have also been used very successfully to compare with line energy calculations designed to determine why particular line directions are preferred by such dislocations in crystals (for a review see [42.45]). The line direction of a dislocation can be obtained by analyzing its direction of projection on two or more topographs recorded with different reciprocal lattice vectors. The use of analytical geometry enables the line direction to be determined either directly from the measured direction of projection [42.46] or indirectly by comparing calculated projected directions of expected dislocation line directions for the material of interest with the measured projection directions [42.47]. Such analysis is most readily carried out on direct images of dislocations since they are the most well defined, although similar analysis can be performed on dynamical or intermediary images.

Determination of Burgers Vector Direction

For sufficiently low dislocation densities, standard Burgers vector analysis, which enables the determination of the direction of the Burgers vector, is readily carried out in the low absorption regime, using the $\boldsymbol{g}\cdot\boldsymbol{b}=0$ criterion for invisibility of screw dislocations, and the combination of $\boldsymbol{g}\cdot\boldsymbol{b}=0$ and $\boldsymbol{g}\cdot\boldsymbol{b}\times\boldsymbol{l}=0$ criterion for invisibility of edge or mixed dislocations (where $\boldsymbol{b}$ is the dislocation Burgers vector and $\boldsymbol{l}$ is the dislocation line direction). These criteria are also used in the analysis

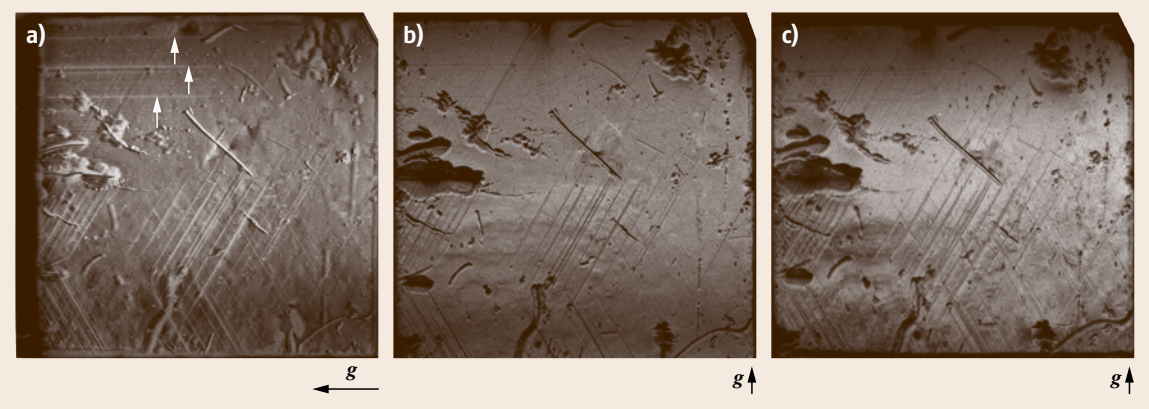


Fig. 42.18a–c Synchrotron white-beam topographs recorded in transmission geometry from a ZnO single crystal containing slip dislocations: **(a)** $\boldsymbol{g}=11\bar{2}0$; **(b)** $\boldsymbol{g}=1\bar{1}00$; **(c)** $\boldsymbol{g}=2\bar{2}01$. Note the disappearance of dislocation segments (indicated by *arrows* on **(a)**) on **(b)** and the weak contrast on **(c)**. The Burgers vector of these dislocations is determined to be $\frac{1}{3}[11\bar{2}0]$ and, from the line directions, it is determined that these are screw dislocations

of TEM images of dislocations. An example of dislocation analysis in a single crystal of zinc oxide (ZnO) is presented in Fig. 42.18. These dislocations were likely formed by deformation process during postgrowth cooling. These dislocations lie in the basal plane of the 2H crystal structure of ZnO and are visible in Fig. 42.18a ($\boldsymbol{g} = 11\bar{2}0$) and invisible on the $1\bar{1}00$ (Fig. 42.18b) and the $2\bar{2}01$ (Fig. 42.18c) reflections, although weak contrast is observed. Application of the $\boldsymbol{g}\cdot\boldsymbol{b} = 0$ criterion to possible $\frac{1}{3}\langle 11\bar{2}0\rangle$ Burgers vectors which lie in the basal plane shows that the dislocations have Burgers vector $\frac{1}{3}[11\bar{2}0]$.

Determination of Burgers Vector Sense and Magnitude

The determination of the actual sense and magnitude of the Burgers vector of a dislocation, once the Burgers vector direction has been determined by $\boldsymbol{g}\cdot\boldsymbol{b}$ analysis, requires more detailed analysis. *Chikawa* [42.48] developed conventional-radiation divergent-beam techniques for the determination of the sense of both edge and screw dislocations through measurement of the sense of tilt of lattice planes surrounding the dislocation core. For the case of screw dislocations, *Mardix* et al. [42.49] subsequently further developed Chikawa's divergent-beam technique to enable the magnitude of this tilt to be measured as a function of distance from the core. This could then be fitted to the corresponding theoretical expression, enabling the Burgers vector magnitude to be determined. *Si* et al. [42.50] developed two methods, one section and one projection, which are analogous to those of *Chikawa* [42.48], and *Mardix* et al. [42.49], but which make use of synchrotron white radiation. In these techniques the lattice tilt surrounding dislocations of mainly screw character was measured with an accuracy that was significantly improved due to the relaxation of the requirement for short specimen–film distances which is inevitable in conventional radiation techniques. *Chen*, *Dudley*, and co-workers [42.51, 52] have similarly determined the dislocation sense of micropipes and closed-core threading screw dislocations in 4H-SiC by comparing grazing-incidence synchrotron white-beam x-ray topographic images of the dislocations with corresponding images simulated by the ray-tracing method.

42.6.2 Contrast from Inclusions

Individual point defects are not visible on x-ray topographs, but when such defects cluster to form a precipitate or inclusion, contrast can be observed. Under low absorption conditions, *direct* or kinematical images of precipitates are formed on x-ray topographs and these typically consist of two dark half-circles separated by a line of no contrast perpendicular to the projection of the diffracted vector. This is simply due to the fact that distortions parallel to a given set of atomic planes are not discernible. An example is shown in Fig. 42.19, which is a transmission synchrotron white beam x-ray topography (SWBXT) image recorded from an AlN crystal. Under higher absorption conditions, dynamical contrast can be observed. When the precipitate is close to the x-ray exit surface, opposite contrast either side of the defect can usually be observed. This contrast will usually reverse with reversal of the sign of the reflection vector. The contrast is produced by tie point migration in the region above the defect. Since the defect is close to the exit surface, the reflecting planes rotate very sharply in order to meet the surface at the preferred angle, and so the curvature becomes too large for the eikonal theory to handle. Consequently, the contrast developed above the defect is *frozen-in*. The black–white contrast not only reverses with the sense of the reflection vector but also with the sense of the strain in the lattice. This can be used to determine the nature of the precipitate. If the contrast is enhanced on the side of positive $\boldsymbol{g}$, then the lattice is under compression, and if reduced it is under tension. This empirical rule was first determined by *Meieran* and *Blech* [42.53].

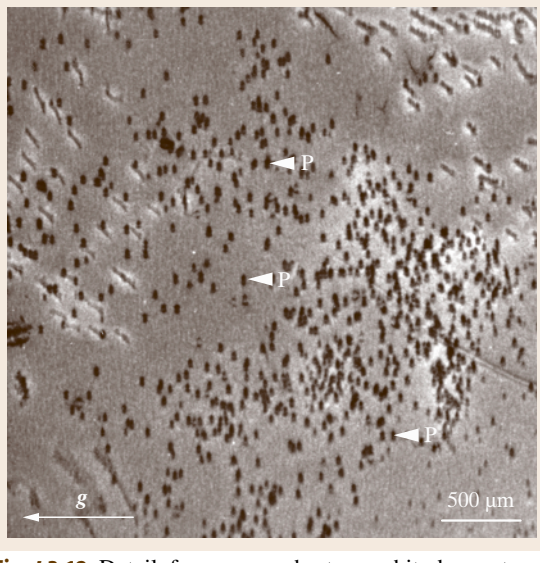
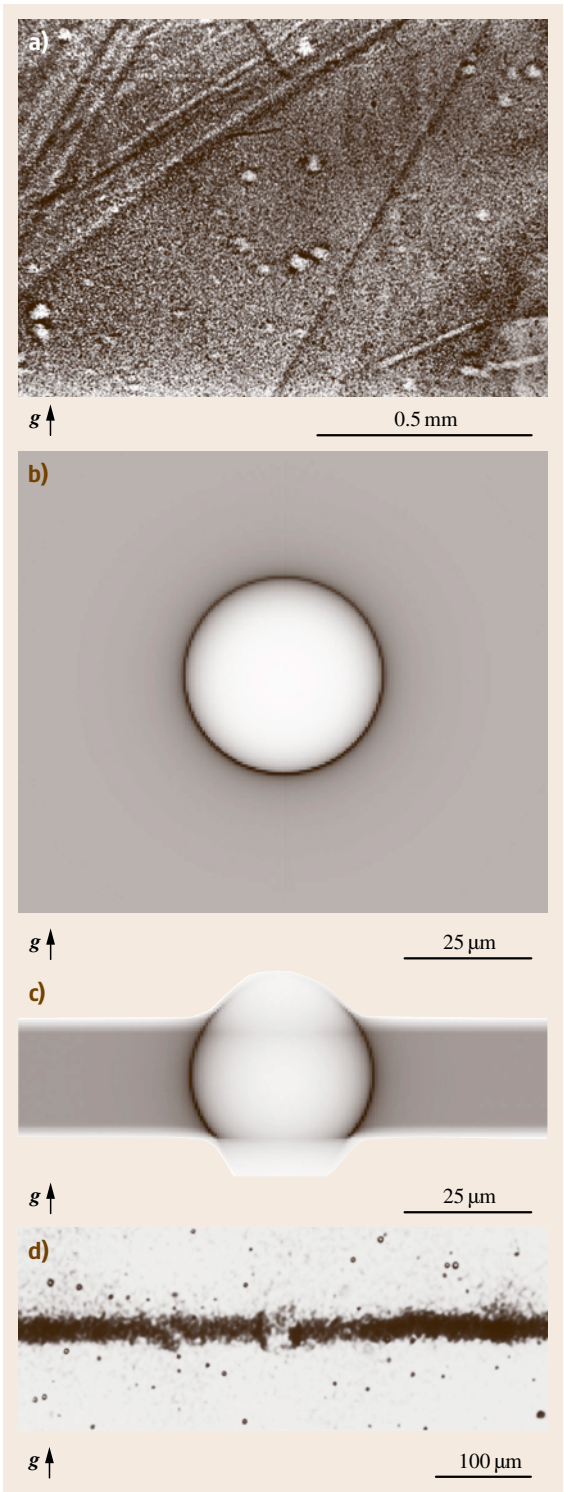


Fig. 42.19 Detail from a synchrotron white-beam topograph recorded in transmission from a AlN single crystal showing precipitate contrast (P) under low absorption conditions



Precipitates/inclusions may also be imaged using the back-reflection geometry, as shown in Fig. 42.20a for a 3C-SiC platelet containing inclusions near the surface [42.54]. The circular spots of inclusions are similar to the images of superscrew (micropipes) in 4H-SiC (Fig. 42.16) except the screw dislocation images have more distinct white contrast than the inclusion images, there being some diffracted intensity inside the inclusion spots making them appear gray rather than white.

Images of inclusions in back-reflection topographs can be understood by considering ray-tracing computer simulations based on the displacement field of a spherical inclusion. The ray-tracing concept is the same as the orientation contrast model (see Sect. 42.5.2). Figure 42.20b shows a simulated image of an inclusion with a diameter comparable to the white spots in Fig. 42.20a. The simulation a 3*c* axial screw dislocation in Fig. 42.16b and that of an inclusion in Fig. 42.20b are similar, in that intensity reflected from the area centered on their defects is concentrated into a dark ring. This effect is complete for the case of the screw dislocation of Fig. 42.16b with intensity being absent from the ring, whereas for the case of the inclusion the intensity within its dark ring is merely depleted.

The difference between the mechanisms of image formation in the back-reflection topography of axial screw dislocations and inclusions can be analyzed more intimately with section topographs. Simulated section topographs may be obtained by utilizing only the diffracting elements of the crystal that fall within the projected path of a narrow strip of incident rays. For a screw dislocation (Fig. 42.15c), on the section topograph, the strip of diffracted intensity is displaced downward and upward at either side of the white circular center of the simulated screw dislocation image to break into two tails. The upper tail originates from the left half of the illuminated dislocation core, while the lower tail corresponds to the right half. The presence of the tails stretching in directions out of their illuminated sides indicates that the diffracted rays are

Fig. 42.20 **(a)** Synchrotron white-beam back-reflection x-ray topographs from a 3C-SiC platelet ($\boldsymbol{g} = 121212$, $\lambda = 1.24$ Å) showing inclusions; **(b)** computer-simulated back-reflection white-beam x-ray topograph ($\theta_B = 80°$) of an inclusion 14 µm beneath the surface of a 3C-SiC sample; **(c)** computer-simulated white-beam back-reflection x-ray section topograph of an inclusion; **(d)** recorded synchrotron white-beam back-reflection x-ray section topograph of an inclusion 14 µm beneath the surface ($\boldsymbol{g} = 12\,12\,12$, $\lambda = 1.24$ Å) ◀

twisted from the incident beam direction. The twisting direction of the rays is opposite to the dislocation sense. Back-reflection section topography is then capable of discerning the senses of axial screw dislocations. The section topograph of an inclusion (Fig. 42.20d) lacks this two-tailedness.

Like the traverse topograph of Fig. 42.20a, the defect image again shows a gray center of depleted intensity bracketed by a dark perimeter. The diffracting planes in the vicinity of a spherical inclusion lack the helical twist that those in the core of a screw dislocation have. A spherical inclusion's strain field bulges the diffracting planes above it into convex curvatures, tilting them radially outward from the defect's center. In the simulated section topograph of an inclusion in Fig. 42.20c, the absence of twist is evident as the image is symmetrical about a line drawn through the center of the defect parallel to the $\boldsymbol{g}$ vector. Diffracted intensity is depleted from the central area of the defect image and concentrated at its edges, forming two curved bars of dark contrast. This is roughly what is seen in the section topograph of an inclusion in Fig. 42.20b.

42.6.3 Contrast Associated with Cracks

X-ray topography is also capable of discerning the deformation fields associated with cracks in single crystals. *George* et al. [42.55] have used x-ray topography to measure the extent of the plastic zone associated with the stress concentrations that can arise at the tip of a crack. *Raghothamachar* et al. [42.56] reported on deformation initiated at cracks in AlN crystals. This was achieved both in static mode using Lang topography [42.55] and synchrotron white-beam x-ray topography [42.56] and in dynamic mode using monochromatic synchrotron-radiation topography [42.57]. In a similar way, the elastic deformation field associated with crack tips can also be characterized. For example, for the case of *mud cracks*, which are often observed in GaN films grown on either SiC or sapphire substrates, the elastic field of the cracks can penetrate down into the substrate and be observable by x-ray topography. *Itoh* et al. [42.58] published an early report of such topographic observations in GaN films grown on sapphire. The presence of these mud cracks is evidence for the existence of tensile stresses in the films. For the case of SiC substrates, these tensile stresses may arise from thermal expansion mismatch, although this cannot explain their existence for the case of sapphire substrates. For the latter case, the tensile stresses are believed to be associated with grain coalescence [42.59]. TEM observation of the interface regions in the GaN/sapphire system indicates the absence of plastic deformation associated with the crack tips, rather just the presence of an elastic deformation field. This elastic deformation field (for both SiC and sapphire cases) behaves with respect to x-ray topographic observation like an edge dislocation, i. e., one can define a displacement vec-

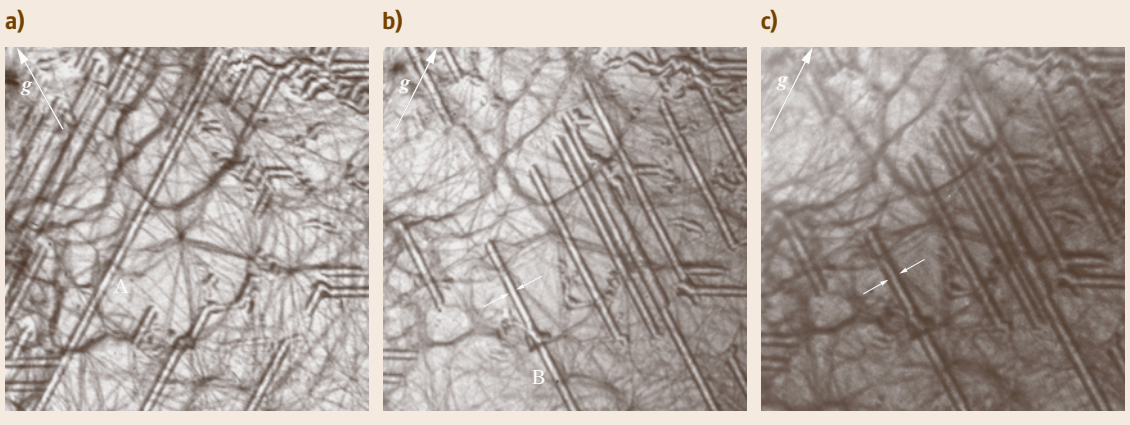


Fig. 42.21a–c SWBXT images recorded from a SiC substrate with GaN epilayers. **(a)** $\boldsymbol{g} = \bar{2}110$, $d = 10$ cm; **(b)** $\boldsymbol{g} = \bar{1}2\bar{1}0$, $d = 10$ cm; **(c)** $\boldsymbol{g} = \bar{1}2\bar{1}0$, $d = 24.7$ cm. The pairs of *dark parallel lines* arise from the elastic deformation associated with *mud cracks* in the GaN film which penetrate into the SiC substrate (intensity from the GaN film is too low to be visible on this image). Crack images A are visible in **(a)** but invisible in **(b)**, while the reverse is true for crack images B. This demonstrates the applicability of the $\boldsymbol{g}\cdot\boldsymbol{u} = 0$ and $\boldsymbol{g}\cdot\boldsymbol{u}\times\boldsymbol{l} = 0$ criteria. The *arrows* on **(b)** and **(c)** highlight the increase in the separation between the two parallel components of the crack images as the specimen–film distance is increased

tor ($\boldsymbol{u}$) associated with the crack which is oriented in the plane of the interface and normal to the long axis ($\boldsymbol{l}$) of the crack. This means that the crack image disappears when the conditions $\boldsymbol{g}\cdot\boldsymbol{u}=0$ and $\boldsymbol{g}\cdot\boldsymbol{u}\times\boldsymbol{l}=0$ are simultaneously satisfied. Figure 42.21 shows topographs of a cracked GaN/SiC sample recorded under various diffraction conditions: $\boldsymbol{g}=\bar{2}110$ and $d=10\,\mathrm{cm}$, where d is the sample–film distance (Fig. 42.21a), $\boldsymbol{g}=\bar{1}2\bar{1}0$ and $d=10\,\mathrm{cm}$ (Fig. 42.21b), and $\boldsymbol{g}=\bar{1}2\bar{1}0$ and $d=24.7\,\mathrm{cm}$ (Fig. 42.21c). All the topographs were taken with the GaN film on the x-ray exit surface. From these figures we can see that all the cracks running along $[\bar{2}110]$ which have $\boldsymbol{u}$ along $[01\bar{1}0]$ are invisible in Fig. 42.21a while all the cracks along $[\bar{1}2\bar{1}0]$ which have $\boldsymbol{u}$ along $[\bar{1}010]$ are invisible in Fig. 42.21b, due to the aforementioned invisibility criteria. Ray tracing, carried out by recording SWBXT images of the crack deformation fields at different specimen–film distances, enables determination of the *sense* of the lattice rotation (i. e., on which side of the interface the *extra half-plane* would lie) as well as the exact location of the misoriented volume with respect to the core of the crack. This is demonstrated in Fig. 42.21b,c. The separation between the parallel components of the images in Fig. 42.21c is larger than that in Fig. 42.21b, indicating that the cracks open up towards the GaN film surface. Further quantitative ray-tracing analysis reveals that the misoriented volumes are located at a distance of 30 μm from the crack core and that the lattice rotations are of the order of 20 arcsec.

42.7 Current Application Status and Development

Photographic films continue to be the detectors of choice for recording x-ray topographs, although large-area charge-coupled device (CCD) detectors can be used when high resolution is not paramount. Currently, holographic films are capable of recording x-ray topographs with submicrometer resolution but require long exposure times even for an intense synchrotron beam. Nuclear emulsions (Ilford plates) which have a grain size below 1 μm are also suitable to record high-resolution x-ray topographs but they are expensive and require special handling and developing procedures. Single-side-coated high-resolution x-ray films (Agfa D3-SC, Kodak SR-45, Fuji IX20, etc.) which have a grain size on the order of 1 μm are usually adequate for recording most x-ray topographs. These have sufficient contrast and resolution and can be developed fast enough to allow real-time observation and feedback for recording multiple topographs. CCD detectors are used instead of films when quick imaging is needed in some dynamic studies; however, the results are somewhat limited due to the poor resolution (25–40 μm).

As an efficient technique to explore the microstructure of various materials, x-ray topography has been extensively applied in both the research and industrial fields. Through observation and analysis of structural features and defects, it plays a key role in the evaluation of materials quality, the investigation of defect formation mechanisms, and the processing–microstructure–properties correlation. For the development of the crystal growth process of any material, x-ray topography is a critical tool that goes hand in hand with the crystal growth technique employed. X-ray topography has also been used to characterize thin films and heterostructures. While conventional x-ray topography can be used as a laboratory tool for quick evaluation of materials, synchrotron-based techniques have the capability for a wide range of experimental modifications, including in situ studies of crystal growth, phase transformations, device performance, etc. Further advances in digitization and processing of images can allow for rapid and automated analysis of x-ray topographs.

Modifications to the x-ray topography technique have also been developed in recent years. One example is the technique of x-ray reticulography [42.60]. Based on this technique, a novel stress measurement procedure has been developed that is capable of noninvasively measuring the complete stress tensor of any single-crystal material. This technique can be applied to a wide range of materials, such as those used in the semiconductor industry for developing or improving the packaging processes for the ever-reducing dimensions of microprocessors, or to measure stresses in as-grown single-crystal materials for comparison with performance and simulations.

References

42.1 H.J. Scheel: The development of crystal growth technology. In: *Crystal Growth Technology*, ed. by H.J. Scheel, T. Fukuda (Wiley, New York 2003) pp. 1–14

42.2 C. Claeys, L. Deferm: Trends and challenges for advanced silicon technologies, Solid State Phenom. **47–48**, 1–16 (1993)

42.3 D.T.J. Hurle: *Handbook of Crystal Growth: Thin Films and Epitaxy* (Elsevier, New York 1994)

42.4 G. Dhanaraj, T. Shripathy, H.L. Bhat: Growth and defect characterization of L-arginine phosphate monohydrate, J. Cryst. Growth **113**, 456–464 (1991)

42.5 S. Kawado, S. Lida, Y. Yamaguchi, S. Kimura, Y. Hirose, K. Kajiwara, Y. Chikaura, M. Umeno: Synchrotron-radiation x-ray topography of surface strain in large-diameter silicon wafers, J. Synchrotron. Radiat. **9**, 166–168 (2002)

42.6 B. Raghothamachar, H. Chen, M. Dudley: Unpublished results (1998)

42.7 M. Dudley, X. Huang: X-ray topography. In: *Microprobe Characterization of Optoelectronic Materials*, ed. by J. Jimenez (Gordon Breach/Harwood Academic, Amsterdam 2003) pp. 531–594

42.8 G. Dhanaraj, X.R. Huang, M. Dudley, V. Prasad, R.H. Ma: Silicon carbide crystals – Part I: Crystal growth and characterization. In: *Crystal Growth Technology*, ed. by K. Byrappa, T. Ohachi (William Andrew/Springer, New York 2003) pp. 181–232

42.9 G.N. Ramachandran: X-ray topographs from diamond, Proc. Indian Acad. Sci. A **19**, 280–294 (1944)

42.10 A. Authier: Contrast of images in x-ray topography. In: *Diffraction and Imaging Techniques in Materials Science*, ed. by S. Amelinckx, R. Gevers, J. van Landuyt (North Holland, Amsterdam 1978) pp. 715–757

42.11 J.E. White: X-ray diffraction by elastically deformed crystals, J. Appl. Phys. **21**, 855–859 (1950)

42.12 V.W. Berg: Über eine röntgenographische Methode zur Untersuchung von Gitterstörungen an Kristallen, Naturwissenschaften **19**, 391–396 (1931), in German

42.13 A.R. Lang: Direct observation of individual dislocations by x-ray diffraction, J. Appl. Phys. **29**, 597–598 (1958)

42.14 C.S. Barrett: A new microscopy and its potentialities, AIME Transactions **161**, 15–65 (1945)

42.15 W.L. Bond, J. Andrus: Structural imperfections in quartz crystals, Am. Mineral. **37**, 622–632 (1952)

42.16 G.H. Schwuttke: New x-ray diffraction microscopy technique for study of imperfections in semiconductor crystals, J. Appl. Phys. **36**, 2712–2714 (1965)

42.17 W. Wooster, W.A. Wooster: X-ray topographs, Nature **155**, 786–787 (1945)

42.18 J.B. Newkirk: Method for the detection of dislocations in silicon by x-ray extinction contrast, Phys. Rev. **110**, 1465–1466 (1958)

42.19 L.G. Schulz: Method of using a fine-focus x-ray tube for examining the surfaces of single crystals, J. Met. **6**, 1082–1083 (1954)

42.20 A. Guinier, J. Tennevin: Sur deux variantes de la methode de luae et leurs applications, Acta Cryst. **2**, 133–138 (1949), in French

42.21 T. Tuomi, K. Naukkarinen, P. Rabe: Use of synchrotron radiation in x-ray-diffraction topography, Phys. Status Solidi (a) **25**, 93–106 (1974)

42.22 A.R. Lang: Point-by-point x-ray diffraction studies of imperfections in melt-grown crystals, Acta Cryst. **10**, 839 (1957)

42.23 U. Bonse, E. Kappler: Röntgenographische Abbildung des Verzerrungsfeldes einzelner Versetzungen in Germanium-Einkristallen, Z. Naturforsch. A **13**, 348 (1958), in German

42.24 D.K. Bowen, B.K. Tanner: *High Resolution X-ray Diffractometry and Topography* (Taylor Francis, London 1998) p. 174

42.25 J. Miltat: White beam synchrotron radiation. In: *Characterization of Crystal Growth Defects by X-ray Methods*, ed. by B.K. Tanner, D.K. Bowen (Plenum, New York 1980) pp. 401–420

42.26 P.G. Neudeck: Electrical impact of SiC structural defects on high electric field devices, Mater. Sci. Forum **338–342**, 1161–1166 (2000)

42.27 P.G. Neudeck, W. Huang, M. Dudley: Breakdown degradation associated with elementary screw dislocations in 4H-SiC p+n junction rectifiers. In: *Power Semiconductor Materials and Devices*, ed. by S.J. Pearton, R.J. Shul, E. Wolfgang, F. Ren, S. Tenconi (Materials Research Society, Warrendale 1998) pp. 285–294

42.28 M. Hart: Synchrotron radiation – Its application to high speed, high resolution x-ray diffraction topography, J. Appl. Crystallogr. **8**, 436–444 (1975)

42.29 X.R. Huang, M. Dudley, J.Y. Zhao, B. Raghothamachar: Dependence of the direct dislocation image on sample-to-film distance in x-ray topography, Philos. Trans. R. Soc. Ser. A **357**, 2659–2670 (1999)

42.30 D.R. Black, G.G. Long: X-ray Topography, Special Publication 0960-10 (National Institute of Standards and Technology (NIST), Washington 2004)

42.31 M. Dudley, S. Wang, W. Huang, C.H. Carter Jr., V.F. Tsvetkov, C. Fazi: White beam synchrotron topographic studies of defects in 6H-SiC single crystals, J. Phys. D Appl. Phys. **28**, A63–A68 (1995)

42.32 M. Dudley, J. Wu, G.-D. Yao: Determination of penetration depths and analysis of strains in single crystals by white beam synchrotron x-ray topography in grazing Bragg–Laue geometries, Nucl. Instrum. Methods B **40/41**, 388–392 (1989)

42.33 N. Kato: The flow of x-rays and materials waves in ideally perfect single crystals, Acta Crystallogr. **11**, 885–887 (1958)

42.34 N. Kato: The energy flow of x-rays in an ideally perfect crystal: comparison between theory and experiments, Acta Crystallogr. **13**, 349–356 (1960)

42.35 G. Borrmann: The extinction diagram of quartz, Phys. Z. **42**, 157–162 (1941)

42.36 G. Borrmann: Absorption of Röntgen rays in the case of interference, Phys. Z. **127**, 297–323 (1950)

42.37 B.K. Tanner: *X-ray Diffraction Topography* (Pergamon, Oxford 1976)

42.38 B.K. Tanner: Contrast of defects in x-ray diffraction topographs. In: *X-ray and Neutron Dynamical Diffraction: Theory and Applications*, ed. by A. Authier, S. Lagomarsino, B.K. Tanner (Plenum, New York 1996) pp. 147–166

42.39 J.E.A. Miltat, D.K. Bowen: On the widths of dislocation images in x-ray topography under low-absorption conditions, J. Appl. Cryst. **8**, 657–669 (1975)

42.40 X.R. Huang, M. Dudley, W.M. Vetter, W. Huang, C.H. Carter Jr.: Contrast mechanism in superscrew dislocation images on synchrotron back reflection topographs. In: *Applications of Synchrotron Radiation Techniques to Materials Science IV*, ed. by S.M. Mini, D.L. Perry, S.R. Stock, L.J. Terminello (Materials Research Society, Warrendale 1998) pp. 71–76

42.41 X.R. Huang, M. Dudley, W.M. Vetter, W. Huang, W. Si, C.H. Carter Jr.: Superscrew dislocation contrast on synchrotron white-beam topographs: An accurate description of the direct dislocation image, J. Appl. Cryst. **32**, 516–524 (1999)

42.42 X.R. Huang, M. Dudley, W.M. Vetter, W. Huang, S. Wang, C.H. Carter Jr.: Direct evidence of micropipe-related pure superscrew dislocations in SiC, Appl. Phys. Lett. **74**, 353–355 (1999)

42.43 B.W. Batterman, H. Cole: Dynamical diffraction of x-rays by perfect crystals, Rev. Mod. Phys. **36**, 681–717 (1964)

42.44 P. Penning, D. Polder: Anomalous transmission of x-rays in elastically deformed crystals, Philips Res. Rep. **16**, 419–440 (1961)

42.45 H. Klapper: Defects in non-metal crystals. In: *Characterization of Crystal Growth Defects by X-ray Methods*, ed. by B.K. Tanner, D.K. Bowen (Plenum, New York 1980) pp. 133–160

42.46 D. Yuan, M. Dudley: Dislocation line direction determination in pyrene single crystals, Mol. Cryst. Liq. Cryst. **211**, 51–58 (1992)

42.47 J. Miltat, M. Dudley: Projective properties of Laue topographs, J. Appl. Cryst. **13**, 555–562 (1980)

42.48 J.I. Chikawa: X-ray topographic observation of dislocation contrast in thin CdS crystals, J. Appl. Phys. **36**, 3496–3502 (1965)

42.49 S. Mardix, A.R. Lang, I. Blech: On giant screw dislocations in ZnS polytype crystals, Philos. Mag. **24**, 683–693 (1971)

42.50 W. Si, M. Dudley, C.H. Carter Jr., R. Glass, V.F. Tsvetkov: Determination of Burgers vectors of screw dislocations in 6H-SiC single crystals by synchrotron white beam x-ray topography. In: *Applications of Synchrotron Radiation to Materials Science*, ed. by L. Terminello, S. Mini, D.L. Perry, H. Ade (Materials Research Society, Warrendale 1996) pp. 129–134

42.51 Y. Chen, M. Dudley: Direct determination of dislocation sense of closed-core threading screw dislocations using synchrotron white beam x-ray topography in 4H silicon carbide, Appl. Phys. Lett. **91**, 141918 (2007)

42.52 Y. Chen, G. Dhanaraj, M. Dudley, E.K. Sanchez, M.F. MacMillan: sense determination of micropipes via grazing-incidence synchrotron white beam x-ray topography in 4H silicon carbide, Appl. Phys. Lett. **91**, 071917 (2007)

42.53 E.S. Meieran, I.A. Blech: X-ray extinction contrast topography of silicon strained by thin surface films, J. Appl. Phys. **36**, 3162 (1965)

42.54 W.M. Vetter, M. Dudley: The contrast of inclusions compared with that of micropipes in back-reflection synchrotron white-beam topographs of SiC, J. Appl. Cryst. **37**, 200–203 (2004)

42.55 G. Michot, K. Badawi, A.R. Halim, A. George: X-ray topographic study of crack-tip dislocation patterns in silicon, Philos. Mag. A **42**, 195–214 (1980)

42.56 B. Raghothamachar, M. Dudley, J.C. Rojo, K. Morgan, L.J. Schowalter: X-ray characterization of bulk AlN single crystals grown by the sublimation technique, J. Cryst. Growth **250**, 244–250 (2003)

42.57 G. Michot, A. George: In situ observation by x-ray synchrotron topography of the growth of plastically deformed regions around crack tips in silicon under creep conditions, Scr. Metall. **16**, 519–524 (1982)

42.58 N. Itoh, J.C. Rhee: Study of cracking mechanism in GaN/α-Al_2O_3 structure, J. Appl. Phys. **58**, 1828–1837 (1985)

42.59 E.V. Etzkorn, D.R. Clarke: Cracking of GaN films, J. Appl. Phys. **89**, 1025–1034 (2001)

42.60 A.R. Lang, A.P.W. Makepeace: Synchrotron x-ray reticulography: principles and applications, J. Phys. D Appl. Phys. A **32**, 97–103 (1999)

43. Defect-Selective Etching of Semiconductors

Jan L. Weyher, John J. Kelly

In the present chapter we first briefly consider mechanisms for the etching of semiconductors (Sect. 43.1) and relate these principles to methods for controlling surface morphology and revealing defects (Sect. 43.2). Section 43.3 describes in some detail defect-sensitive etching methods. Results are presented for the classical (orthodox) method used for revealing dislocations in Sect. 43.3.1. More recently developed open-circuit (photo)etching approaches, sensitive to both crystallographic and electrically active inhomogeneities in semiconductors, are reviewed in Sect. 43.3.2. In particular, attention will focus on newly introduced etchants and etching procedures for wide-bandgap semiconductors.

43.1 **Wet Etching of Semiconductors: Mechanisms** 1454
43.1.1 Chemical Etching 1454
43.1.2 Electrochemical Etching 1454
43.1.3 Electroless Etching 1456
43.1.4 Photogalvanic Etching 1458
43.2 **Wet Etching of Semiconductors: Morphology and Defect Selectivity** 1459
43.2.1 Chemical Etching 1459
43.2.2 Electrochemical Etching 1459
43.2.3 Electroless Etching 1460
43.2.4 Photogalvanic Etching 1461
43.3 **Defect-Selective Etching Methods** 1461
43.3.1 Orthodox Etching for Revealing Dislocations 1461
43.3.2 Electroless Etching for Revealing Defects 1469
References 1473

Wet etching processes are widely used in the fabrication of semiconductors. There are three main application fields for this technique: device pattern formation, polishing, and visualization of damage or defects. This chapter is restricted to the description of defect-selective etching (DSE). Among the methods used for revealing and analyzing defects in semiconductors, such as x-ray topography and diffraction, transmission electron microscopy (TEM), cathodoluminescence (CL), electron-beam-induced current (EBIC), and laser scattering tomography (LST), etching has several favorable features that make the technique attractive for assessment of the quality of single crystals and device structures. These are: simple and relatively low-cost equipment, no limits to the size of samples to be examined, very quick data acquisition, and the broad range of defects that can be revealed and analyzed.

The rapid development of various etching systems and methods for different semiconductors began after the discovery in the 1950s of a direct correspondence between the outcrops of dislocations and etch pits. Numerous review papers appeared in regular journals and books summarizing the current developments in this field [43.1–12]. Although for elemental and *classical* compound semiconductors different etching systems are well known, challenging problems arise due to the technological developments of the new generation of wide-bandgap compound semiconductors (group III nitrides and SiC). These materials are characterized by very high chemical resistance to the majority of the known acid-based etching systems, show strong polarity-related anisotropy of properties, and contain a very high density and diversity of defects, not occurring in classical compound semiconductors (e.g., inversion domains, macro- and nanopipes, pinholes, and extended stacking faults). New etching systems and more sophisticated etching methods as well as their calibration are therefore essential.

43.1 Wet Etching of Semiconductors: Mechanisms

For etching of semiconductors in solution (or in a melt) one can distinguish two mechanisms [43.7, 13]. In *chemical etching* valence electrons are exchanged *locally* between surface bonds of the semiconductor and an active agent in the etchant. When all back bonds are broken, the surface atom passes into solution. The other type of mechanism involves *free* charge carriers (valence-band holes or conduction-band electrons) which, when localized at the surface, cause rupture of surface bonds. If an electrochemical cell, with a voltage source and a counterelectrode, is used to supply charge carriers to the semiconductor, we speak of *electrochemical etching*. Both electrons and holes may cause decomposition of the solid. Alternatively, free charge carriers can be supplied from a chemical species in solution under open-circuit conditions: this is referred to as *electroless etching*. In addition, if minority carriers are required for etching, the semiconductor may be illuminated. In this case we can distinguish *photoelectrochemical etching* and *electroless photoetching*. Finally, there is a hybrid form of photoetching that requires a counterelectrode but no voltage source. This *photogalvanic etching* method has been used extensively for wide-bandgap semiconductors.

In this section we first briefly review the various etching mechanisms. This forms the basis for a consideration of defect-selective etching in Sect. 43.2, a theme that will be further developed in Sect. 43.3.

43.1.1 Chemical Etching

A typical example of a chemical etching reaction is that of InP in concentrated HCl solution [43.14]. By electron exchange In-P and H-Cl bonds are broken and $InCl_3$ and PH_3 are formed as products

$$InP + 3HCl \rightarrow InCl_3 + PH_3 \,. \qquad (43.1)$$

The products are subsequently hydrolyzed in the aqueous solution. Etching of InP in HBr follows a similar mechanism [43.7]. The chemical etching of silicon in alkaline solution is another important reaction [43.15, 16]. These are complex, multistep processes. A general feature is that their rate is controlled by surface kinetics and, as a consequence, etching is strongly anisotropic. For example, the etch rate of the Si (111) face in KOH solution and of the (111) In face of InP in concentrated HCl solution is much lower than that of other crystallographic faces. Unlike electrochemical or electroless etching, the kinetics of chemical etching is not dependent on the electrochemical potential of the sample.

For revealing defects in many semiconductors (Si, SiC, GaN) a NaOH/KOH eutectic is often used [43.12]. Despite the long and widespread application of this system, surprisingly little is known about the etching mechanism. In all these cases the semiconductor is oxidized, i. e., loses valence electrons, during etching. For example, GaN will be converted to Ga^{3+} ($Ga(OH)_3$ or Ga_2O_3) in the melt. OH^- ions cannot act as a reducing agent, i. e., electron acceptor. We suggest that trace water in the eutectic is responsible for bond rupture and enhanced etching at defects. For example, in the case of GaN a possible reaction could be

$$GaN + 3H_2O \rightarrow Ga(OH)_3 + NH_3 \,. \qquad (43.2)$$

Water, strongly bound in the solvation shell of the cations and anions, is available for bond rupture. The product $Ga(OH)_3$ will dissolve in the melt.

43.1.2 Electrochemical Etching

For the kinetics of electrochemical etching one has to distinguish between reactions based on majority and minority carriers [43.7, 13]. In this section we first consider majority-carrier (holes in a p-type or electrons in an n-type semiconductor) and then minority-carrier reactions. To guide the reader we show in Fig. 43.1 a simple electrochemical cell with three electrodes: the semiconductor working electrode (WE), a platinum counterelectrode (CE), and a reference electrode (RE). With the aid of a voltage source (e.g., a potentiostat) the electrochemical potential V of the semiconductor can be varied with respect to that of the reference electrode. As a result, a current flows between the working electrode and the counterelectrode. This current j is measured as a function of V. The semiconductor can be illuminated to give photocurrent.

Majority-Carrier Reactions

If, in a p-type semiconductor, holes are localized at the surface (i. e., the bonding electrons are removed) the surface bonds will be broken. In the case of GaN, three holes are required to dissolve one formula unit of the solid in acidic solution [43.17].

$$GaN + 3h^+ \rightarrow Ga^{3+} + \tfrac{1}{2}N_2 \,. \qquad (43.3)$$

The solid is *oxidized*; the Ga^{3+} ions dissolve and molecular nitrogen escapes to the gas phase. From (43.3)

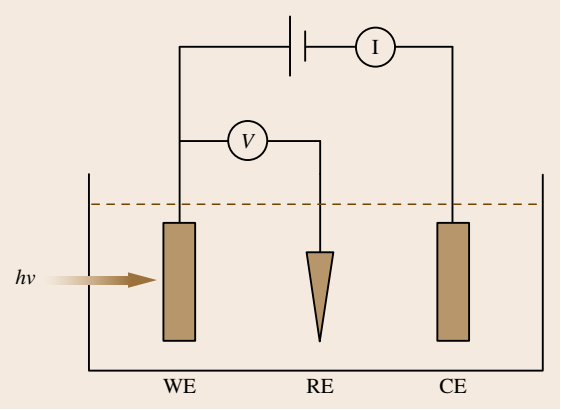


Fig. 43.1 A schematic view of a three-electrode electrochemical cell. The potential V of the semiconductor working electrode (WE) is regulated with respect to that of the reference electrode (RE) using a voltage source. As a result, a current j can flow between the working electrode and a counterelectrode (CE). In a typical experiment j is measured as a function of V either in the dark or under illumination. If electrons flow from WE to CE the current is *anodic* (j_a); an oxidation reaction occurs at the semiconductor. If electrons flow in the opposite direction (from CE to WE) the current is *cathodic* (j_c) and a reduction reaction occurs at the semiconductor

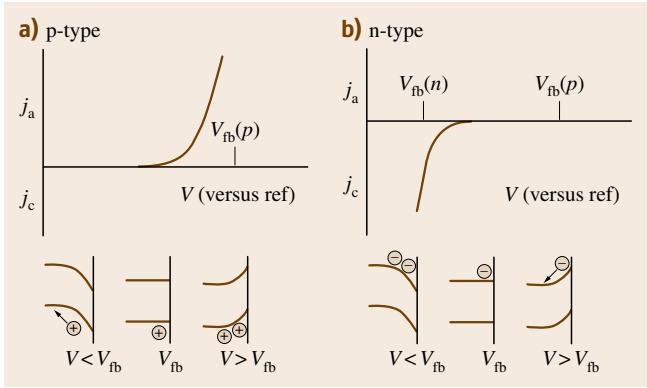


Fig. 43.2a,b Current–potential characteristics for majority-carrier reactions at **(a)** a p-type and **(b)** an n-type semiconductor. In case **(a)** accumulation of valence-band holes at the surface leads to oxidation of the electrode and an anodic current in the external circuit. The flat-band potential ($V_{fb}(p)$) is denoted, as well as energy-band schemes of the semiconductor at the interface with solution. In case **(b)** accumulation of electrons at the surface causes the reduction of H^+ ions or H_2O to hydrogen or, in some cases, the reduction of the semiconductor. Note: the difference in flat-band potential for n- and p-type forms of the same semiconductor corresponds to the difference in Fermi levels of the two types

it follows that the rate of the reaction should depend on the surface hole concentration p_s, which is related to the hole concentration in the bulk p_b by

$$p_s = p_b \exp\left(\frac{eV_{sc}}{k_B T}\right) , \qquad (43.4)$$

where eV_{sc} is the band bending in the semiconductor at the interface with the solution. In an electrochemical experiment the band bending, and thus the reaction rate, can be varied by means of the applied potential V. A change in the applied potential can lead to a potential drop across the space-charge layer of the semiconductor (Fig. 43.2) and across the double layer (Helmholtz layer) in solution. For simplicity, we neglect the latter; this will not markedly affect the argument in this section [43.7, 13]. In that case the band bending is given by

$$eV_{sc} = e(V - V_{fb}) , \qquad (43.5)$$

where V_{fb} is the flat-band potential. At this potential, an important reference point in semiconductor electrochemistry, there is no space-charge layer or electric field present at the semiconductor surface (see the center picture at the bottom of Fig. 43.2a). At potentials considerably negative with respect to the flat-band potential of a p-type semiconductor ($V < V_{fb}(p)$) the surface hole concentration is very low ($p_s \ll p_b$): this corresponds to depletion. Oxidation reactions requiring holes cannot take place and no current flows in the external circuit. At potentials approaching $V_{fb}(p)$ an *anodic* current is measured which increases exponentially with increasing potential (Fig. 43.2a). For almost all semiconductors this process involves oxidation and dissolution of the solid: anodic etching (as in (43.3)). A reduction reaction at the counterelectrode (e.g., hydrogen evolution) provides the charge carriers required to oxidize the semiconductor. The anodic current density can be converted to an etch rate by applying Faraday's law [43.7]. It is clear that the etch rate can be simply regulated via the applied potential.

Because of the absence of holes in the valence band, anodic oxidation is not possible with an n-type semiconductor (in the dark). Instead reactions involving electrons via the conduction band can be expected. In aqueous solution this is generally the reduction of protons at low pH [43.18]

$$2H^+ + 2e^- \rightarrow H_2 , \qquad (43.6a)$$

or of water at higher pH

$$2H_2O + 2e^- \rightarrow H_2 + 2OH^- , \qquad (43.6b)$$

giving hydrogen evolution. Since the surface electron concentration increases exponentially with decreasing potential, i.e., $n_s = n_b \exp(-eV_{sc}/k_BT)$, the rate of the cathodic current associated with these reactions also increases markedly (Fig. 43.2b). (In this case oxidation of water at the counterelectrode provides the electrons required at the semiconductor working electrode.) For a few semiconductors the presence of electrons at the surface can give rise to decomposition of the solid. For example, at negative potential the n-type semiconductor ZnO is electrochemically reduced

$$ZnO + 2H^+ + 2e^- \rightarrow Zn + H_2O , \qquad (43.7)$$

a reaction that competes with hydrogen evolution [43.18]. Another example is n-type InP [43.19]

$$InP + 3H^+ + 3e^- \rightarrow In + PH_3 . \qquad (43.8)$$

Zinc and indium are reactive metals that dissolve readily in acidic and alkaline solution.

Minority-Carrier Reactions

Reactions requiring minority carriers (electrons in a p-type or holes in an n-type semiconductor) can be observed when the electrode is illuminated with suprabandgap light. What is important for etching is the effective spatial separation of the photogenerated electrons and holes (the prevention of recombination) allowing the minority carriers to react at the surface.

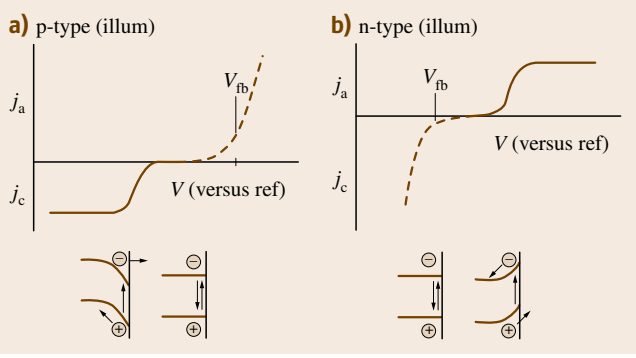


Fig. 43.3a,b The influence of suprabandgap illumination (illum) on the current–potential characteristics of **(a)** a p-type and **(b)** an n-type semiconductor. In case **(a)** the electric field of the space-charge layer drives the photogenerated minority carriers (electrons) to the surface, where a reduction reaction takes place. The holes are registered as a cathodic photocurrent in the external circuit. At positive potentials, electrons and holes recombine. In case **(b)** holes migrate to the surface, causing oxidation of the semiconductor: an anodic photocurrent is measured. In both cases the majority-carrier processes (Fig. 43.2) are shown as *dashed lines*

In a p-type semiconductor electrons and holes generated by light are separated by the electric field of the space-charge layer at negative potentials (see inset in Fig. 43.3a). At the surface the electrons are used for the reduction of H^+ ions or H_2O (43.6a,b). In the case of p-type InP the semiconductor itself is decomposed under illumination (43.8). The photogenerated holes are swept to the back contact and registered as a cathodic photocurrent in the external circuit. The constant limiting photocurrent at negative potential generally depends on the light intensity. At more positive potentials going towards the flat-band condition the electric field at the surface decreases and electron–hole recombination competes with the electrochemical reaction of the electrons. Because of the importance of surface recombination at semiconductor electrodes in solution and the poor kinetics of reactions such as hydrogen evolution, the onset of photocurrent is generally at a potential considerably negative with respect to the flat-band value (up to 0.6 V), in contrast to a Schottky diode.

Holes created by suprabandgap light in an n-type semiconductor can give rise to photoelectrochemical oxidation of the solid (43.3). As in the case of the p-type electrode, electron–hole recombination generally dominates at potentials close to V_{fb} (Fig. 43.3b). The onset of photocurrent, anodic in this case, is at positive potentials where the electric field moves the holes to the surface and the electrons to the back contact. For both cases, n- and p-type, the etch rate depends not only on the applied potential but also on the light intensity.

43.1.3 Electroless Etching

In electrochemical etching, the charge carriers are supplied to the semiconductor–solution interface via an external circuit. The disadvantage of this approach is the need for a counterelectrode, a voltage source, and, in well-defined experiments, a reference electrode (Fig. 43.1). Etching of semiconductors under open-circuit conditions is possible, in principle, if the charge carriers can be supplied from a redox species in solution [43.13]. While electron injection from solution into the conduction band of semiconductors can occur, this generally does not lead to decomposition. On the other hand, extraction of electrons from the valence band (i. e., *injection of holes* into the band) can be used for etching many semiconductors of both p- and n-type

(see the next paragraph). There is another form of electroless etching in which minority carriers are generated by suprabandgap light. This will be dealt with in the paragraph *Electroless Photoetching*.

Electroless Etching in the Dark

The principle of electroless etching is shown in Fig. 43.4 for a p-type semiconductor, e.g., GaAs, in an alkaline solution containing an oxidizing agent (a strong electron acceptor) such as the $Fe(CN)_6^{3-}$ ion [43.20]. Two electrochemical reactions occur: the oxidation of the semiconductor which, in the case of GaAs, requires six valence-band holes

$$GaAs + 10OH^- + 6h^+ \rightarrow GaO_2^- + AsO_3^{3-} + 5H_2O\,, \tag{43.9}$$

and the reduction of the oxidizing agent by hole injection into the valence band (VB)

$$Fe(CN)_6^{3-} \rightarrow Fe(CN)_6^{4-} + h^+(VB)\,. \tag{43.10}$$

The potential dependence of the rates of these two reactions is shown in Fig. 43.4. The form of the anodic current–potential curve (a) for reaction (43.9) has been introduced before. Reduction of the oxidizing agent (43.10) starts at the equilibrium (Nernst) potential of the $Fe(CN)_6^{3-}/Fe(CN)_6^{4-}$ redox couple. The rate increases as the potential is made negative (curve b, Fig. 43.4). In the present case, the rate finally becomes diffusion controlled, i. e., the cathodic current is independent of the applied potential. In an electrochemical experiment, the measured current–potential curve (c) is the sum of the two *partial* curves a and b.

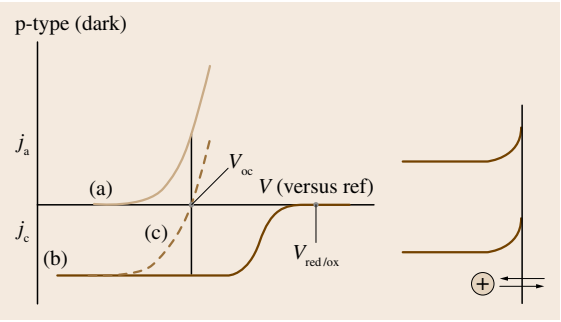


Fig. 43.4 The principle of electroless etching of a p-type semiconductor. Curve a is the anodic current–potential curve for oxidation/dissolution of the electrode. Curve b shows the cathodic current–potential curve for reduction of the oxidizing agent by hole injection into the valence band. This reaction starts at the equilibrium potential $V_{red/ox}$ and its rate becomes diffusion limited at negative potentials. The current measured with such a system, i. e., the sum of the anodic and cathodic contributions, is shown as a *dotted line*. At the open-circuit potential V_{oc} the two *partial* currents are equal ($j_a = j_c$). This is shown in the schematic diagram on the *right*. It is clear that at V_{oc} the semiconductor is etched

From Fig. 43.4 it is clear that at the open-circuit potential, indicated by V_{oc}, the two electrochemical reactions (43.9) and (43.10) occur at a significant rate; the anodic and cathodic currents are equal ($j_a = j_c$). For this system the electroless etch rate is determined by mass-transport-controlled hole injection (curve b). From this it follows that the etch rate can be enhanced by either increasing the concentration of the oxidizing agent in solution or by improving the hydrodynamics of the system.

In a similar way holes injected by $Fe(CN)_6^{3-}$ into the valence band of n-type GaAs will give rise to etching at open-circuit potential. From this discussion it is clear that, for electroless etching of n-type and p-type materials, the oxidizing agent must be capable of extracting electrons from the valence band of the solid; the electron acceptor levels of the oxidizing agent must correspond in energy to the valence band of the solid. For wide-bandgap semiconductors such as GaN and SiC with an energetically low-lying valence band this is no longer possible. Even the strongest oxidizing agents fail to inject holes under normal conditions. Consequently, electroless etching is not possible for these materials.

Electroless Photoetching

As for photoelectrochemical etching that depends on minority carriers, electroless etching may also be promoted by illumination with suprabandgap light [43.21–23]. The principle is shown schematically in Fig. 43.5a. At open-circuit potential a photon generates an electron–hole pair. An electron acceptor in solution (Ox^+) captures the electrons very effectively from the conduction band, allowing the holes to break surface bonds and thus cause etching. The electron acceptor in this case is different from that required for electroless etching in the dark: its acceptor levels must correspond to the conduction band of the solid and thus be higher in energy. For electroless photoetching the rates of reaction of electrons and holes at the surface are equal and must be higher than the rate of electron–hole recombination.

Electroless photoetching can be understood in electrochemical terms on the basis of Fig. 43.5b for an n-type semiconductor. As in Fig. 43.4, the two partial

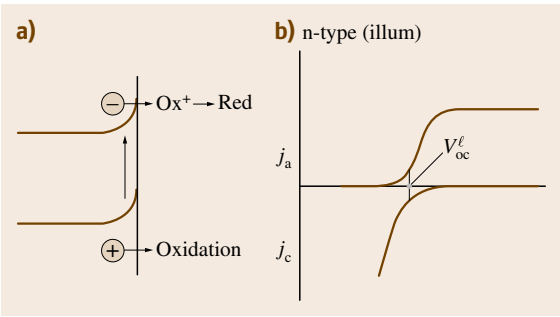


Fig. 43.5a,b The principle of electroless photoetching of an n-type semiconductor. Light generates electron–hole pairs **(a)**. The electrons are captured by an electron acceptor (Ox^+) in solution. The holes cause oxidation and dissolution of the semiconductor. These reactions compete with electron–hole recombination. **(b)** Shows the current–potential characteristics of the two processes: photoanodic oxidation of the semiconductor (by minority carriers) and cathodic reduction of the oxidizing agent (by majority carriers). At the open-circuit potential under illumination (illum) V_{oc}^{ℓ} both reactions occur at the same rate ($j_a = j_c$) and the semiconductor is photoetched

current–potential curves are shown: (a) for the photoanodic oxidation of the semiconductor via the valence band and (b) for the reduction of the electron acceptor via the conduction band (this reaction, involving majority carriers, can occur in the dark). At the open-circuit potential under illumination V_{oc}^{ℓ} both reactions occur and the semiconductor is photoetched. From this description it is clear that the systems shown in Fig. 43.3 will not give rise to photoetching. For example, the photocathodic reduction reaction at the p-type electrode (Fig. 43.3a) is only observed at negative potentials, a range in which the semiconductor cannot be oxidized (see anodic curve). In these cases the kinetics of the minority-carrier reaction is poor and electron–hole recombination dominates in the vicinity of the flat-band potential.

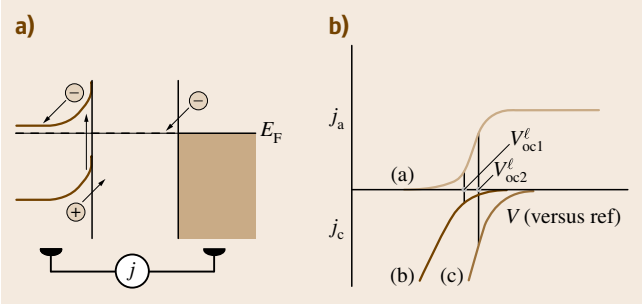


Fig. 43.6a,b The principle of photogalvanic etching. An n-type semiconductor is short-circuited to a metal in solution (without a voltage source). Electrons and holes generated by light **(a)** are separated by the electric field at the semiconductor surface. The holes cause etching. The electrons pass to the metal, where an oxidizing agent is reduced. Since the two electrodes are short-circuited, the Fermi level E_F is constant in the system. **(b)** shows the corresponding electrochemistry. Curve a gives the potential dependence of the minority-carrier reaction (photoanodic oxidation). Curve b describes reduction of the oxidizing agent (via majority carriers). The open-circuit potential is V_{oc1}^{ℓ}. If the surface area or the catalytic activity of the metal is increased (curve c) the open-circuit potential shifts to V_{oc2}^{ℓ} and the photoetch rate is significantly increased

43.1.4 Photogalvanic Etching

For the case in which the rate of the majority-carrier reaction (e.g., electron transfer in the example of Fig. 43.5) is hindered by unfavorable kinetics, photoetching under open-circuit conditions may be promoted by short-circuiting the semiconductor to a noble metal in solution [43.24, 25]. A metal can be chosen that *catalyzes* the majority-carrier reaction (this may be the reduction of oxygen, naturally present in solution, or an added oxidizing agent). The principle is shown in Fig. 43.6a. Illumination creates in the semiconductor electrons and holes that are separated by the electric field of the space-charge layer. The holes react at the surface, causing oxidation and etching. The electrons, instead of having to react at the semiconductor–solution interface, now pass to the metal via the external circuit. Reduction of the oxidizing agent from solution occurs at the metal.

The corresponding electrochemical picture is shown in Fig. 43.6b. Curve a represents the potential dependence of the current of the illuminated n-type semiconductor electrode, while curve b is the cathodic current–potential curve measured in a separate experiment at the metal. Under open-circuit conditions (V_{oc1}^{ℓ}) the two currents must be equal ($j_a = j_c$). From Fig. 43.6 it is clear that photoetching will occur in this case. The ratio of the surface areas of semiconductor and metal exposed to solution is important for the photoetch rate. An increase in the area of the metal leads to an increase in the rate of reduction of the oxidizing agent (curve c). This causes a shift in the open-circuit potential to V_{oc2}^{ℓ} ($> V_{oc1}^{\ell}$) and an increase in the photoetch rate [43.25].

The principle shown in Fig. 43.6 is that of a galvanic cell (the semiconductor is the anode, the metal the

cathode) that operates under illumination. It is, in fact, a solar cell that could supply electrical energy. This type of etching we refer to as *photogalvanic*.

Photogalvanic etching has an advantage over photoelectrochemical etching in that a voltage source is not required. However, the experiment is less well defined. The open-circuit potential, and thus the photoetch rate, depend on what is happening at the metal, i. e., its active area, (catalytic) activity etc. An Ohmic resistance in the system, e.g., a poor contact to the semiconductor, complicates the picture [43.25] and can give rise to a considerable reduction in etch rate.

43.2 Wet Etching of Semiconductors: Morphology and Defect Selectivity

In the kinetics of wet etching of solids two types of limitation of the etch rate can be distinguished:

- Surface kinetics: an interfacial chemical step is rate limiting. In this case the crystallographic orientation of the surface is generally important (etching may be anisotropic). The activation energy, due to a chemical process, is high ($> 40\,\mathrm{kJ/mol}$). These are clearly conditions which may favor defect revealing.
- Mass transport: the rate-limiting step is convective diffusion of active etching species from solution to the surface, or of etching products away from the surface. The activation energy is considerably lower ($< 20\,\mathrm{kJ/mol}$), being due to the temperature dependence of the diffusion coefficient (i. e., of the solution viscosity). Etching in this case has a *polishing* effect: the concentration gradient responsible for dissolution is higher at protrusions and lower at recessed areas of the surface, thereby giving rise to surface leveling. This effect is enhanced if a soluble surface film is involved in the etching process.

In this section we consider briefly the surface morphology and the basis for defect selectivity in the etching systems described above.

43.2.1 Chemical Etching

The chemical etchants considered in Sect. 43.1.1 (concentrated HCl or HBr for InP, concentrated alkaline solution for Si) are kinetically controlled systems: they show strong anisotropy. The atomic structure of the surface is important in determining etching morphology; this is very clear in the Si case [43.15, 16]. Such etchants may be highly defect selective, e.g., HBr for InP [43.26]. Eutectic etching corresponds to this class. Since the etch rate at *disturbed* areas of the surface (e.g., at dislocations) is high, etch pits result in this case. In Sect. 43.3.1 defect revealing in these chemical etchants is considered in more detail.

Chemical etching, on the other hand, may also be mass-transport controlled. Etching of III–V semiconductors (GaAs, InP) in aqueous solutions of halogens (chlorine, bromine, iodine) follows a chemical mechanism. The etch rate may be controlled by either kinetics or diffusion; this is also the case for etchants based on hydrogen peroxide [43.7]. Kinetic control gives defect selectivity [43.26]. Under mass-transport-controlled conditions these etchants are used for surface preparation and polishing. In the case of the halogens, water may be replaced as a solvent by an alcohol; bromine in methanol is a widely used etchant for GaAs.

43.2.2 Electrochemical Etching

The degree of defect selectivity in electrochemical etching depends on whether majority or minority carriers are responsible for etching.

Majority-Carrier Reactions

The kinetics of electrochemical etching of *p-type* semiconductors can be sensitive to surface structure: the rate constants of reactions such as those given by (43.3) may depend on surface properties, e.g., on whether the surface is Ga or N terminated. The defect selectivity of such anodic etching of p-type materials has not been studied much; it seems, however, in most cases to be inferior to that of other methods described in this chapter.

Anodic etching may be mass-transport limited. For example, for GaAs in alkaline solution the exponentially increasing current observed at positive potentials (Fig. 43.2a) levels off, becoming potential independent. This is due to limitation of the dissolution process by mass transport of OH^- ions in solution (43.9). The anodic etch rate depends on both the OH^- concentration and the hydrodynamics [43.20]. Another example is the anodic etching of p-type Si in acidic fluoride solution [43.27]. In both cases the semiconductor is *electropolished*.

Cathodic decomposition of n-type semiconductors via conduction-band electrons (as in (43.7)) is clearly not an attractive approach to etching. However, *de Wit* et al. [43.18] have shown that electrochemical reduction

of polycrystalline ZnO films reveals very clearly the grain boundaries. This indicates that, in certain cases, this approach could perhaps be interesting for defect revealing.

Minority-Carrier Reactions

Crystallographic and other defects usually give rise to electronic states in the bandgap of semiconductors. Such states can act as centers for very effective recombination of electrons and holes. In the onset of the anodic photocurrent–potential curve (Fig. 43.3b) surface recombination competes with the electrochemical reaction: holes required for dissolution of the solid are lost by recombination. This gives a significantly reduced etch rate at the surface defects, which show up as hillocks [43.7, 21, 28–31]. At positive potentials the photocurrent becomes independent of potential: the band bending at the surface is sufficient to separate effectively the photogenerated electrons and holes, thus preventing recombination. The limiting photocurrent can be determined by the light intensity or by mass transport in solution [43.25]. In the latter case electropolishing can be achieved.

43.2.3 Electroless Etching

As in Sect. 43.1.2 we must make a distinction here between electroless etching in the dark and under illumination. In this section we also consider an exceptional class of electroless system: Sirtl and adapted-Sirtl etchants.

Electroless Etching in the Dark

Since electroless etching requires strong oxidizing agents whose acceptor levels show a significant overlap with the valence band of the semiconductor, the rate of hole injection (e.g., (43.10)) is generally high and the etch rate is mass-transport controlled. In such a case one would not expect defect selectivity. However, in previous work [43.7, 20] we have shown that, since defects represent highly reactive areas on the surface, a galvanic cell can be formed between such areas and the more *noble* perfect surface: holes injected over the whole surface are used to etch preferentially defective areas, thus revealing defects as etch pits. This effect is expected when the etch rate is determined by the hole injection reaction (e.g., (43.10)). If, on the other hand, the oxidation reaction (43.9) is rate limiting and diffusion controlled, defects are not revealed. Instead polishing is observed. This subtle difference between the two cases can be understood on the basis of the electrochemistry of the systems [43.20]. In chemical etching systems free charge carriers (holes in the valence band) are not involved. Consequently, galvanic effects are not observed with chemical etchants.

Electroless Photoetching

As in the case of photoelectrochemical etching, electron–hole recombination competes with open-circuit photoetching. Etchants with H_2O_2 as electron acceptor have been used successfully to reveal defects in III–V semiconductors such as GaAs, GaP, and GaAsP [43.7, 21, 32].

Sirtl-Type Etchants

In 1961 *Sirtl* and *Adler* developed an etchant for Si based on aqueous solutions of CrO_3 and HF [43.33]. This *Sirtl etchant* proved very effective in revealing crystallographic defects in various orientations of Si. In addition, the system has been used for the delineation of junctions between layers of different dopant concentration. Subsequently, a modified form of the Sirtl etch was proposed for GaAs [43.34]. This system, termed diluted Sirtl with light (DSL), was highly successful in the characterization of III–V semiconductors. With the appropriate choice of solution composition, all crystallographic defects could be revealed in n-type, p-type, and semi-insulating GaAs with high sensitivity. A special feature of this system is the exposure of defects as hillocks during etching, not only under illumination but also in the dark.

The electrochemistry of Si in Sirtl and of GaAs in DSL etchants has been studied thoroughly [43.35–39] and far-reaching conclusions could be drawn with regard to etching mechanisms. It was shown that, except at very high HF concentrations, the dissolution of both semiconductors follows an electroless mechanism: the holes required for oxidation of the semiconductor are supplied by a Cr VI complex in solution.

In the case of GaAs [43.35–37] three ranges of etchant composition could be defined on the basis of the concentration ratio of HF to CrO_3. For $[HF]/[CrO_3]$ ratios below ≈ 10 and $[HF] < 10\,M$, the dissolution process is kinetically controlled (range A). In the dark p- and n-type etching kinetics are the same. Defects are revealed as hillocks and the defect sensitivity of n-type GaAs in higher. Illumination with suprabandgap light enhances both the etch rate and defect sensitivity of n-type GaAs but has no effect on p-type. For $[HF]/[CrO_3] > 20$ and $[HF] < 10$ (range B), etching is controlled by mass transport of the Cr VI complex in solution. Defect sensitivity is low for both p- and n-type

crystals. For solutions containing a HF concentration above 10 M (range C) a purely chemical attack on Si by HF species very likely occurs; this is analogous to the reaction between InP and HCl in (43.1).

An important feature of the electroless system is the formation of a passivating surface layer that contains both Cr VI and Cr III species. In range A the coverage by the surface film is high. Etching kinetics are determined by film formation (due to hole injection from Cr VI) and film removal (via intermediates of the semiconductor oxidation reaction with the aid of HF). In range B at relatively high HF concentration, the surface coverage decreases to low values. Hole injection by Cr VI, and thus GaAs oxidation/dissolution, are limited by mass transport of the oxidizing agent.

Defects are revealed as hillocks or ridges on the surface of p- and n-type GaAs crystals, both in the dark and under illumination. Defect sensitivity in the dark can be explained by assuming a reduced bond strength at the defect. This results in a local increase in surface coverage by the passivating film and, consequently, a reduction in etch rate. For n-type GaAs under illumination locally enhanced recombination at defects leads to a further decrease in etch rate and improved defect sensitivity. Other morphological features of the system which were explained include surface roughening and growth striations.

A subsequent study of the Si/Sirtl system [43.38, 39] showed features very similar to those observed in DSL, including the importance of the formation of an adsorbed mixed-valence chromium complex during etching.

43.2.4 Photogalvanic Etching

Photogalvanic dissolution of an n-type semiconductor is, in many respects, similar to that of photoelectrochemical dissolution: in the latter, the photocurrent at fixed light intensity is regulated via the applied potential; in the former, the photocurrent is determined by the rate of reaction of the oxidizing agent at the metal electrode, which depends on concentration, and the electrode area and (catalytic) activity. In photogalvanic etching, conditions can be chosen to correspond to the rising part of the photocurrent–potential curve (Fig. 43.3b); in this case competition between dissolution and recombination gives defect selectivity in many systems. Alternatively, a photocurrent limited by mass transport in solution can be achieved by ensuring that the the hole flux to the surface (dependent on light intensity) is much higher than the diffusion of a rate-determining species (e.g., OH^- ions) in solution. In this case photogalvanic polishing is obtained [43.25].

43.3 Defect-Selective Etching Methods

In this section we consider in more detail methods for defect-selective etching (DSE). We distinguish two types of system, namely:

- Orthodox, also called classical or preferential, etchants (strong acids, molten bases); these are, in fact, the chemical systems described in Sects. 43.1.1 and 43.2.1.
- Newer systems based on electroless etching in the dark and under illumination, including photogalvanic etching.

43.3.1 Orthodox Etching for Revealing Dislocations

In this subsection, factors determining the kinetics of formation of pits and their morphology are described, as is the role of thermodynamics, i. e., the elastic energy of dislocations, as well as the calibration of these orthodox etchants and special applications.

Kinetics-Related Conditions

The mechanism of pit formation in semiconductors is governed by kinetically controlled, surface chemical reactions in both acidic and molten-salt etches. The details and some examples of etching systems have been described in Sects. 43.1.1 and 43.2.1. This etching method was discovered and used as early as the 19th century for disclosing the crystallographic symmetry of natural crystals, which was read from the shape of etch pits on the crystal faces. When the association between the pits and dislocations was found, the etch pit density (EPD) became a standard parameter describing the structural quality of semiconductor single crystals and substrates for device producers. Apart from the density of dislocations, much information can be obtained from the morphology of the etch pits, provided the basic requirements for formation of well-developed pits are fulfilled. These criteria were discussed in detail in earlier reviews [43.5, 6, 8] and are summarized in the following. The formation of pits proceeds via repeated nucleation

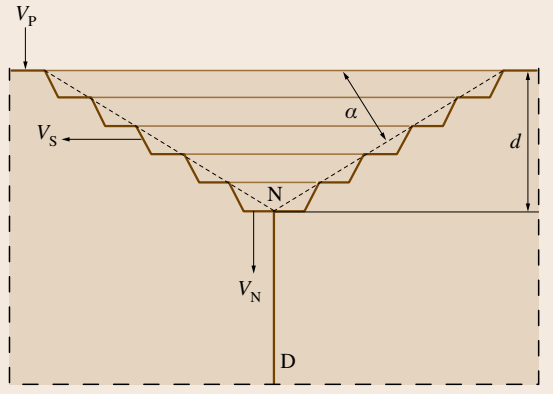


Fig. 43.7 Schematic drawing of an etch pit originating at a dislocation (D)

events at the emerging point of a dislocation on the etched surface, as shown in Fig. 43.7. Once the stable nucleus (N) is formed it grows further by horizontal step movement.

For practical applications, i. e., for obtaining clearly visible etch pits, the rate of formation of nuclei (V_N) must be larger than the step propagation velocity (V_S), while the rate of etching of the perfect surface (V_P) should be as low as possible (preferably zero). These criteria can be achieved by optimizing the etching temperature and the composition of the etchants, but the actual morphology of the etch pits, particularly the inclination angle of the side walls α, also depends on other factors such as the degree of decoration of dislocations (the presence of Cottrell atmospheres) and the type of dislocation. As an example, Fig. 43.8a shows the result of DSE of Ga-polar GaN in molten salts: deep pits are formed with $\alpha = 25-50°$ as derived from atomic force microscopy (AFM) section analysis in Fig. 43.8b. Etching of the same material in hot phosphoric acid or in a mixture of phosphoric and sulfuric acids (HH etch) yields very shallow pits with α below 10°. In the latter etch large shallow pits *sweep away* smaller neighboring pits, which might result in the erroneous estimation of the total density of dislocations. In order to avoid underestimation of the EPD in these acidic etches, the temperature and time have to be carefully optimized [43.40, 41]. Such optimization may allow edge and mixed/screw dislocations from the diverse size of pits to be distinguished [43.41], but from our experience it can be concluded that the procedure has to be repeated for samples grown in different conditions.

Factors Influencing Pit Morphology

There are a number of factors which may influence the morphology and relative size of etch pits formed on dislocations. These are:

- Crystallographic orientation (symmetry) of the etched surface
- Polarity of surfaces in compound semiconductors
- Composition of etching medium
- Changes of chemical composition around dislocations (Cottrell atmospheres) and in the matrix
- Position of dislocations with respect to the surface
- Elastic energy of dislocations represented by their Burgers vectors.

Symmetry. The correspondence between crystallographic symmetry and the shape of etch pits in single crystals is probably one of the best recognized and most

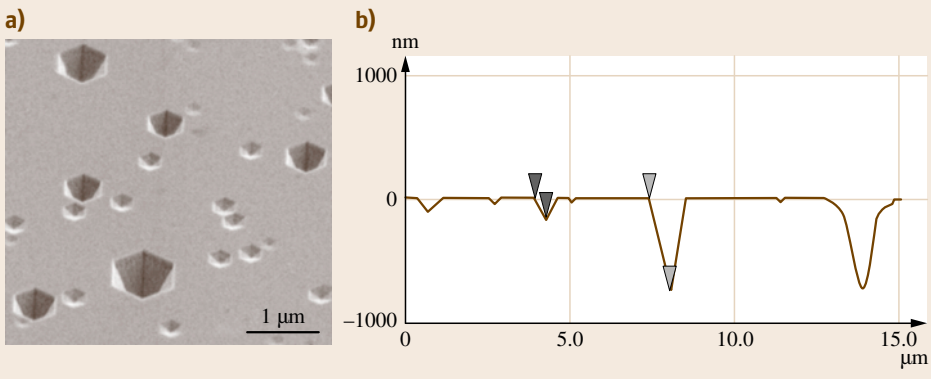


Fig. 43.8 **(a)** SEM image of deep pits formed on the Ga-polar surface of a GaN epitaxial layer during etching in molten KOH + NaOH eutectic. Sample tilt 45°. **(b)** AFM section profile across the pits from **(a)** (courtesy of G. Nowak)

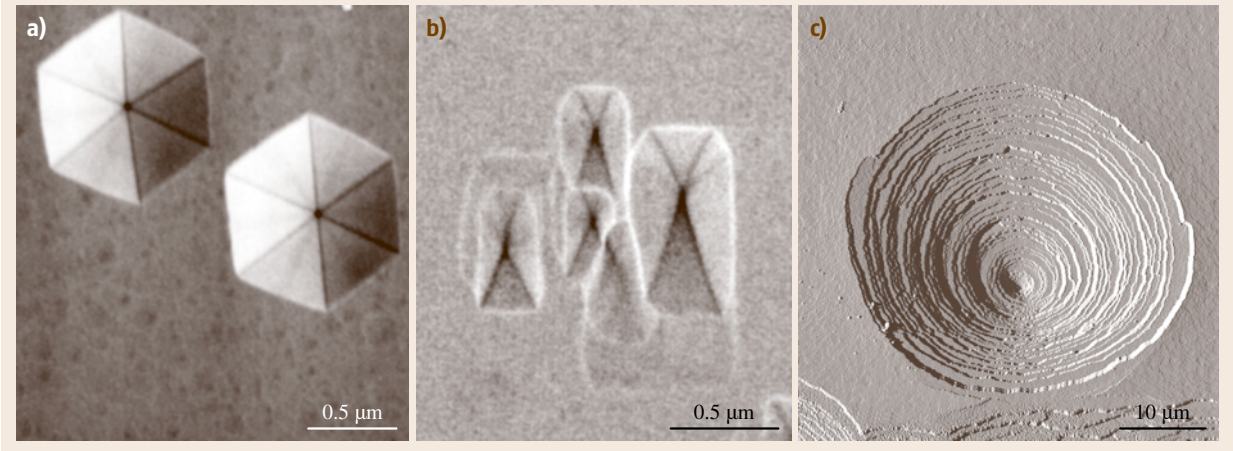


Fig. 43.9 (a,b) SEM and (c) AFM images of GaN single crystals etched on (a) (0001), (b) $(10\bar{1}0)$, and (c) $(000\bar{1})$ surfaces in molten salts (E-etch) (AFM image courtesy of S. Müller)

used advantages of the preferential etching method. Numerous etches are known for different semiconductors that show this characteristic, e.g., Sirtl etch for Si [43.33], molten KOH for GaAs [43.42], and several etching systems for InP [43.26, 43, 44]. A similar correlation has recently been shown for GaN in molten KOH + NaOH eutectic (E-etch after [43.12]) and is demonstrated in Fig. 43.9a,b: well-defined hexagonal and rectangular etch pits are formed during etching of (0001) Ga and $(10\bar{1}0)$ surfaces, respectively. It can be concluded that a properly calibrated etching method constitutes a fast and unambiguous tool for establishing the crystallographic orientation of semiconductor bulk crystals and epitaxial layers.

Polarity. Preferential etching is also sensitive to the bonding direction between atoms forming compound semiconductors. As a rule, polar surfaces show markedly different growth and etching behavior. In zincblende III–V materials (e.g., GaAs and InP) polar {111}B surfaces are very reactive because the presence of the group V atom at this face gives a higher electron density than that of the {111}A face. (In the extreme case the B face would have an electron pair in the dangling bond while this would be missing at the A face.) As a result of this difference it is easy to recognize the polar {111} faces using well-known etches. The sensitivity of etching methods to the atomic configurations of the etched surfaces has an important practical application for recognition of nonequivalent ⟨110⟩ and $\langle\bar{1}10\rangle$ directions on the nonpolar {001} faces of III–V compound semiconductors. For instance, the etch pits formed on the (001) surface of GaAs during etching in molten KOH are elongated in the $[0\bar{1}1]$ direction, as was unequivocally established by calibration with the x-ray Lang technique [43.45]. Consequently, the definition of the orientation and identification flats on the zincblende III–V substrates, given in standard commercial specifications, is based on the shape of the etch pits, as shown in Fig. 43.10.

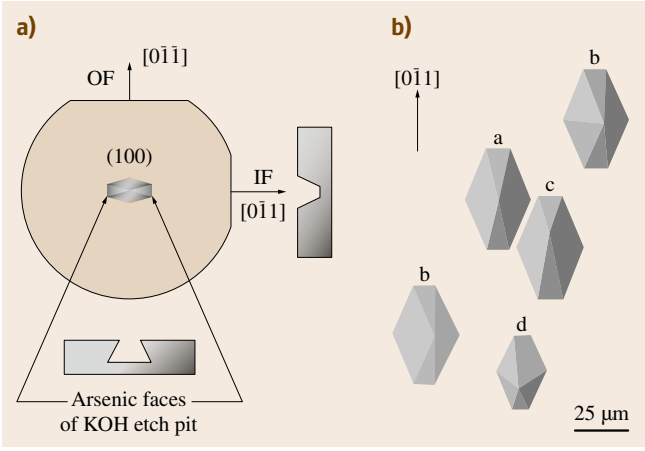


Fig. 43.10 (a) Example of commercial specification of a GaAs substrate with the description of orientation (OF) and identification flats (IF) on the basis of the shape of KOH-related etch pits. (b) Pits formed in molten KOH on the (001) surface of GaAs at dislocations differently inclined with respect to the surface

III–V materials with a wurtzite lattice (GaN, AlN) have very strong polarity-related anisotropy of properties, which shows up, for example, in markedly different chemical reactivity of (0001) Ga and (000$\bar{1}$) N surfaces. Following standard assignment based on comparative hemispherically scanned x-ray photoelectron diffraction (HSXPD), TEM, and etching studies [43.46–48], it was shown that N-polar surfaces of GaN can be dissolved in dilute aqueous solution of KOH/NaOH even at room temperature, while Ga-polar surfaces remain intact in such etching conditions. Fast dissolution of the N-polar surface of GaN in KOH was explained by adsorption of hydroxide ions at the negatively charged dangling bond of each nitrogen atom at this surface and subsequent formation and dissolution of gallium oxide [43.49]. As a result, numerous pyramids are formed on the (000$\bar{1}$) N surface and, once the whole surface is covered by them, the etching process is terminated. As for Si (111) surfaces etched in concentrated KOH solution [43.15, 16], such pyramids are bounded by chemically stable low-index {1$\bar{1}$01} planes and are not related to any specific defects (though for GaN this conclusion is still tentative). Similar behavior was reported for the nitrogen-polar surface of AlN [43.11] and this makes DSE of these surfaces more demanding. Dislocations on the N-polar surface of GaN single crystals can be revealed in molten eutectic [43.12] at relatively low temperatures as compared with the Ga-polar surfaces. In contrast to the Ga-polar surface, the resultant etch pits are circular, with a very irregular terraced morphology (Fig. 43.9a,c). The most convincing example of the difference of the polarity-dependent pit morphology was obtained by revealing dislocations in the vicinity of inversion domains (IDs) in Mg-doped homoepitaxial GaN [43.50]. The upper part of the image in Fig. 43.11a shows hexagonal pits formed on the Ga-polar surface while in the lower N-polar area circular pits are visible.

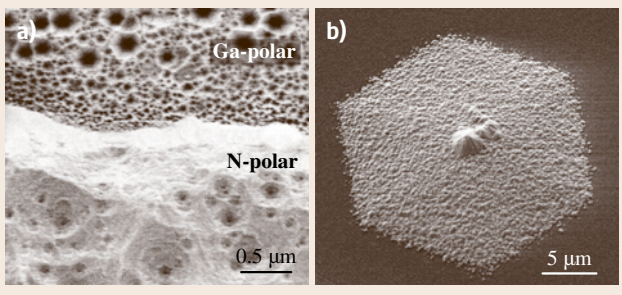


Fig. 43.11a,b SEM images of (**a**) GaN epitaxial layer with the neighboring Ga- and N-polar areas etched in molten E+M etch [43.45] (courtesy of G. Kamler) and (**b**) GaN N-polar epitaxial layer after etching in hot HH [43.46]

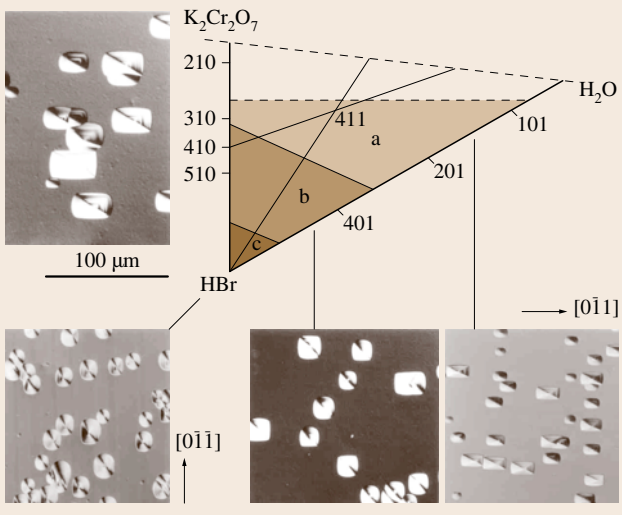


Fig. 43.12 HBr-rich corner of the BCA ternary etch system for InP. The differential interference contrast (DIC) optical images show the morphology of etch pits as a function of composition of etchants (with permission from [43.26])

DSE of N-polar heteroepitaxial GaN layers is more difficult because, in addition to the high reactivity of this side, a higher density and variety of defects (e.g., inversion domains of different diameters, higher level of impurities, i. e., carrier concentration [43.51–53]) are inherent to these materials. Etching in molten salts results in fast dissolution of N-polar heteroepitaxial layers and overall roughness, without the possibility of attributing the tiny etch features to particular defects. It seems more promising to etch in hot acids, which reveals large and small inversion domains (IDs) in the form of protruding pyramidal etch features, as shown in Fig. 43.11b. However, evaluation of dislocation density in this material using orthodox etching methods does not seem to be viable.

Etchant Composition. The morphology of etch pits can be tailored by changing the ratio of constituents of the etching solution, etching conditions (e.g., temperature) or by adding so-called inhibitors. Theoretical considerations and rules were discussed and variable pit morphologies in different etching systems were demonstrated in earlier reviews [43.5,6]. A very clear example of the dependence of the pit morphology on the etchant composition was found in the orthodox HBr-$K_2Cr_2O_7$ (BCA) etching system developed for InP [43.26]. In

the HBr-rich corner of the ternary BCA diagram there are three composition fields, denoted a, b, and c in Fig. 43.12. Etchants from these fields result in the formation of pits with different morphology: the pits formed in region b are square, while etchants from regions a and c yield rectangular pits, elongated in the $[0\bar{1}\bar{1}]$ and $[0\bar{1}1]$ directions, respectively.

The kinetics of DSE can be influenced by adding traces of ions (inhibitors) to the solution. Their role is to *poison* kinks and ledges inside the pits, which slows down horizontal movement of steps during etching (decrease of V_S in Fig. 43.7). This effect can have practical application when kinetics-related conditions of pit formation are not fulfilled, i. e., when the pits are too shallow as in etching of GaN in hot acids. It was recently shown that, by adding Al^{3+} and Fe^{3+} ions to H_3PO_4, merging of pits on GaN heteroepitaxial layers can, indeed, be avoided and well-defined deep pits are formed [43.54].

Effect of Decoration and Composition. In semiconductors, foreign atoms (dopants, impurities) are attracted by dislocations due to the presence of a strain field and charge (Coulomb interaction). The resultant impurity (Cottrell) atmosphere may have a dual influence on the formation of etch pits:

1. By releasing the strain, the elastic energy of the dislocation is diminished, which should decrease the rate of nucleation (V_N in Fig. 43.7).
2. The chemical potential is changed and this may locally change the chemical reactivity in the given etching solution.

As was already pointed out by *Amelinckx* [43.2] it is difficult to predict which of these two factors will prevail. GaAs etching in molten KOH resulted in the formation of larger etch pits on so-called grown-in (i. e., strongly decorated) dislocations than on stress-induced dislocations [43.8]. A similar effect was observed in bulk GaN single crystals on which the size of pits formed on grown-in and indentation-induced dislocations was compared [43.55]. Morphology of pits can also be strongly influenced by heavy doping or alloying, as was shown for In-doped GaAs: the elongated hexagonal pits formed in molten KOH on nondoped GaAs (similar to these shown in Fig. 43.10b) were gradually transformed into regular hexagonal pits with increasing content of In [43.56].

Geometrical Position of Dislocations with Respect to the Surface. This effect seems to be obvious and was recognized already in the early 1960s [43.2] and subsequently well substantiated in numerous papers and reviews, e.g., [43.5, 8, 57, 58]. Most impressive was the work distinguishing five groups of etch pits and thereby five directions of dislocation lines in GaAs epitaxial layers after molten KOH etching [43.59]. These are: $a = \langle 001 \rangle$, $b = \langle 011 \rangle$ or $\langle 101 \rangle$, $c = \langle 112 \rangle$, $d = \langle 211 \rangle$, and $e = \langle 121 \rangle$ directions. Similar results were later obtained on bulk liquid-encapsulated Czochralski (LEC)-grown GaAs crystals and described in a review paper [43.8]. In Fig. 43.10b etch pits are shown in a GaAs substrate demonstrating the presence of four differently inclined dislocations (*a–d* types) in one place.

The majority of dislocations in GaN heteroepitaxial layers are perpendicular to the (0001) surfaces because

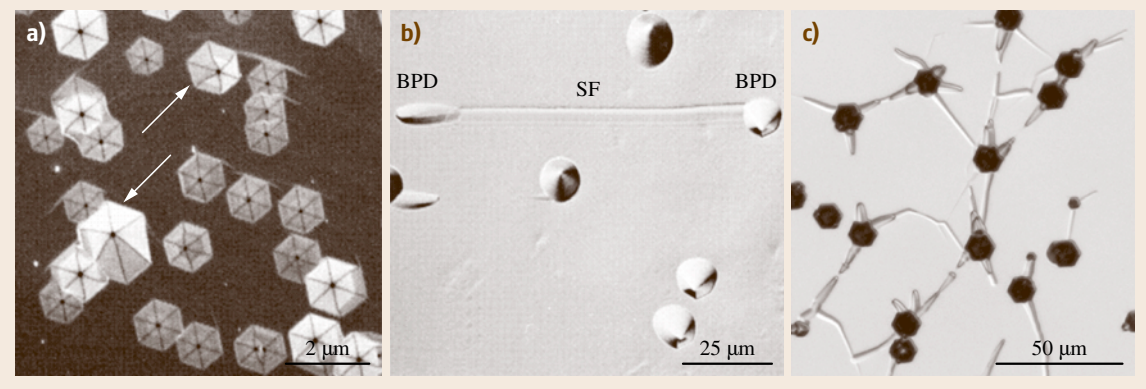


Fig. 43.13 **(a)** SEM image of thick HVPE-grown GaN layer after etching in molten E at 450 °C for 5 min. **(b,c)** DIC images of SiC substrate after etching in molten E at 520 °C for 5 min. The sample in **(b)** is 8° off axis

these are threading dislocations. The etch pits formed on them are, therefore, center-symmetrical, as shown in Fig. 43.9a. The same holds for bulk GaN single crystals and thick hydride vapor phase epitaxy (HVPE)-grown layers, though some inclined dislocations could be found (see two the non-center-symmetrical pits marked by arrows in Fig. 43.13a and the pit on the N-polar surface in Fig. 43.9c). In bulk and epitaxial SiC, so-called basal-plane dislocations are easily recognized from the characteristic morphology of shell-like etch pits. They are frequently observed both in SiC substrates and in 8° off-axis epitaxial layers [43.58, 60, 61]. The characteristic appearance of basal-plane dislocations (BPDs) and a bounding stacking fault (SF) in misoriented SiC is shown in Fig. 43.13b. In Fig. 43.13c, for comparison, pits are shown on an exactly (0001)-oriented SiC surface: hexagonal etch pits are on dislocations perpendicular to the surface (screw dislocations (SDs)), and grooves on dislocations almost parallel to the surface, i.e., basal plane dislocations (BPDs). In Fig. 43.13c the BPDs form dislocation nodes with SDs and a three-dimensional (3-D) network in the bulk of the material.

Thermodynamic Factors: Elastic Energy of Dislocations

From Cabrera's theory on the thermodynamics of pit formation [43.6] it follows that the critical value of the chemical potential difference ($\Delta\mu$) of a stable nucleus of a pit at the outcrop of a dislocation (N in Fig. 43.7) depends inversely on the elastic energy (E_{el}) of the dislocation

$$\Delta\mu = 2\pi^2\Omega\frac{\gamma^2}{E_{el}}\,, \tag{43.11}$$

where γ is the edge free energy and Ω the molecular volume.

The elastic energy value differs for different types of dislocations [43.63]

$$E_s = Gb^2\alpha\,, \tag{43.12}$$

$$E_e = Gb^2\alpha\left(\frac{1}{1-\nu}\right)\,, \quad \text{and} \tag{43.13}$$

$$E_m = Gb^2\alpha\left(1-\nu\cos^2\frac{\theta}{1-\nu}\right)\,, \tag{43.14}$$

for screw, edge, and mixed dislocations, respectively (where G is the shear modulus, b the Burgers vector, α a geometrical factor, ν Poisson's constant, and θ the angle between screw and edge components of the Burgers vector of mixed dislocations). Heteroepitaxial layers of GaN are the best material for considering the influence of elastic energy of dislocations on the formation of pits during orthodox etching: in this material all three types of dislocations, including nanopipes, usually coexist. In addition, numerous TEM studies of cross-sectional specimens have showed that the vast majority of dislocations in epitaxial GaN layers are of the threading type and are perpendicular to the surface. As a result, the influence of the tilt of a dislocation line on the size and/or morphology of pits can be excluded. Figure 43.14a shows the typical surface morphology of a heteroepitaxial GaN sample after

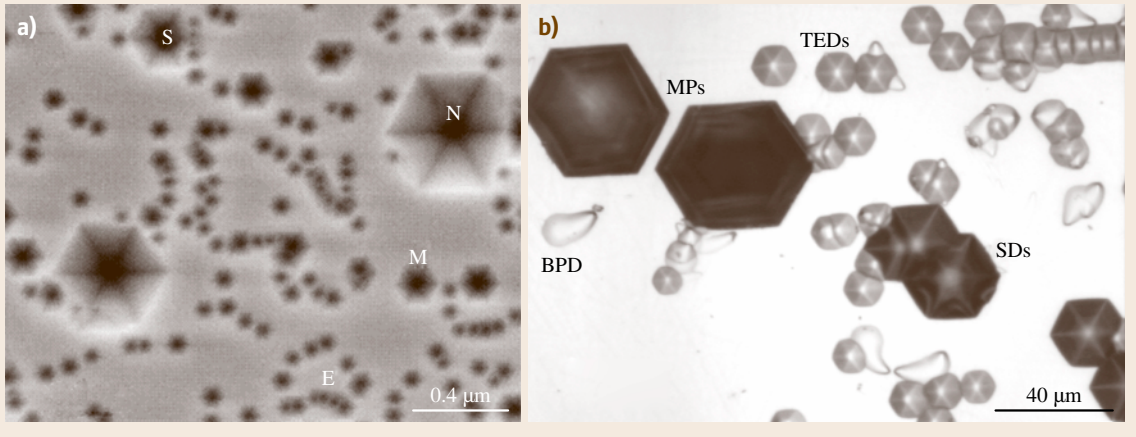


Fig. 43.14 (a) SEM image of heteroepitaxial GaN layer after etching in molten E+M. (b) Optical image of etch pits formed on different defects in 6H-SiC wafer during etching in molten KOH (with permission from [43.62])

etching in molten E+M orthodox etch (eutectic of KOH + NaOH + 10%MgO [43.64]). Four size grades (denoted N, S, M, and E) can be discerned in this material, representing nanopipes and screw-, mixed-, and edge-type dislocations, respectively. The assignment of the pit size to the type of dislocations was done on the basis of recent direct calibration of orthodox etching by TEM [43.64–66].

In the wurtzite lattice of GaN the Burgers vectors of dislocations are

$$\boldsymbol{b}_\mathrm{e} = \frac{1}{3}\langle 11\bar{2}0\rangle\,, \qquad (b_\mathrm{e} = a\,,\ b_\mathrm{e}^2 = a^2)\,,$$

$$\boldsymbol{b}_\mathrm{m} = \frac{1}{3}\langle 11\bar{2}3\rangle\,, \qquad (b_\mathrm{m} = \sqrt{c^2 + a^2}\,,\ b_\mathrm{m}^2 = 3.66a^2)\,,$$

$$\boldsymbol{b}_\mathrm{s} = [0001]\,, \qquad (b_\mathrm{s} = c\,,\ b_\mathrm{s}^2 = 2.66a^2)\,,$$

$\boldsymbol{b}_\mathrm{nano} = n \times \boldsymbol{b}_\mathrm{s}$ where $n = 1, 2, \ldots$, and a, c are the lattice parameters.

In the ideal lattice structure of an etched GaN sample, the size of pits should depend on the magnitude of the Burgers vector of dislocations, i. e., the largest pits should be formed on nanopipes and the smallest on the edge dislocations. This pit size sequence is indeed valid, with the exception of screw dislocations, on which the pits are usually larger than on the mixed ones. The arguments for explaining this seeming discrepancy have been discussed in recent work [43.66] and are based on the fact that both edge and mixed dislocations are characterized by a larger deformation of the lattice, i.e., the contribution of the Poisson-constant-related term in (43.12–43.14), than the screw dislocations. The resultant higher attractive forces may be responsible for more effective decoration of edge and mixed dislocations and, in this way, for a release of strain around these defects. This, in turn, would result in a less favorable energetic condition for the formation of etch pits on edge and mixed dislocations.

The energy of defects has an even more pronounced effect on the size of pits in SiC. Very large differences between the size of pits formed on micropipes, screw, threading, and basal plane dislocations were found after an optimized etching procedure in molten KOH [43.62], as is obvious from Fig. 43.14b.

Calibration of Etching

Each new DSE system requires confirmation of its reliability in revealing all dislocations. Since the size and morphology of pits may vary depending on the type and status of dislocations, as was shown in the previous section, calibration by a direct method may allow one to identify the types of dislocation and provide an unequivocal interpretation of the etch features. Different approaches are used for this purpose, e.g., x-ray topography of etched samples [43.67] (suitable for dislocation density below $10^5\,\mathrm{cm}^{-2}$), comparison with cathodoluminescence (CL) [43.68] or another calibrated etching method [43.40], sequential etching and calibration by TEM (e.g., [43.69, 70]). The latter is the most attractive because, apart from the direct association of the etch pit with the underlying defect, it yields information on the exact type of dislocation and on any additional characteristic features, such as decoration. The method became very popular especially after the introduction of the focused ion beam (FIB) technique, though the conventional cross-sectional approach is still used [43.71, 72]. Figure 43.15 shows a typical set of TEM images (specimen prepared by FIB) of the etch

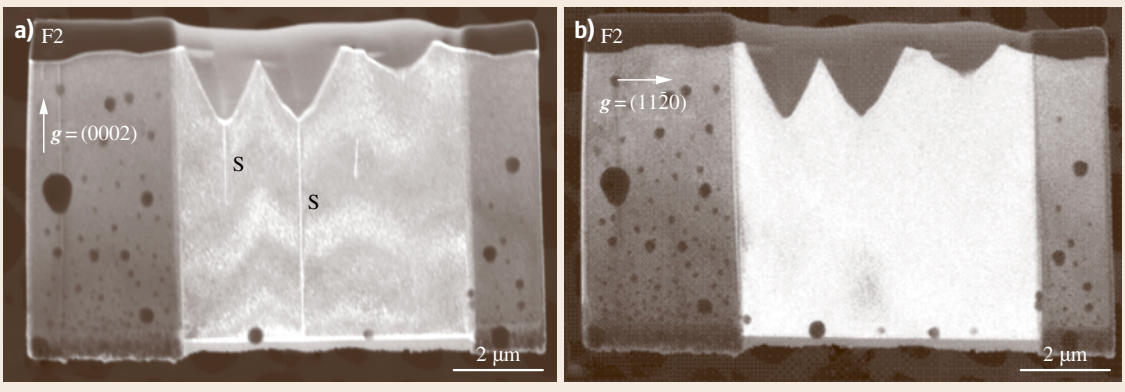


Fig. 43.15a,b Bright-field TEM images taken with different diffraction conditions showing the association of large etch pits with screw dislocations (with permission from [43.59])

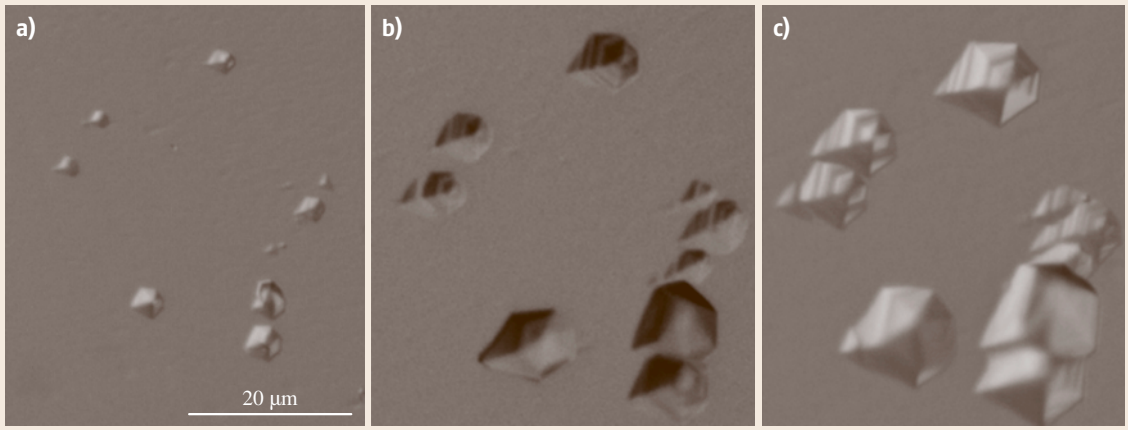


Fig. 43.16a–c DIC images of ZnO single crystal after sequential etching in 0.7% aqueous solution of HCl for **(a)** 30 s, **(b)** 60 s, and **(c)** 90 s

pits and underlying dislocations in thick HVPE-grown GaN. The method is particularly useful for samples with a moderate dislocation density, for which the conventional cross-sectional specimen preparation method is not effective because of the low probability of finding any defects in the thin foil.

The simplest way to verify the reliability of an orthodox etch for revealing dislocations is sequential etching: the pits formed on dislocations develop in size after each subsequent etching step and always have a point bottom indicating the position of the outcrop of the linear defect, as demonstrated by the set of images (made with the same magnification) in Fig. 43.16. Equally effective and simple is simultaneous etching of a two-sided polished sample. This method is, however, limited to relatively thin samples (e.g., substrates for epitaxy) containing threading dislocations.

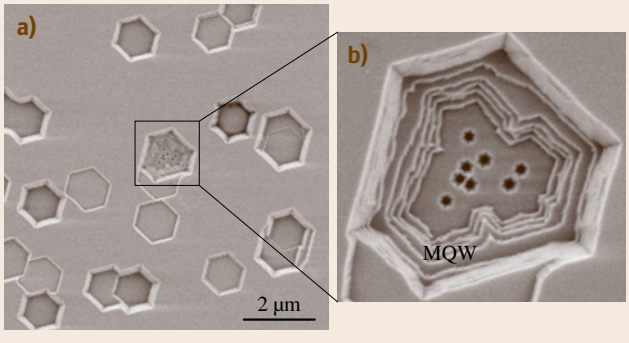


Fig. 43.17a,b SEM images of **(a)** MQW GaN-based laser structure after deep molten E+M etching. In **(b)** an enlarged fragment of image **(a)** is shown (etching and images: courtesy of G. Kamler)

Special Applications of Orthodox Etching

Well-controlled orthodox etching has frequently been used for revealing the origin of defects in epitaxial layers and device structures. For this purpose multiple (sequential) etching and photography are employed, which allows in-depth tracing of the defects in the sample. The results presented in [43.73] illustrate well the potential of this method: the clusters of screw dislocations revealed on the top of thick SiC epitaxial layers (up to several micrometer) were shown to be formed as a result of the closing of micropipes from the substrate. A similar approach was used for tracking the origin of pairs of dislocations in SiC epitaxial layers [43.74] and conversion of BPDs from the substrates into threading edge dislocations (TEDs) in SiC epitaxial layers [43.75]. Application of orthodox etching for examination of device structures is also very attractive since it permits one to establish at which interface in multilayered samples the dislocations are nucleated. The principle of the method is based on the fact that the point-bottomed etch pits formed on dislocations are transformed into flat-bottomed pits when the interface at which nucleation occurred is reached [43.76]. The method was recently used for studying GaN-based laser structures [43.77] and was calibrated by TEM [43.78]. Figure 43.17a shows the result of etching in molten E of the laser structure similar to that from [43.78], in which the dislocations were nucleated at different interfaces (the depth of the flat-bottomed pits is different),

with the exception of a cluster of dislocations propagating from the layers beneath the multiple quantum well (MQW) (see the enlarged central complex pit in Fig. 43.17b).

43.3.2 Electroless Etching for Revealing Defects

In another approach to etching, described in Sects. 43.1.2–43.1.4, free charge carriers (holes) localized at the surface cause the breaking of bonds. Holes can be supplied by an external voltage source (electrochemical etching), by an electron acceptor in solution (electroless etching) or by illumination (photoetching). Electrochemical etching has not been widely used for defect studies. Electroless etching, which operates under open-circuit conditions, has proved very successful. One of the most versatile electroless etchants for GaAs is the CrO_3/HF/H_2O (DSL) system [43.37, 79]. This can be used for revealing dislocations both in the dark and under illumination. For wide-bandgap semiconductors such as GaN and SiC electroless etching in the dark is not possible (Sect. 43.1.3). In this case (photo)electrochemical or photogalvanic etching is an option. The latter is also described in the *Photogalvanic Etching* paragraph for revealing dislocations. These etching approaches can also be used for studying electrically active inhomogeneities in semiconductors. Special applications of electroless etching are dealt with in the final section.

Dislocations

We first consider the CrO_3/HF/H_2O system in the dark and then two open-circuit photoetching approaches.

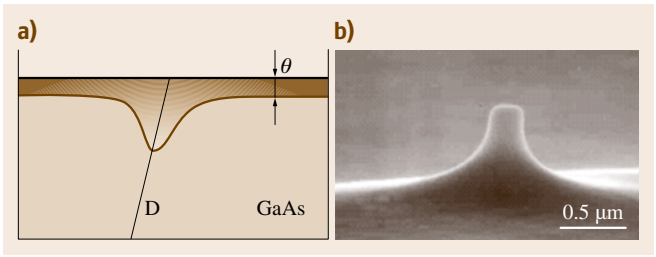


Fig. 43.18 **(a)** Hypothetical section profile of the thickness of the passivating layer θ formed on GaAs surface with an emerging dislocation D, immersed in CrO_3/HF/H_2O etching solution. **(b)** SEM image of the protruding etch feature formed on a dislocation in GaAs during dark etching in the same solution

CrO_3/HF/H_2O in the Dark. In this DS system (without light) the holes required for oxidation and dissolution of the semiconductor are injected into the valence band by hexavalent chromium ions in solution (Sect. 43.2.3). It has been shown that defect-selective etching in this case is determined by a surface passivating layer consisting of a Cr mixed-valence complex. The thickness of the passivating layer θ (or its coverage) determines the final morphological characteristics of the dislocation-related etch features [43.37]. The deformation field around dislocations reduces the Ga–As bond strength and, in this way, locally increases the thickness θ of the passivating layer (Fig. 43.18a). As a result during dark etching the outcrops of dislocations are revealed in the form of nanometer-size hillocks (tips), as shown in Fig. 43.18b. The surrounding Cottrell atmosphere has only a weak influence on the final shape of the etch features at the dislocation; it may contribute by diminishing the dislocation-related deformation field due to

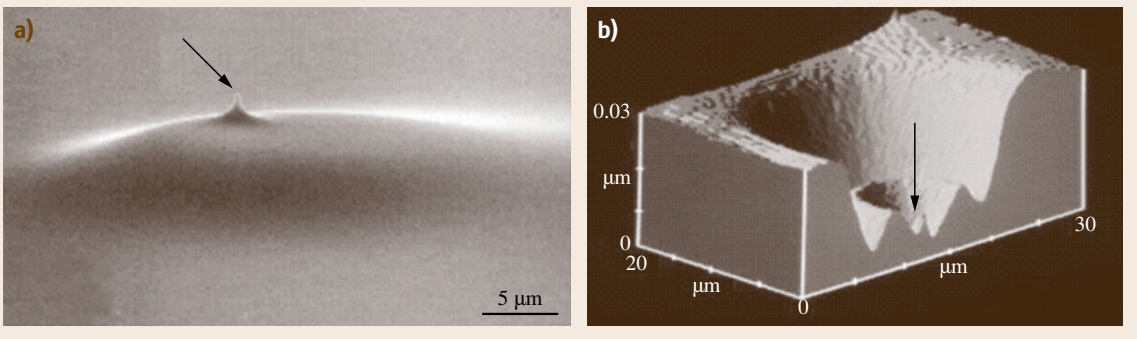
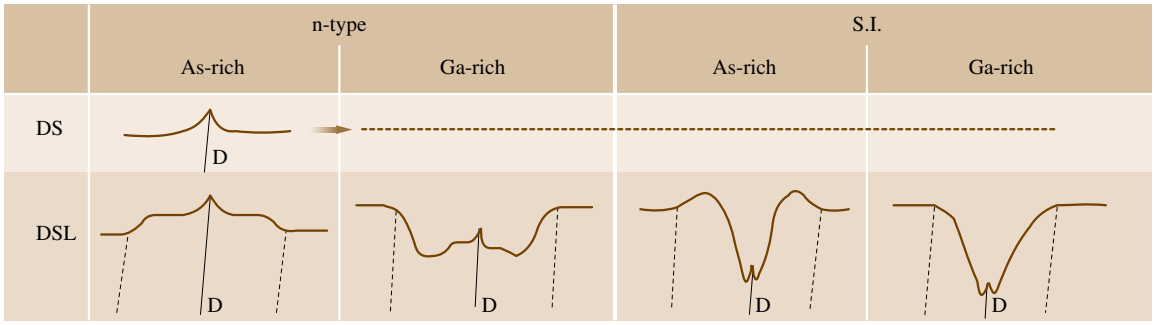
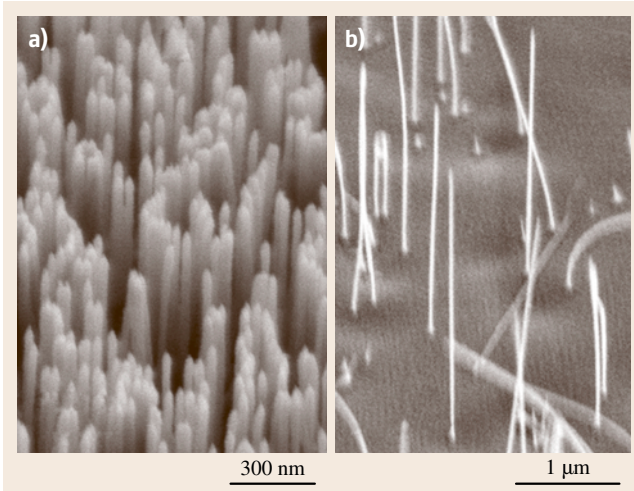


Fig. 43.19 **(a)** SEM and **(b)** phase-stepping microscopy (courtesy of P. Montgomery) images of dislocations and related Cottrell atmospheres, revealed by photoetching of n-type GaAs grown from: **(a)** slightly As-rich and **(b)** Ga-rich melt. The *arrows* indicate the etch features formed on dislocations

Table 43.1 Schematic representation of the surface profiles across the etch features formed on dislocations in GaAs after dark etching (DS) and photoetching (DSL) (with permission from [43.9]) (D: dislocations, dashed line: extent of electrically active zone with properties different from those of the dislocation-free matrix)





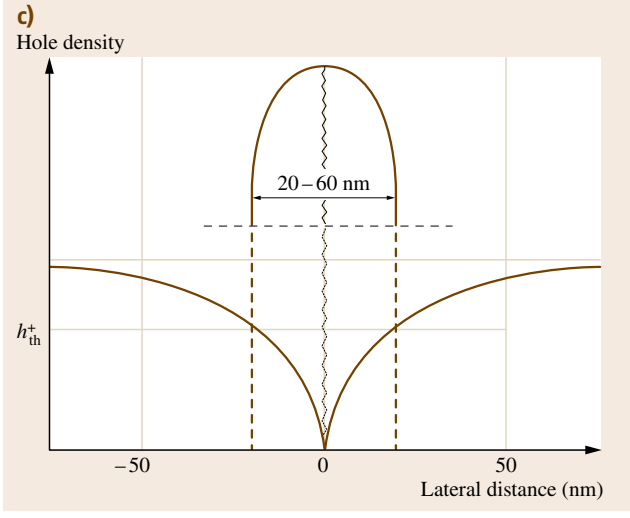


Fig. 43.20a–c SEM images of whisker-like features formed on dislocations during open-circuit photoetching of **(a)** high-dislocation-density MOCVD-grown and **(b)** low-dislocation-density HVPE-grown GaN layers. **(c)** Model of the formation of a whisker on a single dislocation during open-circuit photoetching of n-type GaN (with permission from [43.80]) ◀

the local increased concentration of foreign atoms or native point-type defects.

Photoetching

DSL System. When the surface of the semiconductor is illuminated with suprabandgap light during etching, complex etch features are formed on the dislocation sites. The final morphology of the etch features now depends on the electrical properties of both the dislocation and the surrounding atmosphere. Similar to etching in the dark, submicron protrusions are formed at the outcrops of dislocations (Fig. 43.19a), but now they reflect a cylindrical region depleted of carriers due to the recombination of electrons and holes at the dislocation. The dislocation-related atmospheres are revealed as either hillocks or depressions depending on their electronic nature: when the local recombination is stronger than in the matrix, the etch rate is locally decreased and large hillocks or ridges are formed [43.81] as can be recognized in Fig. 43.19a. On the other hand, the atmosphere containing excessive holes (as in GaAs grown from Ga-rich melt, in which Ga^{++}_{As} double acceptors are present) locally increases the etch rate and depressions are formed around the dislocation-related protrusions (Fig. 43.19b). Such complex etch features are formed in As- and Ga-rich GaAs during DSL photoetching and the electronic nature of defects responsible for the influence

of atmospheres was disclosed by photoluminescence and EBIC studies [43.82, 83]. It was also shown that the section profile across the complex etch features around dislocations constitutes a fingerprint of the type of GaAs, as was discussed in [43.9] and demonstrated by the data in Table 43.1.

Photogalvanic Etching. Electroless etching of wide-bandgap semiconductors (GaN, SiC) can be performed only with the help of ultraviolet (UV) light and a supporting electrode (photogalvanic etching) for the reasons discussed in Sect. 43.1.4. The only widespread etching system used to date is based on aqueous solutions of KOH with dissolved oxygen as electron acceptor [43.24, 25], although more complex etching solutions containing a strongly oxidizing component ($K_2S_2O_8$) are emerging [43.85]. Since dilute KOH solutions (in the range 0.002–0.01 M KOH) are suitable for revealing dislocations, it is necessary to employ stirring during photoetching in order to ensure the supply of OH^- ions for the oxidation reaction at the surface. Other technical details and the limitations of this method have been discussed in several recent papers [43.11, 51, 52, 84, 86]. During etching of GaN in this system the recombination of electrons and holes at the dislocations is very effective; this leads to formation of whisker-like etch features with almost unlimited length (Fig. 43.20a,b). The diameter of the whiskers remains in the tens of nanometer range and represents a tube of the material from which holes are depleted due to recombination of photocarriers at the dislocation (see the model in Fig. 43.20c, in which h_{th}^+ describes the critical number of holes required for dissolution of a GaN molecule as follows from (43.3)). The fact that each whisker contains a dislocation was confirmed by direct TEM calibration of the photoetched samples [43.51, 53, 87–89].

Photogalvanic etching is also effective for revealing different defects in SiC. It was shown that dislocations, stacking faults, macropipes, and chemical inhomogeneities could be visualized in aqueous KOH solutions used with UV light [43.58]. Figure 43.21 shows a characteristic image of protruding features formed on dislocations parallel to the surface and pinned by two micropipes.

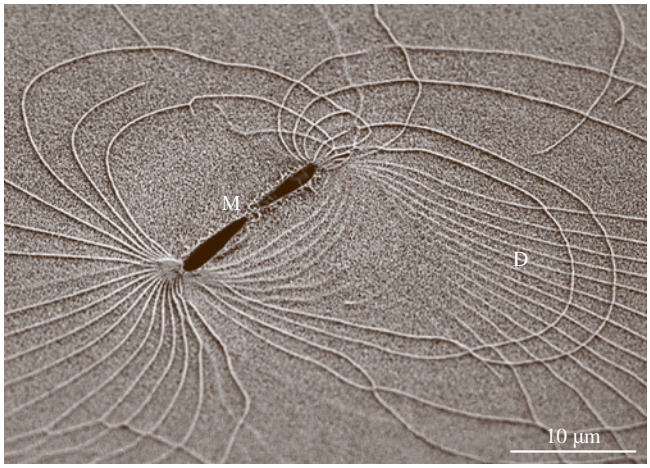


Fig. 43.21 SEM image of etch features formed on micropipes (M) and pinning dislocations (D) parallel to the (0001) Si surface of SiC substrate revealed by photoetching

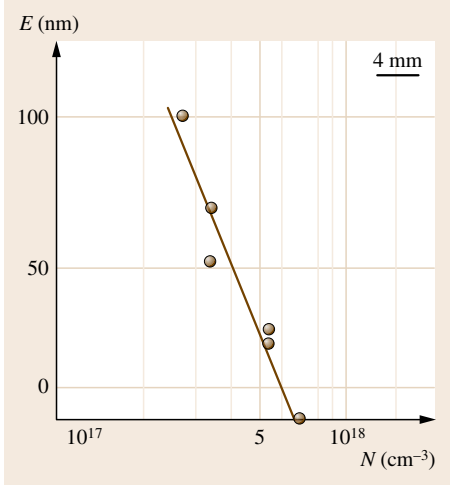


Fig. 43.22 Plot of relative etch depth E versus carrier concentration N for n-type GaAs photoetched in DSL (with permission from [43.84])

Electrically Active Inhomogeneities

Electroless photoetching constitutes an attractive tool for examination of extended electrically active inhomogeneities inherent to compound semiconductors, e.g., growth striations [43.90]. It was shown that well-controlled photoetching is ultrasensitive to very small differences in carrier concentration both in n-type and in semi-insulating (SI) semiconductors and, after calibration by an appropriate method, can be used for quantitative evaluation of local differences in carrier concentration. Figure 43.22 shows a plot of the relative photoetch depth E of Si-doped GaAs as a function of the carrier concentration n; the latter was determined by the EBIC method. This strong inverse dependence of E on n can be understood on the basis of the illuminated Schottky diode. Electron–hole pairs are generated by light to a depth determined by the absorption coefficient of the semiconductor. For photoetching to occur,

the valence-band holes must reach the surface. Two processes contribute. Holes created within the space-charge (depletion) layer are driven to the surface by the electric field. Since the thickness of the depletion is inversely proportional to the square root of the electron density ($W_{sc} \propto 1/n^{1/2}$), this contribution will decrease as the dopant density increases. Holes may also reach the edge of the space-charge layer by diffusion and then migrate to the surface. Photocurrent measurements on GaP Schottky diodes and p–n junctions [43.91] showed a strong inverse dependence of the apparent diffusion length of the minority carriers on the carrier concentration in the range of Fig. 43.22. As a result of these two effects, diffusion and migration, the hole flux to the surface of an illuminated SI or n-type semiconductor may be expected to decreases as the electron density increases. This, of course, will lead to a drop in photoetch rate, as shown for GaAs in Fig. 43.22.

A similar result was subsequently demonstrated for n-type GaN [43.92, 93]. The local photoetch rate again showed an inverse dependence on carrier concentration, determined in this case by Raman measurements. Very convincing experimental evidence of the general validity of this relationship has been found during etching of the N-polar heteroepitaxial GaN layer containing Ga-polar inversion domains (IDs). During simultaneous growth of areas of different polarities more impurities and doping element are incorporated into the N-polar material; it becomes more heavily doped, say n^+, while Ga-polar IDs have lower carrier concentration, say n. During photoetching the latter areas are etched more rapidly, i. e., deep craters are formed at IDs, as shown in Fig. 43.23a and explained by the model in Fig. 43.23b. The experimental details of such peculiar etching behavior have been discussed in recent papers [43.51–53].

Photoelectrochemical measurements on n-type GaN electrodes in the same KOH solution as used for electroless photoetching confirm the explanation given above. As the dopant density of the electrode is increased the onset for photocurrent is shifted to positive potential, i. e., the hole flux to the surface decreases. This trend can explain why bulk GaN single crystals, unintentionally doped (contaminated) with oxygen and carbon to the level of carrier concentration close to $10^{20}\,cm^{-3}$, cannot be photoetched in aqueous KOH solutions.

The relationship between the etch depth and carrier concentration established for GaAs in the DSL etching system has recently been confirmed for photogalvanic etching of GaN in KOH solutions [43.92]. A quantitative correlation of the etch rate and carrier concentration was obtained by micro-Raman determination of the carrier concentration [43.93].

Special Application of Electroless Etching

There are some features of this method which allow analysis of specific defects occurring in compound semiconductors. It was shown that the popular Abra-

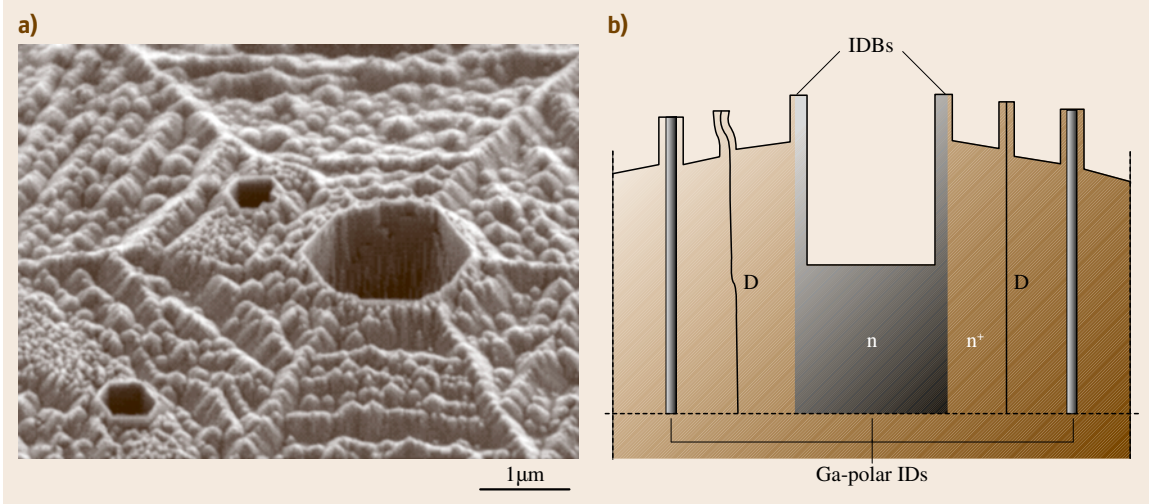


Fig. 43.23 **(a)** SEM image of etch features formed during photoetching of N-polar GaN heteroepitaxial layer containing inversion domains. **(b)** Schematic representation of a cross-section of the photoetched GaN layer from **(a)** (with permission from [43.53])

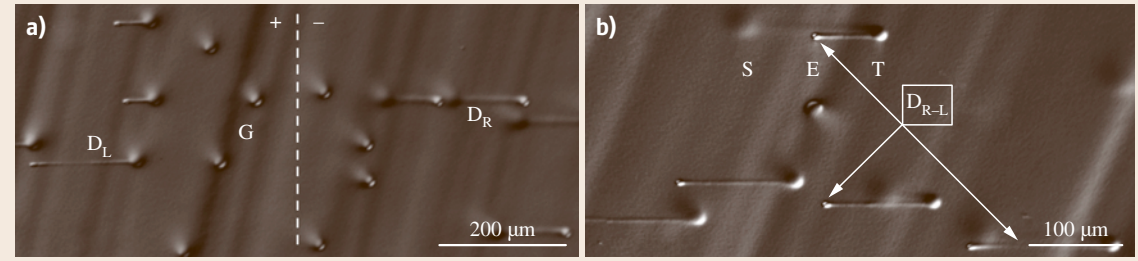


Fig. 43.24a,b DIC optical images of n-type GaAs after open-circuit photoetching

hams and Burocchi (AB) etch for GaAs is characterized by a so-called memory effect: the hillocks and ridges formed on defects remain intact even after very deep etching, even though the defects are no longer at the initial position [43.94]. On the basis of this result a projective etching method was developed and used for analysis of dislocations which moved in GaAs single crystals due to thermal stress during cooling after growth (G–S dislocations after [43.79]). This three-step etching procedure permits one to define the glide/climb system (glide/climb direction and plane) of any displaced dislocation. Numerous examples of these types of dislocations were shown and analyzed previously [43.8–10]. Figure 43.24 shows a set of images which illustrates the possibility offered by the photoetching method in analyzing the behavior of G–S dislocations. In Fig. 43.24a some dislocations moved to the left (denoted D_L) while others moved to the right (denoted D_R): most probably the strain was of the opposite sign, as indicated by "+" and "–" on the left and right side of the dashed line. The fact that in the central part of the image there are some dislocations that did not move [so-called grown-in dislocations (G), after [43.79]] supports this interpretation. In Fig. 43.24b some dislocations (denoted D_{R-L}) clearly *changed their mind* during the movement: first they moved from the starting position (S) to the right, stopped for some time (this can be recognized from the presence of the Cottrell atmosphere at the turning point (T) where the change of the direction of glide/climb occurred) and later moved to the left of the image, where they finally stopped at position E. Worth noting is the different size of the Cottrell atmospheres at positions S, T, and E, which is most probably the result of slower diffusion of decorating point defects due to decreasing temperature between the subsequent stop events of the D_{R-L} dislocations.

References

43.1 H.C. Gatos, M.C. Lavine: Characteristics of the {111} surfaces of the III-V intermetallic compounds, J. Electrochem. Soc. **107**, 427–433 (1960)

43.2 S. Amelinckx: *Direct Observation of Dislocations*, Supplement to Solid State Physics, Vol. 6 (Academic, New York 1964) pp. 15–50

43.3 D.J. Stirland, B.W. Straughan: A review of etching and defect characterization of gallium arsenide substrate material, Thin Solid Films **31**, 139–170 (1976)

43.4 D.C. Miller, G.A. Rozgonyi: Defect characterization by etching, optical microscopy and x-ray topography. In: *Handbook on Semiconductors*, Vol. 3, ed. by S.B. Keller (North-Holland, Amsterdam 1980) pp. 218–246

43.5 R.B. Heimann: Principles of chemical etching – the art and science of etching crystals. In: *Crystals: Growth, Properties and Applications*, Vol. 8, ed. by J. Grabmaier (Springer, Berlin, Heidelberg 1982) pp. 173–224

43.6 K. Sangwal: *Etching of Crystals* (North-Holland, Amsterdam 1987)

43.7 P.H.L. Notten, J.E.A.M. van den Meerakker, J.J. Kelly: *Etching of III–V Semiconductors: an Electrochemical Approach* (Elsevier, Oxford 1991)

43.8 J.L. Weyher: Characterization of compound semiconductors by etching. In: *Handbook on Semiconductors*, Vol. 3, ed. by S. Mahajan (Elsevier, Amsterdam 1994) pp. 995–1031

43.9 J.L. Weyher: Characterization of bulk as-grown and annealed III-Vs by photo-etching and complementary methods, Inst. Phys. Conf. Ser. **146**, 399–408 (1995)

43.10 J.L. Weyher, C. Frigeri, S. Müller: Selective etching and complementary microprobe techniques (SFM, EBIC). In: *Microprobe Characterization of Optoelectronic Materials*, Vol. 17, ed. by J. Jimenez (Taylor & Francis, New York 2003) pp. 595–689, Chap. 8

43.11 D. Zhuang, J.H. Edgar: Wet etching of GaN, AlN and SiC: a review, Mater. Sci. Eng. **R 48**, 1–46 (2005)

43.12 J.L. Weyher: Characterization of wide-band-gap semiconductors (GaN, SiC) by defect-selective etching and complementary methods, Superlattices Microstruct. **40**, 279–288 (2006)

43.13 J.J. Kelly, D. Vanmaekelbergh: Chemical and electrochemical etching of semiconductors. In: *Semiconductor Micromachining*, Fundamental Electrochemistry and Physics, Vol. 1, ed. by S.A. Campbell, H.J. Lewerenz (Wiley, Chichester 1997), Chap. 2

43.14 P.H.L. Notten: The etching of InP in HCl solutions: a chemical mechanism, J. Electrochem. Soc. **131**, 2641–2644 (1984)

43.15 J.J. Kelly, H.G.G. Philipsen: Anisotropy in the wet-etching of semiconductors, Curr. Opin. Solid State Mater. Sci. **9**, 84–90 (2005)

43.16 R.A. Wind, M.A. Hines: Macroscopic etch anisotropies and microscopic reaction mechanisms: using microfabrication to unravel the complicated chemistry of KOH/Si etching, J. Phys. Chem. B **106**, 1557–1569 (2002)

43.17 I.M. Huygens, K. Strubbe, W.P. Gomes: Electrochemistry and photoetching of n-GaN, J. Electrochem. Soc. **147**, 1797–1802 (2000)

43.18 A.R. de Wit, M.D. Janssen, J.J. Kelly: Electrochemical characterization of polycrystalline ZnO layers, Appl. Surf. Sci. **45**, 21–27 (1990)

43.19 A. Gagnaire, J. Joseph, A. Etcheberry, J. Gautron: An ellipsometric study of the electrochemical surface modifications of n-InP, J. Electrochem. Soc. **132**, 1655–1658 (1985)

43.20 P.H.L. Notten, J.J. Kelly: Evidence for cathodic protection of crystallographic facets from GaAs etching profiles, J. Electrochem. Soc. **134**, 444–448 (1987)

43.21 J.J. Kelly, J.E.A.M. van den Meerakker, P.H.L. Notten: Electrochemistry of photoetching and defect-revealing in III-V materials, Dechema-Monogr. **102**, 453–464 (1986)

43.22 J. van de Ven, H.J.P. Nabben: Anisotropic photoetching of III-V semiconductors I. Electrochemistry, J. Electrochem. Soc. **137**, 1603–1610 (1990)

43.23 J. van de Ven, H.J.P. Nabben: Anisotropic photoetching of III-V semiconductors II. Kinetics and structural factors, J. Electrochem. Soc. **138**, 144–152 (1991)

43.24 C. Youtsey, I. Adesida, G. Bulman: Highly anisotropic photoenhanced wet etching of n-type GaN, Appl. Phys. Lett. **71**, 2151–2153 (1997)

43.25 L. Macht, J.J. Kelly, J.L. Weyher, A. Grzegorczyk, P.K. Larsen: An electrochemical study of photoetching of heteroepitaxial GaN: kinetics and morphology, J. Cryst. Growth **273**, 347–356 (2005)

43.26 J.L. Weyher, R. Fornari, T. Görög, J.J. Kelly, B. Erné: HBr-$K_2Cr_2O_7$-H_2O etching system for indium phosphide, J. Cryst. Growth **141**, 57–67 (1994)

43.27 V. Lehmann: *Electrochemistry of Silicon. Instrumentation, Science, Materials and Applications* (Wiley-VCH, Weinheim 2002)

43.28 M.M. Faktor, J.L. Stevenson: The detection of structural defects in GaAs by electrochemical etching, J. Electrochem. Soc. **125**, 621–629 (1978)

43.29 C.R. Elliott, J.C. Regnault: The detection of strucutural defects in indium phosphide by electrochemical etching, J. Electrochem. Soc. **128**, 113–116 (1981)

43.30 A. Yamamoto, S. Tohno, C. Uemura: Detection of structural defects in n-type InP crystals by electrochemical etching under illumination, J. Electrochem. Soc. **128**, 1095–1100 (1981)

43.31 R. Bhat: Photoelectrochemical defect delineation in GaAs using hydrochloric acid, J. Electrochem. Soc. **132**, 2284–2285 (1985)

43.32 L. Blok: Characterization of vapour grown (001) $GaAs_{1-x}P_x$ layers by selective photo-etching, J. Cryst. Growth **31**, 250–255 (1975)

43.33 E. Sirtl, A. Adler: Cromasäure-Flußsäure als spezifisches System zur Ätzgrubenentwicklung auf Silizium, Z. Metallkd. **52**, 529–531 (1961), in German

43.34 J. Weyher, J. van de Ven: Selective etching and photoetching of {100} Gallium Arsenide in CrO_3-HF aqueous solutions. Part I: Influence of composition on etching behaviour, J. Cryst. Growth **63**, 285–291 (1983)

43.35 J. van de Ven, J.E.A.M. van den Meerakker, J.J. Kelly: The mechanism of GaAs etching in CrO_3-HF solutions I. Experimental results, J. Electrochem. Soc. **132**, 3020–3026 (1985)

43.36 J.J. Kelly, J. van de Ven, J.E.A.M. van den Meerakker: The mechanism of GaAs etching in CrO_3-HF solutions II. Model and discussion, J. Electrochem. Soc. **132**, 3026–3033 (1985)

43.37 J. van de Ven, J.L. Weyher, J.E.A.M. van den Meerakker, J.J. Kelly: Kinetics and morphology of GaAs etching in aqueous CrO_3-HF solutions, J. Electrochem. Soc. **133**, 799–806 (1986)

43.38 J.E.A.M. van den Meerakker, J.H.C. van Vegchel: Silicon etching in CrO_3-HF solutions. I: High [HF]/[CrO_3] ratios, J. Electrochem. Soc. **136**, 1949–1953 (1989)

43.39 J.E.A.M. van den Meerakker, J.H.C. van Vegchel: Silicon etching in CrO_3-HF solutions. II: low [HF]/[CrO_3] ratios, J. Electrochem. Soc. **136**, 1954–1957 (1989)

43.40 P. Visconti, D. Huang, M.A. Reshchikov, F. Yun, R. Cingolani, D.J. Smith, J. Jasinski, W. Swider, Z. Liliental-Weber, H. Markoç: Investigation of defects and surface polarity in GaN using hot wet etching together with microscopy and diffraction techniques, Mater. Sci. Eng. B **93**, 229–233 (2002)

43.41 J. Chen, J.F. Wang, H. Wang, J.J. Zhu, S.M. Zgang, D.G. Zhao, D.S. Jiang, H. Yang, K.H. Ploog: Measurement of threading dislocation densities in GaN by wet chemical etching, Semicond. Sci. Technol. **21**, 1229–1235 (2006)

43.42 J. Angilello, R.M. Potemski, G.R. Woolhouse: Etch pits and dislocations in {100} GaAs, J. Appl. Phys. **46**, 2315–2316 (1975)

43.43 K. Akita, T. Kusunoki, S. Komiya, T. Kotani: Observation of etch pits produced in InP by new etchants, J. Cryst. Growth **46**, 783–787 (1979)

43.44 S.N.G. Chu, C.M. Jodluk, A.A. Ballman: New dislocation etchant for InP, J. Electrochem. Soc. **129**, 352–354 (1982)

43.45 S. Komiya, T. Kotani: Direct observation of dislocations in GaAlAs-GaAs grown by the LPE method, J. Electrochem. Soc. **125**, 2019–2024 (1978)

43.46 E.S. Hellman: The polarity of GaN: a critical review, MRS Internet J. Nitride Semicond. Res. **3**, 1–11 (1998)

43.47 M. Seelmann-Eggebert, J.L. Weyher, H. Obloh, H. Zimmermann, A. Rar, S. Porowski: Polarity of (00.1) GaN epilayers grown on a (00.1) sapphire, Appl. Phys. Lett. **71**, 2635–2637 (1997)

43.48 J.L. Rouvière, J.L. Weyher, M. Seelmann-Eggebert, S. Porowski: Polarity determination for GaN films grown on (0001) sapphire and high-pressure grown GaN single crystals, Appl. Phys. Lett. **73**, 668–670 (1998)

43.49 D. Li, M. Sumiya, S. Fuke, D. Yang, D. Que, Y. Suzuki, Y. Fukuda: Selective etching of GaN polar surface in potassium hydroxide solution studied by x-ray photoelectron spectroscopy, J. Appl. Phys. **90**, 4219–4223 (2001)

43.50 G. Kamler, J. Borysiuk, J.L. Weyher, R. Czarnecki, M. Leszczynski, I. Grzegory: Selective etching and TEM study of inversion domains in Mg-doped GaN epitaxial layers, J. Cryst. Growth **282**, 45–48 (2005)

43.51 J.L. Weyher, F.D. Tichelaar, H.W. Zandbergen, L. Macht, P.R. Hageman: Selective photoetching and transmission electron microscopy studies of defects in heteroepitaxial GaN, J. Appl. Phys. **90**, 6105–6109 (2001)

43.52 L. Macht, J.L. Weyher, P.R. Hageman, M. Zielinski, P.K. Larsen: Direct influence of polarity on structural and electro-optical properties of heteroepitaxial GaN, J. Phys. Condens. Matter **14**, 13345–13350 (2002)

43.53 J.L. Weyher, L. Macht, F.D. Tichelaar, H.W. Zandbergen, P.R. Hageman, P.K. Larsen: Complementary study of defects in GaN by photoetching and TEM, Mater. Sci. Eng. B **91/92**, 280–284 (2002)

43.54 M.G. Mynbaeva, Y.V. Melnik, A.K. Kryganovskii, K.D. Mynbaev: Wet chemical etching of GaN in H_3PO_4 with Al ions, Electrochem. Sol.-State Lett. **2**, 404–406 (1999)

43.55 J.L. Weyher, M. Albrecht, T. Wosinski, G. Nowak, H.P. Strunk, S. Porowski: Study of individual grown-in and indentation-induced dislocations in GaN by defect-selective etching and transmission electron microscopy, Mater. Sci. Eng. B **80**, 318–321 (2001)

43.56 H. Ono, J. Matsui: Influence of In atoms on the shape of dislocation etch pits in LEC In-doped GaAs crystals, Jpn. J. Appl. Phys. **25**, 1481–1484 (1986)

43.57 V. Gottschalch, W. Heinig, E. Butter, H. Rosin, G. Freydank: H_3PO_4-etching of {001}-faces of InP, (GaIn)P, GaP, and Ga(AsP), Krist. Tech. **14**, 563–569 (1979)

43.58 J.L. Weyher, S. Lazar, J. Borysiuk, J. Pernot: Defect-selective etching of SiC, Phys. Status Solidi (a) **202**, 578–583 (2005)

43.59 T. Takenaka, H. Hayashi, K. Murata, T. Inoguchi: Various dislocation etch pits revealed on LPE GaAs{001} layer by molten KOH, Jpn. J. Phys. **17**, 1145–1146 (1978)

43.60 T. Ohno, H. Yamaguchi, S. Kuroda, K. Kojima, T. Suzuki, K. Arai: Direct observation of dislocations propagated from 4H-SiC substrate to epitaxial layer by x-ray topography, J. Cryst. Growth **260**, 209–216 (2004)

43.61 D. Siche, D. Klimm, T. Hölzel, A. Wohlfart: Reproducible defect etching of SiC single crystals, J. Cryst. Growth **270**, 1–6 (2004)

43.62 S.A. Sakwe, R. Müller, P.J. Wellmann: Optimization of KOH etching parameters for quantitative defect recognition in n- and p-type doped SiC, J. Cryst. Growth **289**, 520–526 (2006)

43.63 D. Hull, D.J. Bacon (Eds.): *Introduction to Dislocations* (Pergamon, Oxford 1984)

43.64 G. Kamler, J.L. Weyher, I. Grzegory, E. Jezierska, T. Wosinski: Defect-selective etching of GaN in a modified molten bases system, J. Cryst. Growth **246**, 21–24 (2002)

43.65 J.L. Weyher, P.D. Brown, J.L. Rouvière, T. Wosinski, A.R.A. Zauner, I. Grzegory: Recent advances in defect-selective etching of GaN, J. Cryst. Growth **210**, 151–156 (2000)

43.66 J.L. Weyher, S. Lazar, L. Macht, Z. Liliental-Weber, R.J. Molnar, S. Müller, V.G.M. Sivel, G. Nowak, I. Grzegory: Orthodox etching of HVPE-grown GaN, J. Cryst. Growth **305**, 384–392 (2007)

43.67 F. Secco d'Aragona: Dislocation etch for (100) planes in silicon, J. Electrochem. Soc. **119**, 948–951 (1972)

43.68 K. Motoki, T. Okahisa, S. Nakahata, N. Matsumoto, H. Kimura, H. Kasai, K. Takemoto, K. Uematsu, M. Ueno, Y. Kumagai, A. Koukitu, H. Seki: Growth and characterization of freestanding GaN substrates, J. Cryst. Growth **237–239**, 912–921 (2002)

43.69 J.L. Weyher, P.D. Brown, J.L. Rouvière, T. Wosinski, A.R.A. Zauner, I. Grzegory: Recent advances in defect-selective etching of GaN, J. Cryst. Growth **210**, 151–156 (2000)

43.70 M. Albrecht, H.P. Strunk, J.L. Weyher, I. Grzegory, S. Porowski, T. Wosinski: Carrier recombination at single dislocations in GaN measured by cathodoluminescence in a transmission electron microscope, J. Appl. Phys. **92**, 2000–2005 (2002)

43.71 K. Shiojima: Atomic force microscopy and transmission electron microscopy observations of KOH-

etched GaN surfaces, J. Vac. Sci. Technol. B **18**, 37–40 (2000)
43.72 K. Engl, M. Beer, N. Gmeinwieser, U.T. Schwarz, J. Zweck, W. Wegscheider, S. Miller, A. Miler, H.-J. Lugauer, G. Brüderl, A. Lell, V. Härle: Influence of an in situ-deposited SiN_x intermediate layer inside GaN and AlGaN layers on SiC substrates, J. Cryst. Growth **289**, 6–13 (2006)
43.73 I. Kamata, H. Tsuchida, T. Jikimoto, K. Izumi: Structural transformation of screw dislocations via thick 4H-SiC epitaxial growth, Jpn. J. Appl. Phys. **39**, 6496–6500 (2000)
43.74 S. Ha, H.J. Chung, N.T. Nuhfer, M. Skowronski: Dislocation nucleation in 4H silicon carbide epitaxy, J. Cryst. Growth **262**, 130–138 (2004)
43.75 Z. Zhang, T.S. Sudarshan: Evolution of basal plane dislocations during 4H-silicon carbide homoepitaxy, Appl. Phys. Lett. **87**, 161917–1–161917–3 (2005)
43.76 G. Kamler, J. Borysiuk, J.L. Weyher, A. Presz, M. Wozniak, I. Grzegory: Application of orthodox defect-selective etching for studying GaN single crystals, epitaxial layers and device structures, Eur. Phys. J. Appl. Phys. **27**, 247–249 (2004)
43.77 J.L. Weyher, G. Kamler, G. Nowak, J. Borysiuk, B. Lucznik, M. Krysko, I. Grzegory, S. Porowski: Defects in GaN single crystals and homo-epitaxial structures, J. Cryst. Growth **281**, 135–142 (2005)
43.78 G. Kamler, J. Smalc, M. Wozniak, J.L. Weyher, R. Czarnecki, G. Targowski, M. Leszczynski, I. Grzegory, S. Porowski: Selective etching of dislocations in violet-laser diode structures, J. Cryst. Growth **293**, 18–21 (2006)
43.79 J.L. Weyher, J. van de Ven: Selective etching and photoetching of GaAs in CrO_3-HF aqueous solutions. Part III: Interpretation of defect-related etch figures, J. Cryst. Growth **78**, 191–217 (1986)
43.80 J.L. Weyher, L. Macht: Defects in wide band-gap semiconductors: selective etching and calibration by complementary methods, Eur. Phys. J. Appl. Phys. **27**, 37–41 (2004)
43.81 C. Frigeri, J.L. Weyher: Electron beam induced current and photoetching investigations of dislocations and impurity atmospheres in n-type LEC GaAs, J. Appl. Phys. **65**, 4646–4653 (1989)
43.82 J.L. Weyher, C. Frigeri, P.J. van der Wel: Complementary DSL, EBIC and PL study of grown-in defects in Si-doped GaAs crystals grown under Ga- and As-rich conditions by LEC method, J. Cryst. Growth **103**, 46–53 (1990)
43.83 J.L. Weyher, P.J. van der Wel, C. Frigeri: Spatially resolved study of dislocations in Si-doped LEC GaAs by DSL, PL and EBIC, Semicond. Sci. Technol. **7**, A294–A299 (1992)
43.84 J.L. Weyher, L. Macht, G. Kamler, J. Borysiuk, I. Grzegory: Characterization of GaN single crystals by defect-selective etching, Phys. Status Solidi (c) **0**(3), 821–826 (2003)
43.85 J.A. Bardwell, J.B. Webb, H. Tang, J. Fraser, S. Moisa: Ultraviolet photoenhanced wet etching of GaN in $K_2S_2O_8$ solution, J. Appl. Phys. **89**, 4142–4149 (2001)
43.86 B. Yang, P. Fay: Etch rate and surface morphology controle in photoelectrochemical etching of GaN, J. Vac. Sci. Technol. B **22**, 1750–1754 (2004)
43.87 C. Youtsey, L.T. Romano, I. Adesida: Gallium nitride whiskers formation by selective photoenhanced wet etching of dislocations, Appl. Phys. Lett. **73**, 797–799 (1998)
43.88 C. Youtsey, L.T. Romano, R.J. Molnar, I. Adesida: Rapid evaluation of dislocation densities in n-type GaN films using photoenhanced wet etching, Appl. Phys. Lett. **74**, 3537–3539 (1999)
43.89 S. Lazar, J.L. Weyher, L. Macht, F.D. Tichelaar, H.W. Zandbergen: Nanopipes in GaN: photoetching and TEM study, Eur. Phys. J. Appl. Phys. **27**, 275–278 (2004)
43.90 C. Frigeri, J.L. Weyher, L. Zanotti: Study of segregation inhomogeneities in GaAs by means of DSL photoetching and EBIC measurements, J. Electrochem. Soc. **136**, 262–266 (1989)
43.91 M.L. Young, D.R. Wight: Concentration dependence of the minority carrier diffusion length and lifetime in GaP, J. Phys. D Appl. Phys. **7**, 1824–1837 (1974)
43.92 J.L. Weyher, R. Lewandowska, L. Macht, B. Lucznik, I. Grzegory: Etching, Raman and PL study of thick HVPE-grown GaN, Mater. Sci. Semicond. Process. **9**, 175–179 (2006)
43.93 R. Lewandowska, J.L. Weyher, J.J. Kelly, L. Konczewicz, B. Lucznik: Calibration of the PEC etching of GaN by Raman spectroscopy, J. Cryst. Growth **307**, 298–301 (2007)
43.94 D.J. Stirland, R. Ogden: A dislocation "etch-memory" effect in gallium arsenide, Phys. Status Solidi (a) **17**, K1–K4 (1973)

44. Transmission Electron Microscopy Characterization of Crystals

Jie Bai, Shixin Wang, Lu-Min Wang, Michael Dudley

Since the first observation of dislocations published in 1956, transmission electron microscopy (TEM) has become an indispensable technique for materials research. TEM not only provides very high spatial resolution for the characterization of microstructure and microchemistry but also elucidating the mechanisms controlling materials properties. The results of TEM analyses can also shed light on possible ways for improving the crystal quality. With the recent development of the electron exit wave reconstruction technique, the resolution of TEM has exceeded the typical Scherzer point resolution of ≈ 0.18 nm and observation of dislocation cores with an accuracy of 10 pm has been achieved. Most TEM studies are carried out in a static status; however, dynamic studies using in situ heating, in situ stressing, and even in situ growth can be conducted to study the development, interaction, and multiplication of defects.

44.1 **Theoretical Basis of TEM Characterization of Defects** 1477
44.1.1 Imaging of Crystal Defects Using Diffraction Contrast 1478
44.1.2 Phase-Contrast High-Resolution Transmission Electron Microscopy (HRTEM) 1482
44.1.3 Diffraction Techniques 1484
44.1.4 STEM, EELS, and EFTEM in Microanalysis 1489
44.1.5 FIB for TEM Sample Preparation 1493
44.2 **Selected Examples of Application of TEM to Semiconductor Systems** 1493
44.2.1 Studies of Conventional Heteroepitaxial Semiconductor Systems 1494
44.2.2 TEM Studies of Large-Mismatch Heteroepitaxial Systems 1500
44.2.3 Application of STEM, EELS, and EFTEM 1509
44.3 **Concluding Remarks: Current Application Status and Development** 1514
References 1515

44.1 Theoretical Basis of TEM Characterization of Defects

Modern TEM has three basic operation modes: imaging, diffraction, and spectroscopy analyses. In a transmission electron microscope, the objective lens takes the electrons emerging from the exit surface of the specimen, disperses them to create a diffraction pattern in the back focal plane, and recombines them to form an image in the image plane. When the lenses of the imaging system are adjusted so that the back focal plane of the objective lens coincides with the object plane of the intermediate lens, the diffraction pattern is projected onto the viewing screen. Alternatively, if the image plane of the objective lens works as the object plane of the intermediate lens, an image will be projected onto the viewing screen.

The contrast in the images recorded in imaging mode can be formed by several mechanisms:

1. Amplitude contrast (generally called diffraction contrast). The most commonly used bright field (BF) and dark field (DF) imaging techniques form images of dislocations with either the transmitted or diffracted beam under a two-beam condition. By excluding other beams, such images are formed by amplitude contrast.
2. Phase contrast and Z-contrast. When the transmitted and diffracted beams are made to combine, preserving both amplitude and phase information, a lattice image of the planes that are diffracting or even struc-

tural images of the individual atom columns may be resolved directly (high-resolution TEM, HRTEM).

3. Mass thickness contrast. Incoherent (Rutherford) elastic scatter of electrons can form mass thickness contrast. Such contrast is generally weaker and overshadowed by the stronger effects of electron diffraction, except in cases where there are large differences in atomic number or when diffraction is weak. The investigation of microstructures of crystals mainly includes defects and interfaces. The imaging and analysis of defects are generally carried out with diffraction contrast under two-beam conditions while studies of interfaces are generally conducted under multibeam phase-contrast mode (HRTEM).

In diffraction mode, patterns such as selected-area diffraction patterns, Kikuchi patterns, and convergent beam diffraction patterns can provide crystallographic information such as the orientation, crystallographic symmetry, phase, strain, etc. When a nearly parallel electron beam illuminates the specimen and a particular area in the first image is selected by an aperture, a selected-area diffraction (SAD) pattern is formed; otherwise, if the beam converges onto a small area of the specimen, convergent-beam electron diffraction (CBED) consisting of diffraction discs will occur. On the other hand, large-angle convergent-beam diffraction (LACBED) is a relatively newly established technique which is extensively applied to the analysis of various defects. A LACBED image is formed by moving the specimen out of the object plane of a CBED setting and forming a combined image from both real space (the shadowed specimen image) and reciprocal space (diffracted Bragg lines).

The spectroscopy analyses in a modern analytical TEM include energy-dispersive x-ray spectroscopy (EDS) and electron energy-loss spectroscopy (EELS) that allow quick analysis of material chemistry on the nanoscale. With the newly developed energy-filtered TEM (EFTEM), quick elemental mapping not relying on the scanning technique is possible.

44.1.1 Imaging of Crystal Defects Using Diffraction Contrast

Defects in crystals can be described in terms of translational vectors which represent displacements of atoms from their regular positions in the lattice. Assuming the general displacement vector is $\boldsymbol{R}$, the Howie–Whelan equations describing the change in amplitude of the direct beam ϕ_0 and the amplitude of the diffracted beam ϕ_g can be written as [44.1]

$$\frac{d\psi_0}{dz} = \frac{i\pi}{\xi_0}\phi_0 + \frac{i\pi}{\xi_g}\phi_g \exp(2\pi i \boldsymbol{s} z + 2\pi i \boldsymbol{g}\cdot\boldsymbol{R}),$$

$$\frac{d\psi_g}{dz} = \frac{i\pi}{\xi_g}\psi_0 \exp(-2\pi i \boldsymbol{s} z - 2\pi i \boldsymbol{g}\cdot\boldsymbol{R}) + \frac{i\pi}{\xi_g}\psi_g ,$$

where ξ_0 and ξ_g are the extinction coefficients for the direct and diffracted beam, s is the deviation vector, z is the depth of the defect in the specimen, and $\boldsymbol{g}$ is the reflecting vector.

When an objective aperture is used to exclude either the diffracted electrons or transmitted electrons under a two-beam condition, a bright-field (BF) or dark-field (DF) image is formed. By excluding other beams, such images are formed by amplitude contrast. The contrast can be used to determine the displacement field of the defect. Diffraction contrast and the appearance of features in BF and DF images depend sensitively on the deviation from the Bragg condition (deviation parameter s). Maximum transmission occurs in BF when s is small and positive, and it is under these conditions that most BF images are usually obtained [44.2].

Defects Characterized with Diffraction Contrast

Dislocations. In practical Burgers vector analysis of dislocations, the sample is tilted to a particular two-beam position with the deviation parameter s being set to a positive value. The presence of the dislocation bends the planes on one side into Bragg orientation and forms bright–dark line pair contrast in BF or DF images. A conventional technique to study the Burgers vector is $\boldsymbol{g}\cdot\boldsymbol{b}$ analysis ($\boldsymbol{g}$ is the reflecting plane while $\boldsymbol{b}$ is the Burgers vector of the dislocation) using their null-contrast properties. For dislocations with different nature, the $\boldsymbol{g}\cdot\boldsymbol{b}$ null-contrast rule works differently. Pure screw dislocations fully lose their contrast when the reflection plane satisfies $\boldsymbol{g}\cdot\boldsymbol{b} = 0$. For pure edge dislocations, $\boldsymbol{g}\cdot\boldsymbol{b} = 0$ and $\boldsymbol{g}\cdot(\boldsymbol{b}\hat{}\boldsymbol{u}) = 0$ should be simultaneously satisfied to make the dislocation invisible. For mixed-type dislocations, $\boldsymbol{g}\cdot\boldsymbol{b}$, $\boldsymbol{g}\cdot\boldsymbol{b}_e$, and $\boldsymbol{g}\cdot\boldsymbol{b}_e\hat{}\boldsymbol{u}$ are all required to be zero to minimize contrast (it is not possible to make it completely invisible). In practice, dislocations generally show faint contrast when $\boldsymbol{g}\cdot\boldsymbol{b} = 0$ even if $\boldsymbol{g}\cdot\boldsymbol{b}_e$ and $\boldsymbol{g}\cdot\boldsymbol{b}_e\hat{}\boldsymbol{u}$ are not zero. Thus, the conventional method to determine the Burgers vector of dislocations is by finding two reflections $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$ for which the invisibility criterion holds, and the Burgers vector may be determined using: $(\boldsymbol{g}_1\hat{}\boldsymbol{g}_2) \parallel \boldsymbol{b}$. It is recommended that at least three consistent cases of effective invisibility are found with $w < 1.0$ (where

w is the parameter describing the deviation from the Bragg condition $w = \xi_g s$) and using low-index reflections to avoid confusion. A detailed description of dislocation analysis is presented by *Edington* [44.3]. For strongly anisotropic materials, computer simulation must be used to determine both the direction and magnitude of the Burgers vector by employing variation of $\boldsymbol{g}, \boldsymbol{s}, \boldsymbol{u}, \boldsymbol{b}$ and the foil normal.

Stacking Faults. A stacking fault is a planar defect at which the regular stacking sequence of the crystal is locally interrupted and a relative shift $\boldsymbol{R}$ of the top part of the crystal with respect to the lower part is introduced. For a column passing through the faults, a phase shift $2\pi \boldsymbol{g} \cdot \boldsymbol{R}$ is produced between the waves emitted by the two parts of the crystal located on each side of the fault. Therefore, a phase factor $\alpha = 2\pi \boldsymbol{g} \cdot \boldsymbol{R}$ is introduced into the main and diffracted beam amplitudes. When the fault is inclined to the specimen surface, contrast takes the form of light and dark fringes parallel to the line of intersection of the fault plane with the surface. Thus the observation and analysis of fringes formed by stacking faults inclined to the TEM foil normal can be used to determine the fault vector. This is the so-called $\boldsymbol{g} \cdot \boldsymbol{R}$ analysis used to characterize stacking faults. When $\boldsymbol{g} \cdot \boldsymbol{R}$ is zero or an integer, no contrast is presented and the fault is invisible. For other values of $\boldsymbol{g} \cdot \boldsymbol{R}$, however, the fault produces contrast. The precise form of the fringe contrast depends on the diffraction conditions employed, and this enables fault vector and type to be determined [44.5].

Polarity. Polarity is a consequence of the noncentrosymmetrical structure frequently encountered in compound semiconductors. Polarity reversal and symmetry of both (0001) wurtzite-type and (111) zincblende-type structures have been studied extensively [44.6]. One of the techniques to characterize polarity is the multiple dark-field TEM technique. *Serneels* et al. [44.7] conducted the theoretical calculations and predicted that, under multiple-beam conditions along a noncentrosymmetric zone axis, the inverted region of the crystal should be different in brightness from the surrounding matrix. The difference arises from the violation of Friedel's law. Taking GaN/AlN as an example, when multiple dark-field images are taken along a nonsymmetrical zone axis (say, $\langle 11\bar{2}0 \rangle$ or $\langle 10\bar{1}0 \rangle$) with $\boldsymbol{g} = 0002$ and $\boldsymbol{g} = 000\bar{2}$, an inversion domain will show either brighter or darker contrast than the matrix material [44.8]. *Jasinski* et al. applied this method to characterize the V-like inversion domains in InN films [44.9]. Dark-field images recorded with (0002) and $(000\bar{2})$ reflections under multiple diffraction conditions show a reversal defect-matrix contrast, as illustrated in Fig. 44.1. This method does not determine if the domain has Ga or N-polarity, but it may prove the presence of inversion.

High-Resolution Diffraction Contrast Imaging

For the characterization of crystal defects, the geometry and the character of the individual defects are two critical features of interest. In many studies, such as those involving the dissociation of dislocations, defects of high density without strain field overlap, defects with very small size etc., examination of images with high resolution is required. Apart from instrumental limits, the resolution is also determined by factors such as diffraction contrast. Two ways of achieving high-resolution diffraction contrast are described below.

Weak-Beam Dark-Field (WBDF) Technique. Since its first report by *Cockayne* et al. [44.4], the weak-beam dark-field (WBDF) technique has become a convenient and important method to achieve high-resolution diffraction contrast images of crystal defects. In par-

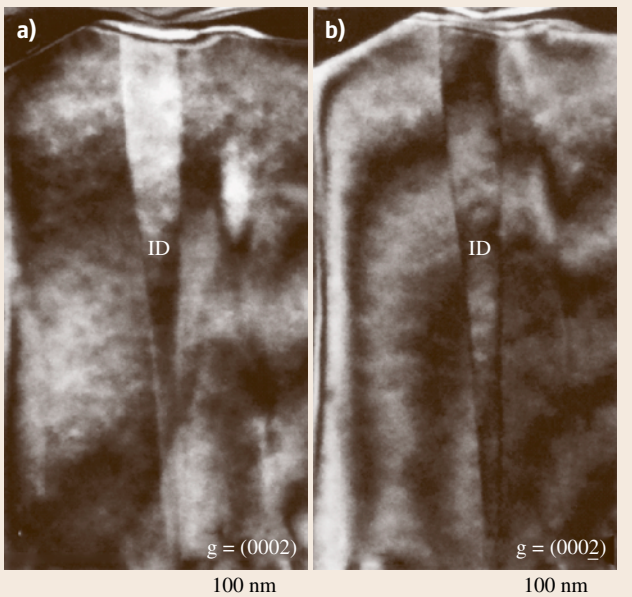


Fig. 44.1a,b Two-beam dark-field TEM images of an inversion domain in InN film ((**a**) recorded with $\boldsymbol{g} = 0002$; (**b**) recorded with $\boldsymbol{g} = 000\bar{2}$). A domain-matrix contrast reversal can be noticed (after [44.4], © AIP 1969)

ticular, it is widely employed to get sharp images of dislocation lines, resolving pairs of dislocations, and in precisely locating the positions of dislocation cores combined with simulation. The major advantage of the weak-beam approach over the two-beam dynamical technique lies in its improved resolution and the high contrast of the image. With the WBDF technique, only the diffraction contrast from the core of the defect contributes to the image. Consequently, systematic analysis of a given crystal defect is possible at a resolution approaching the limit of the microscope. Though the resolution of WBDF is not as high as the direct lattice resolution technique, it does have additional advantages such as allowing thicker specimens, reduced requirement of lens aberration, better contrast, and the validity of $\boldsymbol{g}\cdot\boldsymbol{b}$ analysis. In this technique, the TEM sample is tilted to a large, positive value of s, so that only the crystal planes close to the dislocation core are bent into a diffraction condition with $s \approx 0$ while neighboring lattice planes are away from the Bragg condition. By doing this, a sharp image of the near-core region of the dislocation can be recorded.

High-Order Reflection Method. The image width of a defect is approximately $\zeta_g/3$, where ζ_g is the extinction distance for the reflection $\boldsymbol{g}$ [44.10]. Thus the image can be made narrower and the resolution increased by using high-order reflection with long extinction length. This technique was introduced by *Bell* and *Thomas* and is especially applicable at high voltages [44.11]. In this technique the specimen is tilted to put the high-order reflection into strong diffracting condition and form the image with transmitted beam. Compared with WBDF, this technique allows shorter exposure time and hence minimizes the risk of losing resolution because of mechanical or other instabilities. The disadvantage is that the image resolution is improved at the cost of contrast. If reflections of too high an order are employed, the defect will no longer be visible.

Quantitative Determination of the Indices of Line Features

Determination of the crystallographic direction of linear defects such as straight dislocation lines, needle-shaped precipitates, etc. can be very important when trying to understand their formation mechanisms. In a similar way, knowing the direction of the lines of intersection between planar defects or between inclined planar defects and the sample surface can also help in understanding their origins. For the case of a dislocation, in order to fully characterize this defect, both its line direction and Burgers vector (magnitude and direction) need to be determined. TEM can be conveniently used to determine the line direction by tilting the specimen to at least two different orientations and measuring the angle, on the given micrograph, between the direction normal to the dislocation line direction and the trace of the reflecting plane (normal to $\boldsymbol{g}$). Stereographic projection analysis can then be used to deduce the line direction $\boldsymbol{u}$ from these measurements as described by *Edington* [44.12]. This simply involves plotting the two (or three) normals to the dislocation line direction on a stereographic projection and the subsequent drawing of a great circle through these normals. The pole of this great circle is then the line direction of the dislocation in the crystal. Figure 44.2 is a schematic showing this stereographic analysis. The various points are plotted on a standard projection.

The dashed line shows the great circle corresponding to the trace of the plane perpendicular to the incident beam direction, $\boldsymbol{B}_1\cdot\boldsymbol{g}_1$ is the $\boldsymbol{g}$ vector used for this particular image, $\boldsymbol{g}_1\times\boldsymbol{B}_1$ is the intersection of the trace of the reflecting plane on the image plane, and $\boldsymbol{D}_1$ is

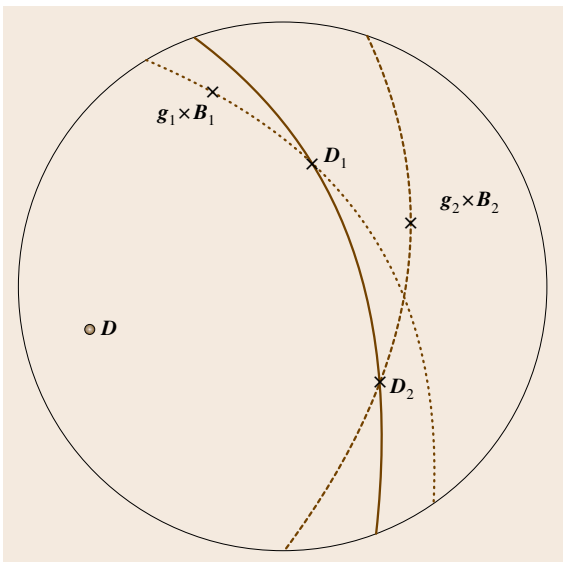


Fig. 44.2 Schematic showing the stereographic analysis. $\boldsymbol{g}_1\times\boldsymbol{B}_1$ and $\boldsymbol{g}_2\times\boldsymbol{B}_2$ are the intersections between the traces of the $\boldsymbol{g}$ vectors utilized with the respective image planes (planes perpendicular to $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$); $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ are the intersections between the traces of the planes perpendicular to the dislocation line direction observed on the two images. $\boldsymbol{D}$ is the pole of the great circle containing $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$

the intersection of the trace on the image plane of the plane normal with the dislocation line. δ_1 is the angle between $\boldsymbol{g}_1 \times \boldsymbol{B}_1$ and $\boldsymbol{D}_1$ measured from the TEM image and plotted on the great circle corresponding to the trace of the plane normal to $\boldsymbol{B}_1$. The dotted line is the great circle corresponding to the trace of the plane normal to another incident beam direction $\boldsymbol{B}_2$, with $\boldsymbol{g}_2$ being the reference $\boldsymbol{g}$ vector and $\boldsymbol{D}_2$ being another observed direction normal to the dislocation line which is oriented at the measured angle δ_2 from $\boldsymbol{g}_2 \times \boldsymbol{B}_2$. The solid line is the great circle that runs through $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$; the pole of this great circle is the dislocation line direction $\boldsymbol{D}$.

Unfortunately, such graphical techniques can be tedious, inconvenient, and imprecise. *Bai* [44.13] introduced an analytical, vector version of this technique wherein the trace of the $\boldsymbol{g}$ vector on a given image is expressed as the vector $\boldsymbol{g} \times \boldsymbol{B}$. With suitably designed coordinate systems being defined, measurement of the angle (δ) between ($\boldsymbol{g} \times \boldsymbol{B}$) and the normal to the dislocation line direction enables it to be expressed as a vector $\boldsymbol{D}_\mathrm{n}$. Repeating this measurement for a second image with different $\boldsymbol{g}$ vector and beam direction $\boldsymbol{B}$ generates another vector $\boldsymbol{D}_\mathrm{n}$ and the cross-product of two such vectors enables the line direction of the dislocation in the crystal to be determined.

The three coordinate systems utilized in the calculation (shown in Fig. 44.3) are defined as follows [44.14]:

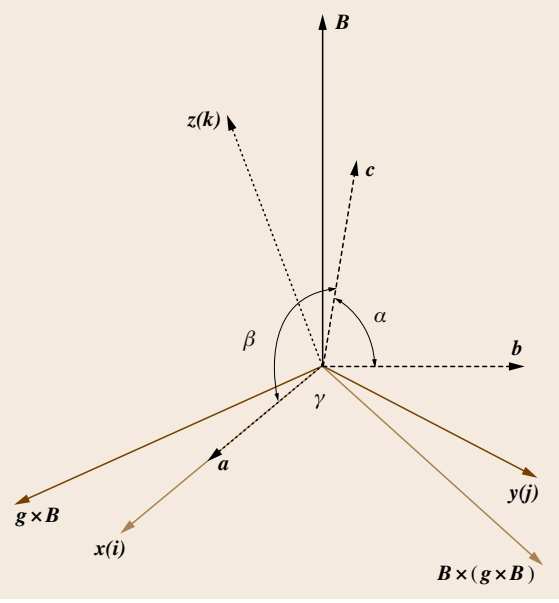


Fig. 44.3 Coordinate systems for the vector analysis of the determination of the line direction of a linear feature

1. A Cartesian coordinate system consisting of the beam direction ($\boldsymbol{B}$), $\boldsymbol{g} \times \boldsymbol{B}$, and $\boldsymbol{B} \times (\boldsymbol{g} \times \boldsymbol{B})$
2. The crystal coordinate system comprising the three basis vectors $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ and three angular parameters α, β, γ of the unit cell for the crystal being studied
3. A Cartesian coordinate system $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{z}$, in which the axis $\boldsymbol{x}$ is along $\boldsymbol{a}$, $\boldsymbol{b}$ lies in the $\boldsymbol{xy}$ plane, and $\boldsymbol{c}$ is in the upper half-space (since $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ roughly form a right-handed relationship).

The normal of the line feature which lies at an angle δ to the trace of the $\boldsymbol{g}$ vector can be represented as $\boldsymbol{D}_\mathrm{n} = (\boldsymbol{g} \times \boldsymbol{B}/|\boldsymbol{g} \times \boldsymbol{B}|) \cos\delta + [\boldsymbol{B} \times (\boldsymbol{g} \times \boldsymbol{B})/|\boldsymbol{B} \times (\boldsymbol{g} \times \boldsymbol{B})|] \sin\delta$. $\boldsymbol{B}$ and $\boldsymbol{g}$ are expressed as vectors in crystal system and therefore are the vectors $\boldsymbol{g} \times \boldsymbol{B}$ and $\boldsymbol{B} \times (\boldsymbol{g} \times \boldsymbol{B})$. All of these vectors then have to be expressed as vectors referred to the Cartesian coordinate system $\boldsymbol{xyz}$ in order to carry out the necessary vector calculations.

The expressions enabling the basis vectors $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ to be referred to the Cartesian coordinate system $\boldsymbol{xyz}$ are

$$
\begin{aligned}
&\boldsymbol{a} = a_x\boldsymbol{i} + a_y\boldsymbol{j} + a_z\boldsymbol{k} = a\boldsymbol{i}\,,\\
&a_x = a\,, \quad a_y = a_z = 0\,,\\
&\boldsymbol{b} = b_x\boldsymbol{i} + b_y\boldsymbol{j} + b_z\boldsymbol{k} = b(\cos\gamma\boldsymbol{i} + \sin\gamma\boldsymbol{j})\,,\\
&b_x = b\cos\gamma\,, \quad b_y = b\sin\gamma\,, \quad b_z = 0\,,\\
&\boldsymbol{c} = c_x\boldsymbol{i} + c_y\boldsymbol{j} + c_z\boldsymbol{k}\,,\\
&c_x = c\cos\beta\,, \quad c_y = c\frac{(\cos\alpha - \cos\beta\cos\gamma)}{\sin\gamma}\,,\\
&c_z = \sqrt{c^2 - c_x^2 - c_y^2}\,,\\
&V_\mathrm{c} = (\boldsymbol{a} \times \boldsymbol{b}) \cdot \boldsymbol{c} = abc_z \sin\gamma\,,\\
&a_x^* = \frac{b_y c_z}{V_\mathrm{c}}\,, \quad a_y^* = -\frac{b_x c_z}{V_\mathrm{c}}\,, \quad a_z^* = (b_x c_y - b_y c_x)\,,\\
&b_y^* = \frac{a_x c_z}{V_\mathrm{c}}\,, \quad b_z^* = -\frac{a_x c_y}{V_\mathrm{c}}\,,\\
&c_z^* = \frac{a_x b_y}{V_\mathrm{c}}\,,\\
&b_x^* = c_x^* = c_y^* = 0\,.
\end{aligned}
$$

$\boldsymbol{B}_1(U_1, V_1, W_1)$ and $\boldsymbol{g}_1(h_1, k_1, l_1)$ can be expressed in the $\boldsymbol{xyz}$ system as

$$
\begin{aligned}
&\boldsymbol{B}_1 = B_x^1\boldsymbol{i} + B_y^1\boldsymbol{j} + B_z^1\boldsymbol{k}\,,\\
&B_x^1 = U_1 a_x + V_1 b_x + W_1 c_x\,, \quad B_y^1 = V_1 b_y + W_1 c_y\,,\\
&B_z^1 = W_1 c_z\,,\\
&\boldsymbol{g}_1 = g_x^1\boldsymbol{i} + g_y^1\boldsymbol{j} + g_z^1\boldsymbol{k}\,,\\
&g_x^1 = h_1 a_x^*\,, \quad g_y^1 = h_1 a_y^* + k_1 b^* y\,,\\
&g_z^1 = h_1 a_z^* + k_1 b_z^* + l_1 c_z^*\,.
\end{aligned}
$$

Then $\boldsymbol{g}_1 \times \boldsymbol{B}_1$ and $\boldsymbol{B}_1 \times (\boldsymbol{g}_1 \times \boldsymbol{B}_1)$ can be written as

$$
\begin{aligned}
(\boldsymbol{g}_1 \times \boldsymbol{B}_1)_x &= g_y^1 B_z^1 - g_z^1 B_y^1 \,, \\
(\boldsymbol{g}_1 \times \boldsymbol{B}_1)_y &= g_z^1 B_x^1 - g_x^1 B_z^1 \,, \\
(\boldsymbol{g}_1 \times \boldsymbol{B}_1)_z &= g_x^1 B_y^1 - g_y^1 B_x^1 \,, \\
[\boldsymbol{B}_1 \times (\boldsymbol{g}_1 \times \boldsymbol{B}_1)]_x &= B_y^1 (\boldsymbol{g}_1 \times \boldsymbol{B}_1)_z - B_z^1 (\boldsymbol{g}_1 \times \boldsymbol{B}_1)_y \,, \\
[\boldsymbol{B}_1 \times (\boldsymbol{g}_1 \times \boldsymbol{B}_1)]_y &= B_z^1 (\boldsymbol{g}_1 \times \boldsymbol{B}_1)_x - B_x^1 (\boldsymbol{g}_1 \times \boldsymbol{B}_1)_z \,, \\
[\boldsymbol{B}_1 \times (\boldsymbol{g}_1 \times \boldsymbol{B}_1)]_x &= B_x^1 (\boldsymbol{g}_1 \times \boldsymbol{B}_1)_y - B_y^1 (\boldsymbol{g}_1 \times \boldsymbol{B}_1)_x \,.
\end{aligned}
$$

The normal of the dislocation line $\boldsymbol{D}_1$ in this case can be written as

$$
\begin{aligned}
\boldsymbol{D}_1 &= D_x^1 \boldsymbol{i} + D_y^1 \boldsymbol{j} + D_z^1 \boldsymbol{k} \,, \\
D_x^1 &= \cos\delta_1 (\boldsymbol{g}_1 \times \boldsymbol{B}_1)_x + \sin\delta_1 [\boldsymbol{B}_1 \times (\boldsymbol{g}_1 \times \boldsymbol{B}_1)]_x \,, \\
D_y^1 &= \cos\delta_1 (\boldsymbol{g}_1 \times \boldsymbol{B}_1)_y + \sin\delta_1 [\boldsymbol{B}_1 \times (\boldsymbol{g}_1 \times \boldsymbol{B}_1)]_y \,, \\
D_x^1 &= \cos\delta_1 (\boldsymbol{g}_1 \times \boldsymbol{B}_1)_z + \sin\delta_1 [\boldsymbol{B}_1 \times (\boldsymbol{g}_1 \times \boldsymbol{B}_1)]_z \,.
\end{aligned}
$$

Similarly, we can get $\boldsymbol{D}_2 = D_x^2 \boldsymbol{i} + D_y^2 \boldsymbol{j} + D_z^2 \boldsymbol{k}$. Then the dislocation line $\boldsymbol{D} = \boldsymbol{D}_1 \times \boldsymbol{D}_2$ can be written as

$$
\begin{aligned}
\boldsymbol{D} &= D_x \boldsymbol{i} + D_y \boldsymbol{j} + D_z \boldsymbol{k} \,, \\
D_x &= D_y^1 D_z^2 - D_z^1 D_y^2 \,, \\
D_y &= D_z^1 D_x^2 - D_x^1 D_z^2 \,, \\
D_z &= D_x^1 D_y^2 - D_y^1 D_x^2 \,.
\end{aligned}
$$

Now the index of $\boldsymbol{D}(D_x, D_y, D_z)$ is in the Cartesian coordinate system $\boldsymbol{xyz}$. We can convert this to a vector referred to the crystal system using the following relationships

$$
\begin{aligned}
\boldsymbol{U} &= D_x \boldsymbol{p} + D_y \boldsymbol{p}' + D_z \boldsymbol{p}'' \,, \\
\boldsymbol{V} &= D_y \boldsymbol{q}' + D_z \boldsymbol{q}'' \,, \\
\boldsymbol{W} &= D_z \boldsymbol{r}'' \,,
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{p} &= \frac{1}{a} \,, \\
\boldsymbol{p}' &= -\frac{b_x}{a b_y} \,, \\
\boldsymbol{p}'' &= -\frac{1}{a}\left[-\frac{c_y b_x}{b_y c_z} + \frac{c_y}{c_z}\right] \,, \\
\boldsymbol{q}' &= \frac{1}{b_y} \,, \\
\boldsymbol{q}'' &= -\frac{c_y}{b_y c_z} \,, \\
\boldsymbol{r}'' &= \frac{1}{c_z} \,.
\end{aligned}
$$

The main source of error in this calculation relates to the accuracy of the determination of $\boldsymbol{B}$, the measurement of the angle δ, and the integralization of the decimal indices.

Generally, a $\boldsymbol{g}$ vector in a low-index selected-area diffraction pattern with $\boldsymbol{B}$ as zone axis is used in this process. The error in determination of $\boldsymbol{B}$ and $\boldsymbol{g}$ and hence δ can be avoided by carefully adjusting the zone axis $\boldsymbol{B}$ along the electron beam. The 180° ambiguity does not create a problem because the coordinates $\boldsymbol{g} \times \boldsymbol{B}$ and $\boldsymbol{B} \times (\boldsymbol{g} \times \boldsymbol{B})$ go with the assigned $\boldsymbol{g}$. The critical point is the sense of δ. If $\boldsymbol{D}_\mathrm{n}$ is anticlockwise from $\boldsymbol{g} \times \boldsymbol{B}$, care must be taken regarding the sense of rotation sense of δ with respect to the trace of the known $\boldsymbol{g}$. An accurate calibration of the rotation between the image and the diffraction pattern is required for the application of this technique.

44.1.2 Phase-Contrast High-Resolution Transmission Electron Microscopy (HRTEM)

For dislocations in semiconductor materials, the individual dislocations exhibit significant deviations from average predictions due to various reasons and exhibit an unusual large data scattering that is commonly of unknown origin. Therefore, it is desirable to characterize individual dislocations with truly atomic resolution and to compare the results with theoretical calculations. Phase-contrast high-resolution TEM to date has broken the 0.1 nm barrier and it is possible to combine theory and experiment to accurately identify atomic column positions with better than 10 pm precision [44.15].

As opposed to the amplitude caused contrast observed in conventional TEM (where the image is formed by one beam), HRTEM images are formed with combined transmission and diffracted beams and the contrast is therefore a composite composed of both amplitude and phase information. There is no relative displacement between the location of a defect and the contrast variation caused by the defect in the HRTEM image, which is an advantage compared with diffraction imaging. However, there are limitations in direct interpretation of the HRTEM image in terms of the sample structure and composition. Since HRTEM images are sensitive to factors such as specimen thickness and orientation, objective lens defocus, spherical and chromatic aberration, etc., precise interpretation of the images may require extensive simulations. The effect of the interference between the beams can be predicted with the phase contrast transfer function (CTF) of the

objective lens and therefore a HRTEM image can be simulated. Since the CTF is focusing dependent, it is crucial to choose the optimum defocus to fully exploit the capabilities of electron microscopy in HRTEM mode. However, there is no simple optimized solution. Better experimental results can be achieved by conducting a specimen exit wave reconstruction with a series of through-focal images [44.16]. In this technique, a series of about 20 pictures is shot under the same imaging conditions with the exception of the focus, which is incremented between each take. Together with exact knowledge of the CTF the series allows for computation of the phase change. As mentioned earlier, amplitude change also contributes to image contrast. Consequently, sample thickness plays a critical rule in the appearance of a lattice image. When the sample is thin enough that amplitude variations do not contribute to the image, the HRTEM image is formed purely by phase contrast and it shows directly a two-dimensional projection (down some low-index direction) of the crystal with defects and all.

A serious limitation to the interpretation of high-resolution images is the limited signal-to-noise ratio. By noise we refer to any contribution to image intensity which is not essentially a part of the signal, e.g., lattice image. Major contributions to noise are disordered surface layers, due to either contamination or specimen thinning damage, and shot noise in the electron beam or recording media. High noise-to-signal ratio could lead to errors in interpretation of crystal/amorphous interfaces.

There are two resolution limits in axial HRTEM which apply to imaging of the simple weak-phase object. The point-to-point resolution, attainable near the Scherzer defocus condition, is $d_{pp} \approx 0.66(C_s\lambda^3)^{1/4}$ (where C_s is the spherical aberration coefficient and λ is the electron wavelength). The information limit is $d_{in} \approx 2\sqrt{E/(C_c\lambda\Delta E)}$, where C_c is the coefficient of chromatic aberrations. For a totally unknown weak-phase object one can identify detail to d_{pp} simply and to d_{in} by careful image simulation or reconstruction from a through-focal series [44.17].

Dislocation Core Observation by HRTEM

Recent developments in electron microscopy have enabled microscopists to resolve the dislocation core structure down to the atomic level. For diamond cubic and zincblende semiconductors resolving dumbbell atom columns along the dislocation line requires resolution better than the typical Scherzer point resolution of ≈ 0.18 nm of most high-resolution transmission electron microscopes. By reconstructing an electron exit wave through-focal series, *Xu* et al. resolved a 30° partial dislocation in GaAs:Be with accuracy of the atomic position at the dislocation core region within 10 pm [44.18]. A modern aberration-corrected TEM can easily achieve point-to-point resolution of 0.1 nm.

Interfaces of Epitaxial Systems

HRTEM plays a valuable role in the characterization of interfaces in epitaxial systems. The investigation generally includes studies of the misfit dislocations, measurement of rigid shifts, and identification of the roughness and extent of the interface. The major requirement for characterization of interfaces is that both crystals should be well aligned on their appropriate zone axes and the interface should be aligned so that it is edge-on to the incident beam. The microscope should be well aligned and operate near the Scherzer focus and the thickness on both sides of the interface should be considerably less than half an extinction thickness (if possible).

For conventional epitaxial structures, e.g., SiGe on Si, low-index zones such as [100] and [110] are typically used, but [111] and [112] can provide further useful projections when three-dimensional views are required in order to specify the interface structures fully. For heterostructures of III–V compounds such as GaAs/$Al_xGa_{1-x}As$, it is extremely difficult to determine the location of the interface under the imaging conditions traditionally required for high-resolution imaging, namely, thin crystals and optimum defocus, and hence impossible to determine the interface sharpness. In order to highlight the compositional variations it is necessary to find thickness and defocus combinations which accentuate differences in the images from the two constituent materials. To get distinct interface, dark-field images formed by the 200 diffraction spot can be used to precisely determine the interface. This is due to the fact that, in the kinematical approximation, the intensity in the 200 diffraction beam is proportional to $(f_{III} - f_V)^2$, where f_{III} and f_V are the atomic scattering amplitudes of the group IIIA and group V elements, respectively. When this diffracted beam is strongly excited, the $Al_xGa_{1-x}As$ layers appear much brighter than the GaAs layers. Such images provide an accurate measure of the layer thickness and of the flatness of the interface [44.19]. Another method to get distinct interface is to use the ⟨100⟩ projection rather than the ⟨110⟩ > projection because the former has four chemically sensitive {002} beams compared with the latter with two of them [44.20]. The interfaces

of epitaxial systems can also be well studied by the high-resolution high-angle annular dark-field scanning TEM (HAADF-STEM) technique, which is extremely sensitive to the atomic number of the atoms in the high-resolution image (Sect. 44.1.4).

Polarity

Liliental-Weber et al. studied the atomic structure of Mg-rich hexagonal pyramids in GaN film which exhibit opposite polarity from the matrix material [44.21]. By taking a series of HRTEM images in a through-focal sequence, the electron wave exiting the specimen was reconstructed and atomic resolution was achieved. With such a high resolution, the exchange of Ga and N sublattices inside and outside the defect can be unambiguously determined. In this case a special crystal projection needs to be chose, e.g., [11$\bar{2}$0], such that two different atoms with different atomic numbers can be resolved. Figure 44.4 shows the reconstructed images taken from different regions related to the pyramidal defect. The inversion of the polarity within the defect (BC) compared with the matrix (AB) can be clearly observed.

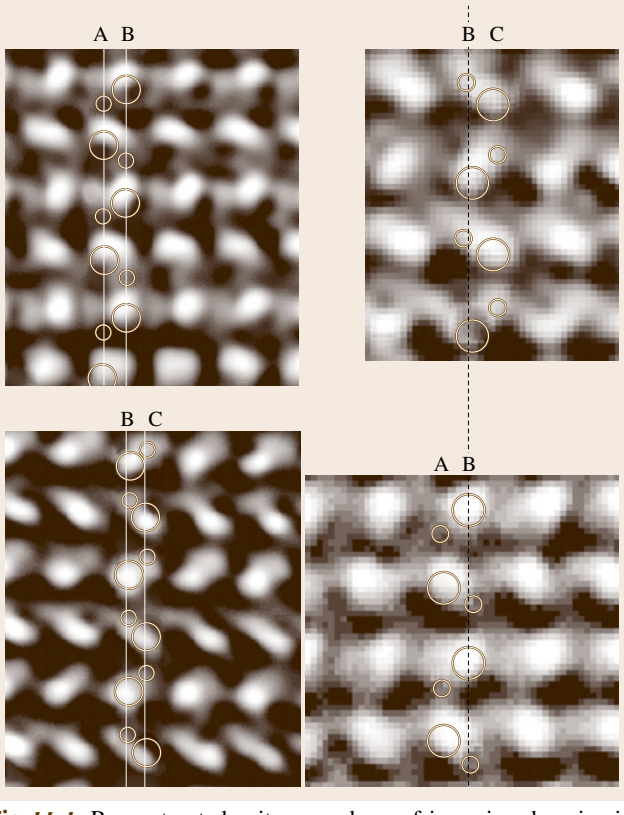


Fig. 44.4 Reconstructed exit wave phase of inversion domains in Mg-rich GaN. Note the change of polarity within the defect (BC) compared with the matrix (AB) (after [44.21], © Elsevier 2005)

44.1.3 Diffraction Techniques

As mentioned previously, a parallel electron beam gives rise to diffraction patterns composed of sharp spots. Generally such a diffraction pattern is achieved by selecting a small region using an aperture inserted in the image plane of the objective lens, forming a selected-area diffraction (SAD) pattern. When the electron beam converges to image a small area of the specimen, a CBED pattern comprising diffraction discs forms. Compared with the limited spatial resolution of conventional SAD (usually down to a fraction of a micrometer), CBED gives resolution down to nanometer. Moreover, CBED discs contain more information than simple diffraction spots in SAD. The application of CBED includes phase identification, symmetry determination (point and space group), thickness measurement, strain and lattice parameter measurement, structure factor determination, etc.

Selected-Area Diffraction (SAD)

A SAD pattern can be treated as the magnified image of a planar section through the reciprocal lattice taken normal to the incident beam direction. It can be used to:

1. Identify the orientation of the specimen
2. Find the proper $\boldsymbol{g}$ vector for defect analysis
3. Identify the structure of defects such as precipitates, twins, stacking faults, etc.
4. Determine the orientation relationship between phases
5. Determine the long-range order parameters.

The accuracy of analysis of a diffraction pattern depends upon the accuracy of measurement. Some factors of importance are:

1. The shape factor of the features, which determines the shape of the intensity distribution about the rel-points
2. Instrument alignment and beam divergence
3. Specimen perfection
4. Curvature of the reflecting sphere and relative orientation of the specimen
5. Double diffraction, giving rise to reflections of zero structure factor.

Several examples of the application of the SAD technique are described in the following.

Twins in Diamond Cubic and Zincblende Structure Semiconductor Crystals. Twinning is a phenomenon which leads to the existence of domains in a crystal which have orientation relationships usually described by a simple symmetry operation. The most commonly observed type of symmetry operation involves a mirror image of the structure in the twinning plane, although the 180° (or equivalently 60°) rotation of the structure about the normal to the twin plane is also quite common in semiconductors with diamond cubic and zincblende structure. Twins can be produced either during growth or by mechanical deformation. The existence of twins can be discerned using the SAD technique, which is sensitive to the differences in orientation between a *twin* and *matrix*. If the twinned domains lay side by side with the untwined *matrix* regions and the incident beam straddles both regions then a simple superposition of two diffraction patterns will occur, one from the matrix and one from the twin. For example, for a (110)-oriented diamond cubic crystal, a twin might be created by 180° rotation about the $(\bar{1}11)$ twin plane normal. This produces twin domains which have $(\bar{1}\bar{1}0)$ surface orientation (i. e., the same as the matrix) but which have different in-plane orientation. This is shown in the stereographic projections in Fig. 44.5, where it can be seen that the orientation relationship between the two structures can not only be considered from the point of view of the 180° rotation about the $(\bar{1}11)$ twin plane normal but can also be considered either as a mirror across $(1\bar{1}2)$ or as a rotation by 70°32′ about the (110) plane normal. The diffraction pattern associated with such side-by-side twins should therefore consist of the (110) matrix patterns with a second *twin* pattern being produced from the first by either a mirror operation across $(1\bar{1}2)$ or as a rotation by 70°32′ about the (110) zone axis. However, if the twin and matrix domains overlap one another the diffraction pattern becomes slightly more complicated, with extra spots appearing at 1/3 intervals along the ⟨111⟩ directions. Figure 44.6 shows an example SAD pattern of (110) Ge epifilm with overlapping matrix and twin domains. Extra spots located at $\boldsymbol{n}\times\boldsymbol{g}/3$ can be clearly observed. Some of these extra spots can arise from multiple diffraction effects [44.22]. However, the full pattern can be understood by considering the tripling of the periodicity which occurs in this direction when the twin and matrix structures are overlapped [44.23]. Further details regarding analysis of twins can be found in the book by *Edington* [44.24].

Atomic Ordering of Semiconductor Alloys. For alloys containing different kinds of atoms which have attractive interactions, the desire to maximize the number of

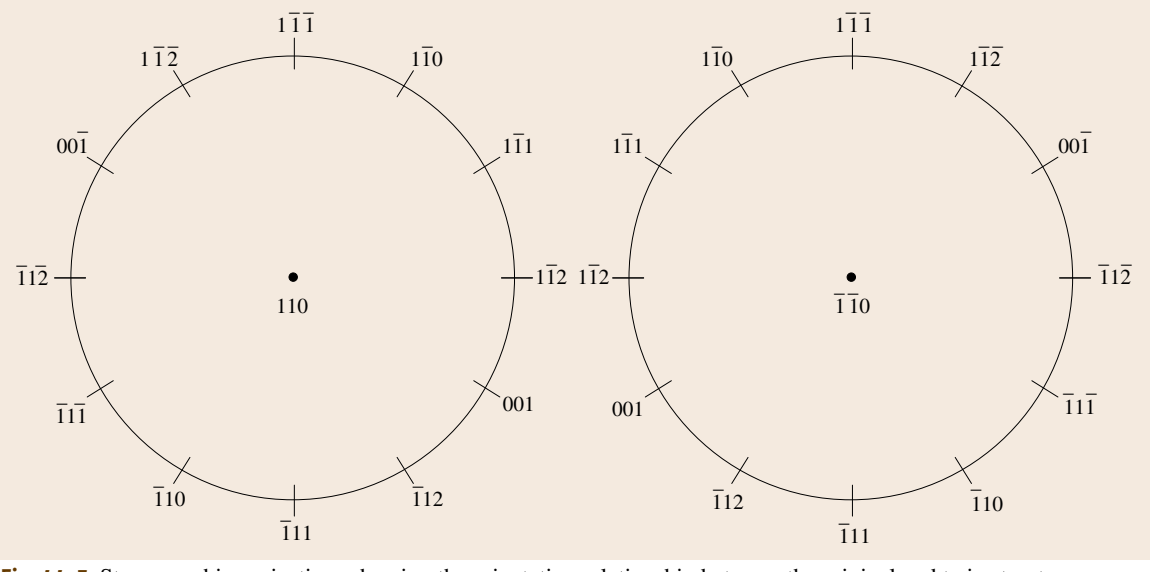


Fig. 44.5 Stereographic projections showing the orientation relationship between the original and twin structures

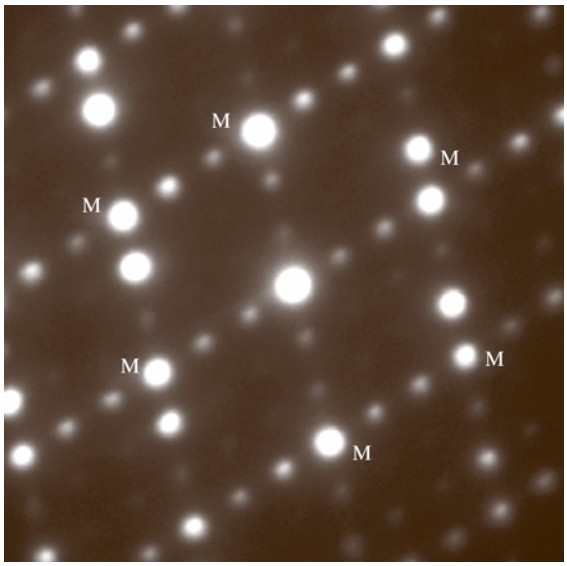


Fig. 44.6 SAD pattern of (110)Ge epifilm with overlapping matrix and twin domains, the main diffraction spots from matrix are indicated by "M"

unlike neighbors can lead to the formation of an ordered structure. Consequently, the usual structure factor rules determining the allowed reflections are relaxed and superstructure reflections in between the original Bragg reflections can be introduced. The intensity of such reflections is related to the difference between the atomic scattering factors of the atoms involved, as opposed to the intensity of fundamental reflections, which is related to their sum. A comprehensive review is given by *Marcinkowski* [44.25]. Studies of such long-range order structure can be carried out using the SAD technique to analyze the superstructure reflections. TEM specimens in various tilted orientations are required to obtain SAD patterns along different directions [44.26,27]. The extra diffraction spots are indexed and compared with those of the standard ordered structure in order to determine the long-range order. With the help of computer simulation, the structural ordering can be determined using smaller mutual tilts [44.28]. In various III–V alloy semiconductors such as AlGaAs, In(Al)GaP, InGaAsP, InAlAs, GaAsSb, GaAsP, and InAsSb, several types of ordered structures such as CuAu, CuPt, chalcopyrite, and famatinite, have been observed. In crystals grown on (001) substrates, observed ordered structure is mostly of CuPt type in which a doubling of the periodicity of the column III or V atoms occurs on the {111} plane [44.26].

Conventional Convergent-Beam Electron Diffraction (CBED)

When the incident electron waves impinging on the specimen comprise a convergent cone of rays, this generates a continuous range of Ewald spheres in reciprocal space. This causes higher-order Laue zones (HOLZs) to become visible in diffraction patterns, such that three-dimensional diffraction information is exhibited on the diffraction pattern. The fact that such information can be obtained from areas as small as a few nanometers in diameter is a major advantage of CBED. The diffraction discs in a CBED pattern are created exactly from the same area, and each disc contains the intensities determined by the orientation of the incident beam with respect to the (*hkl*) lattice plane. Several examples of the application of CBED are presented in the following sections.

Thickness Determination. CBED is a convenient method to determine the TEM specimen thickness. In this technique, the sample is tilted to form a two-beam condition. The intensity fringes (K-M fringes) in the diffracted discs due to various s are related to the sample thickness by $s_i^2/n_k^2 + 1/(\xi_g^2 n_k^2) = 1/t^2$ (where s_i is the diffraction deviation for the i-th fringe, n_k is an integer, ζ_g is the extinction length for the reflection, and t is the thickness of the sample). By choosing a proper n_k to plot a straight line of s_i^2/n_k^2 versus $1/n_k^2$, the thickness t can be determined by the intercept of the line [44.29]. For such measurement, the region of the foil selected should be flat without distortion and the beam must be focused at the plane of the specimen.

Polarity. There are several TEM methods for polarity determination: CBED, TEM-EDS, HRTEM, bend contour analysis, and EELS. Their relative capabilities to determine polarity depend strongly upon the species of crystal and the morphology of specimens to be examined. Since the CBED pattern is sensitive to both defects sample thickness in the illuminated area of the specimen, the use of this method is sometimes limited. *Mitate* et al. studied the polarity of GaN, ZnO, AlN and GaAs with both TEM-EDS (energy-dispersive x-ray spectroscopy) and the conventional CBED methods [44.30]. The results show that the CBED method is useful for ZnO and GaN, while the TEM-EDS method must be used for AlN and GaAs and can also be used for GaN and ZnO. It is clearly indicated that the TEM-EDS method is of use for the case where the difference in atomic scattering factor between the two constituent elements is small, while the CBED method is useful for

the case where the difference is large. In the presence of foil bending, it may be difficult to perform CBED experiments because any displacement of the electron beam on the sample is accompanied by a change of the local crystal orientation. In this case, the polarity can be determined from bend contours.

Sphalerite Structure. *Taftø* and *Spence* [44.31, 32] proposed a method of determining the deviations from centrosymmetry by taking advantage of the strong coherent multiple scattering normally present in CBED. In this technique, the TEM foil is tilted so that both the 200 reflection and two high odd-index reflections are simultaneously excited. When illuminating with a convergent electron beam, the effect of the dynamical interaction of these two weak reflections with the diffracted 200 beam is shown as cross-like contrast which intersects with the broad 200 Bragg line. Either bright or dark contrast suggests that the interference between the two weak reflections with the 200 reflection is either constructive of destructive and thus the type of the reflection *hkl* can then be revealed. An example image showing the detailed features of the $\bar{2}00$ and 200 diffraction discs for a thickness of 100 nm is shown in Fig. 44.7a,b. Figure 44.7c shows the two double-scattering paths which contribute to the 200 reflection.

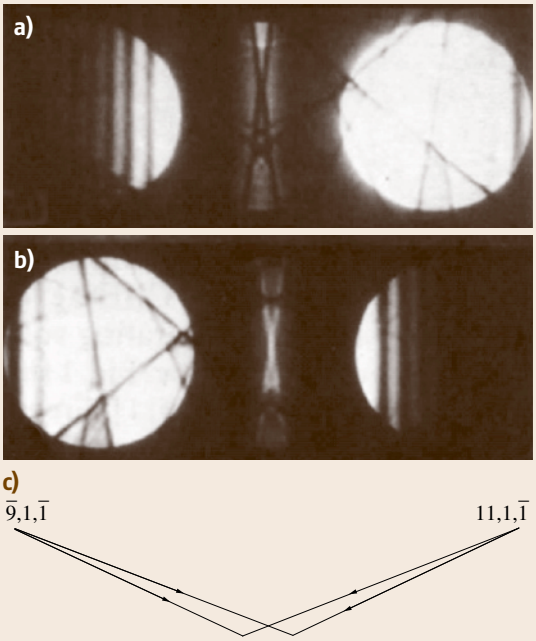


Fig. 44.7a–c $\bar{2}00$ **(a)** and 200 **(b)** diffraction discs showing the destructive and constructive interaction between the high-order odd reflections and $\bar{2}00/200$ reflection; **(c)** two double-scattering paths which contribute to the 200 reflection

Wurtzite Structure. For polarity determination of wurtzite crystals such as GaN, conventional CBED is generally carried out along the $\langle 01\bar{1}0\rangle$ axis. The intensity distributions within the $+\boldsymbol{g}$ and $-\boldsymbol{g}$ diffraction discs for polar directions such as [0002] and [000$\bar{2}$] are different and can be used for determination of atom distributions within the unit cell. To ensure the accurate determination of the polarity, the rotation angle and the 180° ambiguity between the image and the diffraction pattern need to be taken into account. Computer simulation of the intensity distribution taking into account the sample thickness, Debye–Waller factor, and absorption coefficient is necessary to verify the results. A good match between the experimental and simulated pattern is required for at least two different sample thicknesses to ensure that the interpretation is correct. Figure 44.8 shows a TEM image of an inversion domain boundary in an ammonothermally grown bulk GaN crystal [44.33]. The CBED patterns taken on either side of the boundary are shown as insets. An inversion

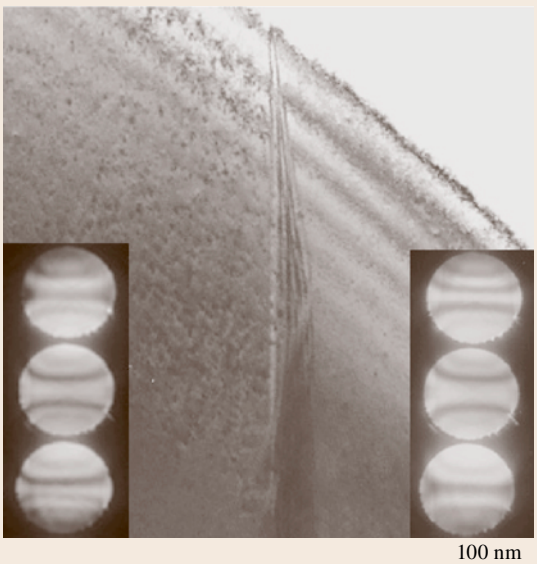


Fig. 44.8 TEM image showing a inversion domain boundary in a GaN bulk crystal; *insets* show the inversion of the intensity distribution in 0002 diffraction discs on either side of the boundary