Crystal Structure

It is useful to begin any discussion of crystal structure with an elementary definition of some crystallographic terms. First of all, crystals are made up of identical, repeating arrangements of atoms called *unit cells*. By definition, a unit cell is the smallest volume of a crystalline solid that exhibits the symmetry properties of the whole crystal. Just seven basic crystal systems are known. These are, cubic, hexagonal, tetragonal, orthorhombic, trigonal, monoclinic, and triclinic. These may be further classified into fourteen Bravais lattice types:

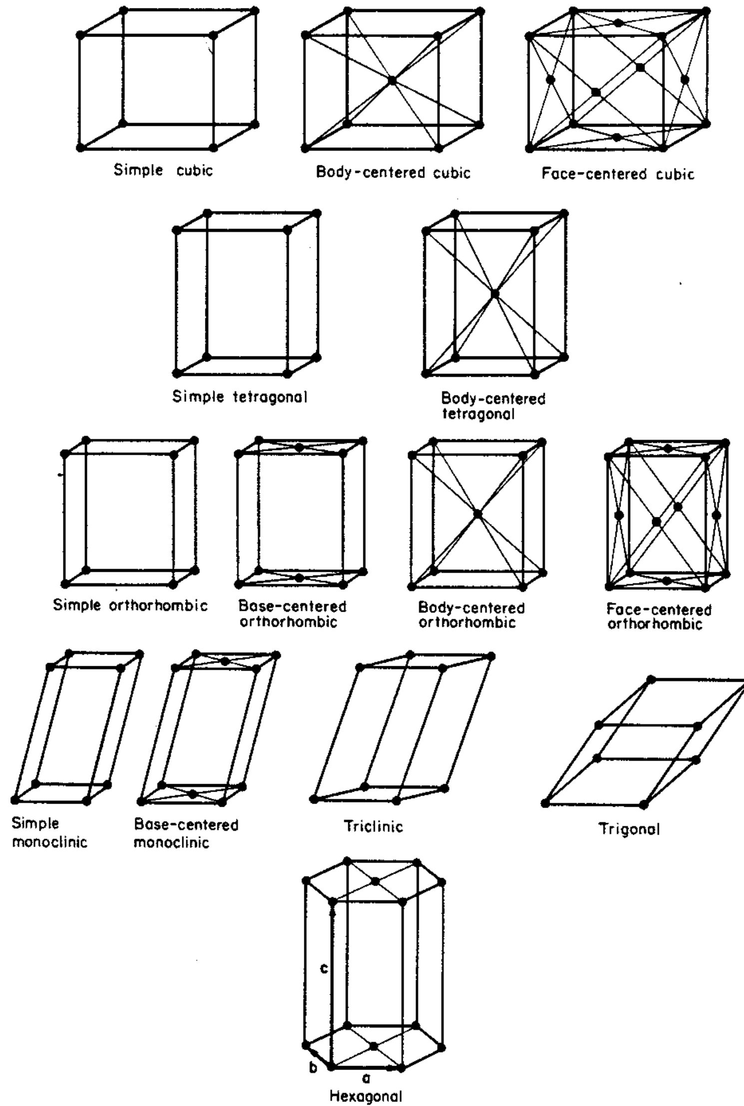


Fig. 11: The fourteen Bravais lattices

Of these types, for semiconductors the face centered cubic (FCC) system is of most importance. Fortunately, cubic lattices are also the easiest Bravais lattices to visualize and understand.

It is clear from the elementary structure of the Bravais lattices that each unit cell has several *lattice points*. In terms of the actual physical structure of a solid material, each lattice point is associated with a *basis group*. Thus, the lattice basis group for a particular crystal is a definite group of atoms associated with each lattice point. In the simplest case (which, for example, occurs in the case of some elemental metals) the basis group consists of just a single atom, in which case one atom occupies each lattice point. Of course, the lattice basis group must be identical for all lattice points. Obviously, for compound materials the basis group must consist of more than one atom since it cannot be defined as one kind of atom at one lattice point and another kind of atom at a different lattice point. Thus, one finds that in general, the basis group of a lattice consists of a definite repeating group of atoms. In some cases, the basis group may be identified with an actual covalent molecule that maintains its identity even when the lattice breaks up during melting or sublimation, (*e.g.*, as in the case of ice, water, and water vapor). However, in many cases, the basis group also breaks up with the lattice itself during change of phase (*e.g.*, as in the case of metallic or ionic solids). Finally, basis groups and unit cells should not be confused. Both are repeating groups of atoms, however, the basis group does not exhibit all of the symmetry properties of the whole crystal.

It is found that even for some elemental materials, the basis group consists of more than one atom. This is precisely the case for elemental silicon for which the lattice basis group consists of two atoms. To understand why this is so, one observes that the Bravais lattice for silicon is easily identified from powder x-ray diffraction patterns as FCC. However, one also recalls from previous consideration of electronic structure, that silicon has tetrahedral coordination due to the tetrahedral geometry of the sp^3 hybrid orbitals. Clearly, if a single silicon atom is inserted at each point of an FCC lattice, the resulting atomic coordination is not tetrahedral. However, if a two-atom basis group is inserted in the FCC lattice, tetrahedral coordination can be realized. The result is the *diamond cubic* structure, which can be thought of as two interpenetrating FCC lattices offset one quarter of the unit cell dimension in each direction:

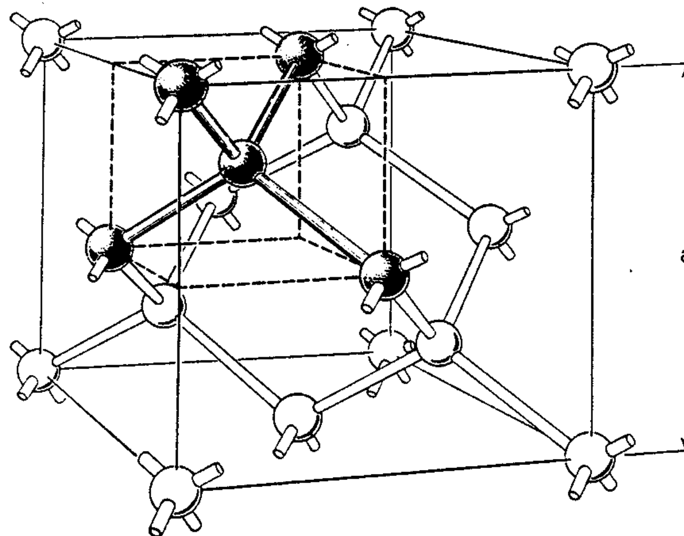


Fig. 12: Diamond cubic crystal structure

Obviously, the archetype of this structure is elemental carbon, *viz.*, diamond. It is well known that the diamond cubic structure has a high degree of symmetry including several mirror planes and two, three, and fourfold rotation axes. This is a direct consequence of the high degree of symmetry associated with tetrahedral coordination. In general, the group of symmetry properties serves as a unique specification of any crystal structure. In passing, one should note that the compound semiconductor, gallium arsenide, GaAs, has essentially the same structure as elemental silicon or germanium. However, for GaAs, the lattice basis group must consist of two different types of atoms, *viz.*, one Ga and one As, instead of two identical atoms, *i.e.*, either Si or Ge. This also corresponds to a cubic crystal structure, similar to the diamond cubic structure, called *zincblende* (after the naturally occurring mineral form of zinc sulfide, ZnS).

Another important term one often encounters in connection with semiconductor materials is *crystal orientation*. To see what is meant by crystal orientation, one returns to the basic unit cell which in the most general case, *i.e.*, a triclinic system, must be described by six independent *lattice parameters* (three lengths and three angles). However, again because of its high symmetry, a cubic system requires only a single lattice parameter, *a*, namely, the unit cell length. (Clearly, for a cubic system, the unit cell length is the same in all three directions and all the angles are 90°.) Thus, one naturally defines a coordinate system for the solid with the lattice parameter as the basic unit of length. Thus, in this "direct" space, any point within the crystal having all integer coordinates corresponds to a corner lattice point, *i.e.*, vertex, of a unit cell.

Crystal orientation is generally specified by *Miller indices*, (*h*,*k*,*l*). The exact definition of Miller indices as coordinates in "reciprocal" space is quite technical in nature and is beyond the scope of the present course; however a simple definition may suffice in terms of the three direct space coordinate axis intercepts. Since three points uniquely determine a geometric plane, the three axis intercepts define a *crystallographic plane*. Perhaps, the simplest example is provided by a plane that intersects each axis at a distance of one lattice parameter from the origin, *i.e.*, the plane has unit intercepts on each coordinate axis. This is primitively identified as the [111] crystallographic plane. In contrast, if one considers a plane that is parallel to the *y* and *z* axes, but has unit intercept on the *x* axis, how should this plane be designated? Clearly, from a strictly geometric point of view, the corresponding *y* and *z* axis intercepts for such a non-intersecting plane have "receded to infinity". It is found that a consistent designation results if the Miller indices, *h*, *k*, and *l*, are identified as reciprocals of the *x*, *y*, and *z* axis intercepts instead of the intercepts themselves. Thus, the reciprocal of ∞ formally corresponds to zero and the plane in question is consequently designated as [100]. Using this scheme, a plane that is parallel to the *z* axis, but has unit intercepts on the *x* and *y* axes is designated as [110]. Clearly, the case of [111] remains unchanged since unity is self-inverse. Indeed, it turns out that the definition of Miller indices as reciprocals of coordinate axis intercepts quite naturally follows from a general description of crystal diffraction phenomena in terms of three dimensional Fourier transforms. A further important observation to be made at this point is that since the coordinate origin in direct space is always translatable by an integral number of lattice parameters, the designation [*hkl*] really defines a *family* of planes rather than a single unique plane. (These families of planes take on particular importance for description of crystal diffraction phenomena.) Furthermore, although crystallographic planes can be arbitrarily specified, they are

generally only useful when they correspond to a plane of atoms having a regular pattern, *i.e.*, a two dimensional lattice.

Digressing briefly, it is important to note that within the general context of semiconductor processing, the designation of crystallographic planes (and crystal orientation) is generic in nature. Of course, in a strict sense Miller indices, [100], [010], and [001] must denote different families of crystallographic planes; however, all of these planes correspond to a face of the unit cell. Accordingly, again, as a consequence of the high degree of symmetry characteristic of the diamond cubic structure, all of these planes are electrically and structurally equivalent. Thus, any crystallographic plane corresponding to a face of the unit cell is generically called a "[100]" plane. Likewise, any crystallographic plane that "cuts" a unit cell face diagonally and is also parallel to an edge of the unit cell is called a "[110]" plane. Similarly, any crystallographic plane that intersects three non-adjacent vertices of the unit cell is called a "[111]" plane. These generic planes are illustrated below:

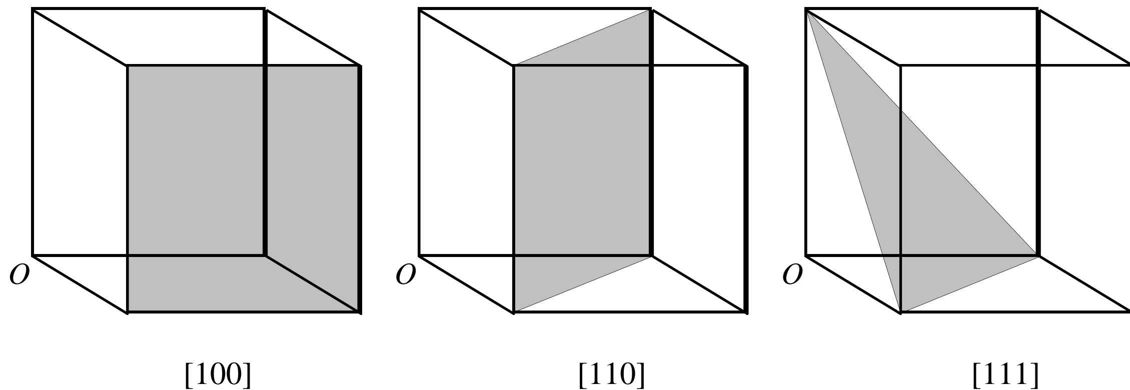


Fig. 13: Representative crystallographic planes

Clearly, in each unit cell, there must be three different, *i.e.*, non-parallel, [100] planes, six different [110] planes, and four different [111] planes.

In practice, physical crystals are generally designated by the orientation of their surfaces. In the case of semiconductor crystals, *i.e.*, silicon substrates, this is the crystallographic plane parallel to the surface used for device fabrication. Thus, a silicon crystal, *i.e.*, a substrate, which has a surface parallel to the unit cell face, is designated [100]. Similarly, if the surface can be thought of as intersecting three opposite corners of the unit cell, then the crystal is designated [111]. These two orientations are essentially the only ones ever used for device fabrication with [100] being much, much more common since it is used exclusively for fabrication of CMOS devices. A small amount of [111] material is still used for fabrication of bipolar devices for which very shallow doping is desirable. In principle, [110] substrates could be manufactured; however such material is of no real practical use and, as such, is very rare.

Clearly, the atomic arrangement for each of these surfaces, *i.e.*, [100], [111], and [110], is different. Directly related to this is the density of *dangling bonds* that are left behind if one breaks or "cleaves" a crystal parallel to a particular crystallographic plane. Naturally, this density is directly related to the surface energy of a crystal of a particular orientation. Therefore, one can primitively estimate relative surface energy by simply

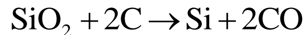
counting the number of bonds broken per unit cell if one cleaves the crystal along a given plane and then dividing by the surface area of that plane within the unit cell. (In reality, this simple picture is generally complicated by the occurrence of *surface reconstruction*, which lowers surface energy by overlapping neighboring dangling bonds.) One finds that in the diamond cubic structure, for a [100] plane, four bonds per unit cell must be broken in order to cleave the crystal. Obviously, the unit cell area associated with this cleavage just equals the square of the lattice parameter, a^2. Similarly, for cleavage parallel to a [111] plane, only three bonds per unit cell must be broken. Obviously, a [111] plane forms an equilateral triangle within the unit cell with each side having a length of $a\sqrt{2}$. The height of the equilateral triangle is $a\sqrt{3/2}$, thus, the corresponding area is $a^2\sqrt{2\times 3/2}/2 = a^2\sqrt{3}/2$. Finally, for [110] cleavage, again only three bonds must be broken. The corresponding area of a [110] plane within the unit cell is, of course, $a^2\sqrt{2}$. One, therefore, can determine primitive dangling bond densities, n_{100}, n_{111}, and n_{110} as follows:

$$n_{100} = 4/a^2 \qquad n_{111} = 2\sqrt{3}/a^2 \cong 3.8/a^2 \qquad n_{110} = 3\sqrt{2}/2a^2 \cong 2.1/a^2$$

Clearly, the dangling bond density for a [110] plane is much lower than either for a [100] or a [111] plane. Therefore, it follows that the crystal binding energy across a [110] surface is particularly low. Hence, a diamond cubic crystal should be much more easily broken parallel to a [110] surface than parallel to either a [100] or [111] surface. It turns out that [110] surfaces do indeed correspond to the natural *cleavage planes* of a silicon crystal. This can be demonstrated rather dramatically by breaking silicon substrates of different orientation from the center using a diamond-tipped scribe. One finds that [100] substrates naturally separate into quarters and [111] substrates into sixths.

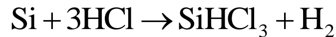
Crystal Growth

Considering more practical matters, integrated circuits are fabricated on single crystal silicon substrates which are mirror-like, polished circular disks called *wafers* (4"/100 mm dia. 525 μm thick, 6"/150 mm dia. 675 μm thick, 8"/200 mm dia., and 12"/300 mm dia., *etc.*). At present, the largest wafers commonly used in integrated circuit manufacturing are 300 mm although there is still substantial usage of smaller sizes, *viz.*, 200 mm. In addition, substrates of 450 mm (18") are under development by wafer manufacturers and are currently being used by a few manufacturers. Wafers are cut as slices from large single crystal ingots of silicon called *boules*. Of course, this silicon ultimately comes from quartzite sand, which is a naturally occurring mineral form of silicon dioxide, SiO_2. Typically, the raw oxide is reduced to *metallurgical grade silicon* in an electric furnace using carbon as the reducing agent:

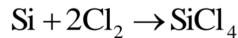


$$SiO_2 + 2C \rightarrow Si + 2CO$$

This material still contains substantial impurity and to obtain *electronic grade silicon*, metallurgical grade silicon is reacted with hydrogen chloride or chlorine:



$$Si + 3HCl \rightarrow SiHCl_3 + H_2$$



$$Si + 2Cl_2 \rightarrow SiCl_4$$

This produces volatile chlorides, *i.e.*, trichlorosilane ($SiHCl_3$) or tetrachlorosilane ($SiCl_4$), which are then carefully distilled and reduced again to silicon by pyrolysis in pure hydrogen (or some other high quality reducing agent). The resulting electronic grade polycrystalline material is quite pure and has as little as 0.05 ppb (50 ppt) of residual boron as the most common impurity.

There are two major crystal growth techniques. These are the *Czochralski* or CZ process and the *float zone* or FZ process. In the CZ process, electronic grade polycrystalline silicon is placed in a quartz crucible surrounded by graphite heat shielding and then heated to the molten state in an inert atmosphere by electrical heating elements. A seed rod having proper orientation is dipped into the melt and then controllably withdrawn. Naturally, the seed is much smaller in diameter than the desired crystal and, thus, the initial stage of CZ growth requires solidification outward from the seed to establish the desired diameter. Once diameter is established, the boule is slowly and controllably withdrawn such that molten silicon solidifies with the desired orientation and a large crystal is built up or "pulled". During the growth process, both the boule and the crucible are rotated to enhance uniformity. Of course, all of this requires very precise measurement and control of temperature and heat flux, which generally can be achieved for large diameter crystals only by sophisticated computer feedback and control. Typically, a CZ apparatus can hold many kilograms of molten silicon. (This, of course, depends on the size of the wafers being produced and has greatly increased over the last forty years.) The resulting boule is substantially free of crystal defects; however, it does contain oxygen contamination arising from the quartz crucible and carbon contamination from elsewhere (graphite shields, susceptor, *etc.*). In contrast, in the FZ method, a solid

electronic grade polycrystalline silicon rod is recrystallized and then refined using a "needles eye" furnace. Thus, the FZ method is essentially a classical *zone refining* technique. It produces the highest purity silicon available; however FZ silicon is generally quite brittle. Indeed, oxygen and carbon contamination in CZ silicon both enhances mechanical strength and allows for the application of internal gettering methods. For these reasons, the vast majority of silicon wafers used in integrated circuit fabrication are manufactured using the CZ process.

The Czochralski Process

From the point of view of integrated circuit manufacturing, it is desirable for the starting silicon wafers not to be intrinsic, but rather to be uniformly doped with some shallow level impurity, *i.e.*, B, P, As, Sb, *etc.* Thus, as well as crystal orientation, background doping (hence, majority carrier type and resistivity) is generally specified when wafers are purchased. Therefore, it is usual for the melt, hence, the grown ingot to be intentionally "contaminated" with a known quantity of shallow level dopant impurity. Therefore, control of dopant concentration and distribution during a CZ growth process is of fundamental importance.

As an initial description of Czochralski growth, it is usual to assume *rapid stirring* conditions, which implies that any excess impurity that might exist in the immediate vicinity of the growth interface is quickly dispersed into the melt. Consequently, such conditions imply that the melt is thoroughly mixed and, accordingly, concentration of any impurity is uniform throughout the melt, *i.e.*, right up to the freezing interface. Physically, this corresponds to very slow growth of the crystal such that the rate that impurity is dispersed throughout the melt is large in comparison to the rate that impurity is incorporated into the freezing crystal. Furthermore, one can safely assume that impurity diffusion within the solid crystal itself is unimportant since the diffusion coefficient of impurity in the liquid is many orders of magnitude larger than the corresponding solid diffusion coefficient. The CZ growth process can be represented pictorially as follows:

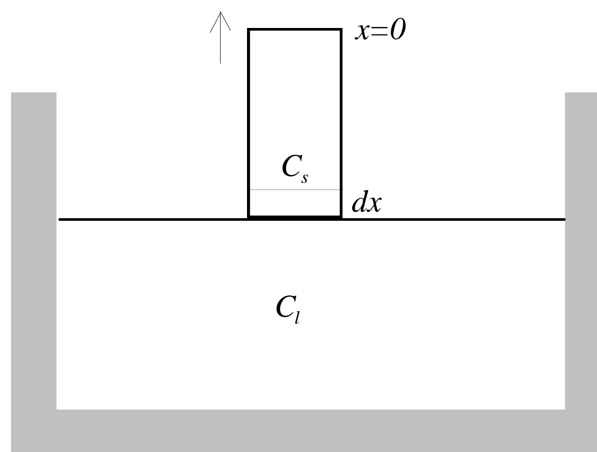


Fig. 14: Schematic diagram of the Czochralski (CZ) process

The arrow denotes the pulling direction. The dimension, x, is the length of crystal pulled from the melt, hence, the mass of crystal solidified, W, is just ρAx, such that A is the cross sectional area of the ingot (assumed to be uniform) and ρ is the density of silicon. Naturally, C_l is the volume concentration of solute, *i.e.*, dopant impurity, atoms in the melt and C_s is the volume concentration of solute atoms in the solid crystal. Clearly, at any given point during crystal growth and under the assumption of rapid stirring, C_l is uniform throughout the melt; however, C_s is a function of the position along the boule, *i.e.*, a function of x. Accordingly, if S is defined as the total number of solute atoms within the melt, then the differential number of solute atoms lost from the melt due to freezing a differential length of crystal is:

$$dS = -C_s A dx = -\frac{C_s}{\rho} dW$$

Here, the negative sign is formally included to indicate that atoms are lost from the melt to the growing crystal. Clearly, for a differential change in the length of the boule, *i.e.*, differential pull distance, dx, a corresponding differential mass of solid, dW, is added to the crystal. Obviously, these differential quantities are related simply as follows:

$$dW = \rho A dx$$

At an arbitrary point in the crystal growth for which a crystal of length, x, and mass, W, has been solidified, the mass of the remaining melt is just the difference, $W_o - W$, such that W_o is the initial mass of the melt before any crystal has been solidified, *i.e.*, pulled. Thus, it follows that the concentration of solute in the melt is given by:

$$C_l = \frac{S\rho}{W_o - W}$$

One formally solves this expression for S and combines it with the differential expression to obtain:

$$\frac{dS}{S} = -\frac{C_s dW}{\rho\left(\frac{C_l}{\rho}(W_o - W)\right)} = -\frac{C_s}{C_l}\frac{dW}{W_o - W}$$

This is a differential expression that relates the concentration of impurity atoms in the melt to the mass of the crystal at any stage during the growth process.

At this point, one might naively assume that the concentrations of impurity in the melt and in the solid are exactly equal; however, this is not the case. At a definite temperature such situations are described by a thermodynamic distribution equilibrium characterized by a constant coefficient, K, which is formally defined in the present case as the concentration ratio, C_s/C_l. In physical terms, one can regard this equilibrium as consequence of the fact that impurity atoms do not "fit" into the crystalline silicon lattice

as well as silicon atoms. Therefore, one expects that impurity atoms will be incorporated into the freezing solid at a lower intrinsic rate than are silicon atoms themselves and, thus, impurity atoms will tend to be "rejected" back into the melt. Accordingly, K can be expected to be generally less than one. As is established by experimental observations and shown in the following table, this is generally found to be the case:

Dopant	K
B	0.72
P	0.32
As	0.27
Sb	0.020
Ga	0.0072
Al	0.0018
In	0.00036

Table 1: Distribution Coefficients for shallow level impurities in silicon

Therefore, the preceding expression takes the form:

$$\frac{dS}{S} = -K\frac{dW}{W_o - W}$$

This expression is readily integrated directly as follows:

$$\int_{S_o}^{S} \frac{dS'}{S'} = -K\int_{0}^{W} \frac{dW'}{W_o - W'}$$

As with other parameters, S_o is defined as the initial number of solute atoms in the melt, which, of course, is determined by the simple formula:

$$S_o = \frac{C_o W_o}{\rho}$$

Here, C_o is just the initial impurity concentration in the melt. For convenience, the integration variable W' is formally replaced with a new variable, w, defined as $W_o - W'$, hence:

$$\int_{S_o}^{S} \frac{dS'}{S'} = K\int_{W_o}^{W_o - W} \frac{dw}{w}$$

Clearly, the indicated integration is elementary, thus:

$$\ln S - \ln S_o = K(\ln(W_o - W) - \ln W_o)$$

Moreover, it follows immediately from the elementary properties of logarithms that:

$$\ln\left(\frac{S}{S_o}\right) = K \ln\left(\frac{W_o - W}{W_o}\right)$$

Obviously, one inverts the logarithm on each side to obtain:

$$\frac{S}{S_o} = \left(\frac{W_o - W}{W_o}\right)^K$$

This is more conveniently expressed in terms of initial melt weight and impurity concentration, which are generally known before crystal growth starts, hence:

$$S = \frac{C_o W_o}{\rho}\left(\frac{W_o - W}{W_o}\right)^K$$

Furthermore, S has been previously related to melt concentration, C_l:

$$\frac{C_l}{\rho}(W_o - W) = \frac{C_o W_o}{\rho}\left(\frac{W_o - W}{W_o}\right)^K$$

Therefore, it immediately follows that:

$$C_l = C_o\left(\frac{W_o - W}{W_o}\right)^{K-1}$$

Of course, it is concentration of dopant in the solid crystal that is really of interest, but this is trivially obtained from the distribution equilibrium:

$$C_s = KC_o\left(\frac{W_o - W}{W_o}\right)^{K-1}$$

Thus, the impurity concentration in the solid and the melt is determined at all stages of crystal growth under rapid stirring conditions. This expression may be recast as a function of ingot length as follows:

$$C_s = KC_o\left(1 - \frac{\rho A x}{W_o}\right)^{K-1}$$

Physically, this equation describes variation of impurity concentration along the length of a CZ grown ingot under rapid stirring conditions. Clearly, at the "seed end", *i.e.*, $x=0$, of the boule, impurity concentration just corresponds to the simple distribution equilibrium. However, as the growth process proceeds, impurity is rejected from the growing crystal, and the concentration of the melt increases. This causes the concentration of impurity to increase in the boule as a function of distance from the seed end. This is shown in the following figure for several values of K (the individual curves are labeled by the corresponding value of K):

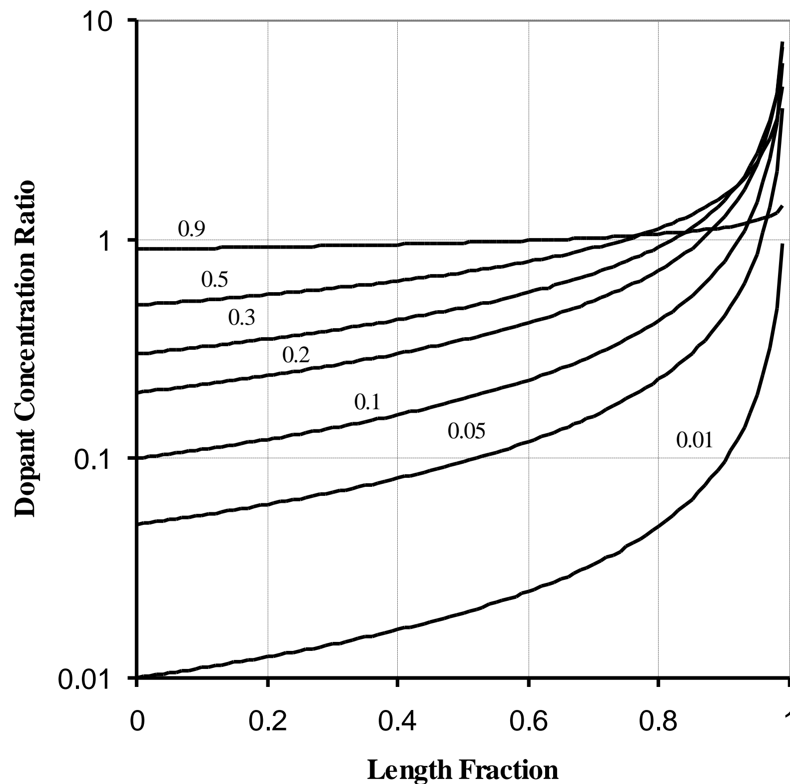


Fig. 15: Doping profile of a CZ crystal assuming rapid stirring

By definition, the seed end of the boule corresponds to a length fraction of zero. In contrast, the "butt end" is opposite of the seed end (*i.e.*, a length fraction nearly unity) and contains the last material solidified. Of course, dopant concentration ratio as indicated in the figure is just C_s/C_o. Clearly, the closer that K is to unity, the more uniform the doping profile obtained.

As might be expected, for realistic values of mixing, growth, and transport rates, the rejection rate of impurity at the growing crystal interface exceeds the rate at which impurities can be transported back into the bulk of the melt. Hence, the rapid stirring condition is not satisfied and breaks down. Consequently, in the vicinity of the growth interface impurity concentration builds up above the value observed in the bulk melt. Accordingly, this causes the crystal to be doped more heavily than would be expected

under rapid stirring conditions. This behavior can be understood by assuming that a stagnant boundary layer of thickness, δ, exists between the solid/melt interface and the bulk of the melt. Therefore, if D represents the impurity diffusion coefficient within the melt and if R is the "instantaneous" pull rate, *i.e.*, $d\Delta x/dt$, then assuming nearly steady-state conditions, the impurity concentration in the stagnant region near the growth interface may be described by the expression:

$$D\frac{d^2C}{d\Delta x^2} + R\frac{dC}{d\Delta x} = 0$$

Here, C is local solute concentration within the melt and Δx is distance from the growth interface. Clearly, C must converge to the uniform concentration, C_l, if Δx becomes large. Physically, the first term in the preceding expression accounts for diffusion of impurity atoms away from the growth interface back into the melt, and is a direct consequence of Fick's Second Law. Likewise, the second term accounts for removal of impurity from the melt due to freezing. Mathematically, this term has the appearance of a "net drift" due to some external "potential", but instead physically describes net motion of material due to pulling the crystal. Indeed, there must indeed be a net motion of impurity atoms with respect to the growth interface simply due to solidification, *i.e.*, in principle, all impurity atoms eventually must "pass through" the growth interface. This amounts to an overall relative motion of impurity atoms that can be regarded equivalently as being due to movement of the growth interface with respect to a stationary melt, or to movement of the melt with respect to a stationary growth interface, or as some combination of the two. (Terms describing convective flow of the melt are absent since the boundary layer is assumed to be stagnant.) Naturally, under conditions of nearly steady state, diffusion and freezing terms must effectively balance each other. For simplicity, C' is identified with the derivative, $dC/d\Delta x$, thus, the preceding equation becomes:

$$\frac{dC'}{d\Delta x} = -\frac{R}{D}C'$$

Clearly, this expression can be trivially integrated; however, instead of a definite integral it is convenient to construct an indefinite form as follows:

$$C' = (C'_0 - C'_\infty)e^{-R\Delta x/D} + C'_\infty$$

Here, C'_0 and C'_∞ are defined, respectively, as solute concentration gradients at the growth interface ($\Delta x = 0$) and far away from the growth interface ($\Delta x \to \infty$). Naturally, numerical values of C'_0 and C'_∞ correspond to imposition of suitable boundary conditions. Accordingly, as asserted previously, far away from the growth interface concentration of dopant impurity can be expected to be uniform, which implies that C'_∞ vanishes. Thus, the preceding formula can be further simplified:

$$C' = C'_0 e^{-R\Delta x/D}$$

To determine C'_0 diffusion and rejection fluxes of impurity atoms at the growth interface are assumed to be equal. Of course, this is consistent with the original differential equation describing the effects of diffusion and freezing in the stagnant boundary layer, and implies that the solute concentration profile across the growth interface and boundary layer remains essentially in a steady state (or at least only very slowly varying one). Thus, for a unit area of growth interface, one can write:

$$-DC'_0 = R(\overline{C}_l - C_s)$$

Clearly, the term on the left just comes from Fick's First Law, which relates diffusion flux linearly to concentration gradient. Obviously, the number of impurity atoms per unit volume rejected back into the melt must just be the difference of impurity concentrations in the melt and in the solid exactly at the growth interface. (By definition, $\overline{C}_l$ is the impurity concentration exactly at the growth interface.) Therefore, the term on the right is the rejection flux and is just the product of the interfacial concentration difference and the instantaneous pull rate, *R*; hence, it follows that:

$$C' = \frac{dC}{d\Delta x} = -\frac{R}{D}(\overline{C}_l - C_s)e^{-R\Delta x/D}$$

It is a simple matter to integrate this expression across the entire boundary layer to obtain the expression:

$$C_l - \overline{C}_l = -\frac{R}{D}(\overline{C}_l - C_s)\int_0^{\delta} d\Delta x \exp\left(-\frac{R}{D}\Delta x\right) = (\overline{C}_l - C_s)\left(\exp\left(-\frac{R}{D}\delta\right) - 1\right)$$

Here, the solute concentration at the boundary layer edge, *i.e.*, $\Delta x = \delta$, has been assumed to be the uniform bulk concentration, C_l. Of course, this condition is not strictly realized unless $\Delta x \to \infty$; however, it is reasonable to assume that δ is sufficiently large so that the concentration for $\Delta x = \delta$ is only negligibly different from C_l. This expression is easily rearranged into a more convenient form:

$$\frac{C_l - C_s}{\overline{C}_l - C_s} = e^{-R\delta/D}$$

At this point, one redefines the ratio, C_s / C_l, as an effective segregation coefficient, K_e, since an enriched boundary layer lies between the bulk and the growth interface. (Of course, the ratio, $C_s / \overline{C}_l$, must equal the actual segregation coefficient, *K*, since $\overline{C}_l$ is defined as the concentration at the actual growth interface.) Therefore, the preceding expression becomes:

$$\frac{\frac{1}{K_e}-1}{\frac{1}{K}-1}=e^{-R\delta/D}$$

Obviously, one solves for K_e as follows:

$$\frac{1}{K_e}=\left(\frac{1}{K}-1\right)e^{-R\delta/D}+1$$

Upon formally taking the reciprocal the desired expression is obtained, thus:

$$K_e=\frac{K}{K+(1-K)e^{-R\delta/D}}$$

This result is applied to an actual CZ growth process by just substituting K_e into expressions obtained previously for rapid stirring conditions. The parameter, $R\delta/D$, is called *normalized growth parameter* and in practice is determined empirically for a given crystal growing apparatus. Clearly, if the normalized growth parameter is made sufficiently large (for example, by increasing the pull rate), then K_e tends toward unity, and this will result in a more uniform distribution of dopant along the length of the boule. Conversely, if the pull rate and, therefore, the normalized growth parameter becomes small, then K_e tends toward K. Obviously, this just corresponds to a return to rapid stirring conditions.

Both D and R can be adjusted by a judicious choice of process conditions, *e.g.*, temperature and pull rate. What about δ? It turns out that δ is a function of the rotation rate of the boule for which an empirical relationship has been determined experimentally:

$$\delta=1.8D^{1/3}\nu^{1/6}\omega^{-1/2}$$

Here, ν is the viscosity of the melt and ω is rotation rate. The pull rate, R, is closely related to the actual growth rate of the crystal, however, the instantaneous growth rate may differ from R because of thermal fluctuations, supercooling, *etc.* Indeed, if the pull rate is relatively small, the instantaneous growth rate may become negative (this is called *re-melting*). This can adversely affect both defect structure and doping distribution on a microscopic scale. In particular, if re-melting is not strongly suppressed by the use of a sufficiently large pull rate, the crystal may exhibit defect "swirl patterns" and dopant striations. Moreover, convective transport within the melt may also redistribute impurities non-uniformly. This is especially significant in the case of oxygen which is dissolved from the quartz crucible at the melt periphery. Indeed, because of these kinds of variations it is difficult to produce large diameter, lightly doped CZ wafers, *i.e.*, with a resistivity exceeding 100 Ω cm. Recently, immersion of the crucible during crystal growth in a strong magnetic field (the *magnetic Czochralski* or MCZ process) has been

found to allow control of convective transport, which improves uniformity of large diameter crystals (>300 mm). As wafer sizes increase this may be expected to become an industry standard. Additionally, if the pull rate is low, the solidified crystal may be held for quite a long time above 950°C. This may allow sufficient time for thermally generated "microdefects" to form. The formation of such microdefects is effectively quenched if the pull rate exceeds a rate of roughly 2 mm/min.

Clearly, the normalized growth parameter and, hence, the effective segregation coefficient can be modified by changing rotation and pull rates during crystal growth. Indeed, it is common commercial practice to program growth parameters so as to obtain a uniform impurity concentration over a large fraction of an ingot. Accordingly, crystal growth proceeds in distinct phases: First, during an initial growth phase, as asserted previously, crystal diameter is built up to the desired dimension. Next, programmed pull and rotation rates are applied. This results in a crystal of constant impurity concentration over a large fraction of its length. Of course, at some point the melt is substantially exhausted and it becomes impossible to sustain a uniform composition. This defines a third growth phase during which the crystal is rapidly pulled out of the melt.

Zone Refining

As noted at the outset, in CZ growth there can be considerable carbon and oxygen contamination that comes from the quartz and graphite components of the process equipment. In general, this contamination causes no problem and perhaps may be beneficial since, as observed previously, ultrapure silicon is actually mechanically more fragile than ordinary CZ silicon and, also, as will be made evident subsequently, oxygen contamination may be used to good effect to set up an internal gettering scheme. However, there are some cases for which ultrahigh purity is desired, *viz.*, 1-10 KΩ cm). This material is most conveniently fabricated by zone refining. Again, to reiterate, within the industry substrates fabricated this way are called float zone (FZ) wafers.

To understand zone refining, suppose that just as in CZ growth, some impurity dissolved in molten silicon is in equilibrium with solid crystal. Once again, the distribution (or segregation) coefficient, K, is defined:

$$K = \frac{C_s}{C_l}$$

Of course, C_s is the impurity concentration within the solid and C_l is the impurity concentration in the liquid. Moreover, again, just as for Czochralski growth it is possible to maintain the system in a steady state, but not in rigorous thermal equilibrium, such that an effective distribution coefficient, K_e, is applicable instead of absolute distribution coefficient, K.

In practice, the basic technique of zone refining is to pass a solid piece of material, *e.g.*, an ingot of silicon, through a circular heating element, *viz.*, "needles eye". This creates a molten zone that slowly moves from one end of the ingot to the other. (Of course, migration of the molten zone along the ingot can be accomplished either by moving the ingot through a fixed heating element or by moving the heating element holding the ingot in a fixed position.) In any case, as the molten zone migrates and as a

consequence of the distribution equilibrium, impurities are collected in the molten material and "swept" preferentially to one end of the ingot. Consequently, the rate that the molten zone passes through the ingot is analogous to the pull rate in the Czochralski process. Thus, the effective distribution coefficient should be essentially determined by the zone migration rate. Within this context, one expects that K_e should be significantly larger than K (or even, perhaps, approach unity) if the molten zone is moved very rapidly through the ingot and, in contrast, should approach or essentially equal K if the molten zone is moved very slowly. For clarity, it is instructive to consider a single pass zone refining process, which may be represented pictorially as follows:

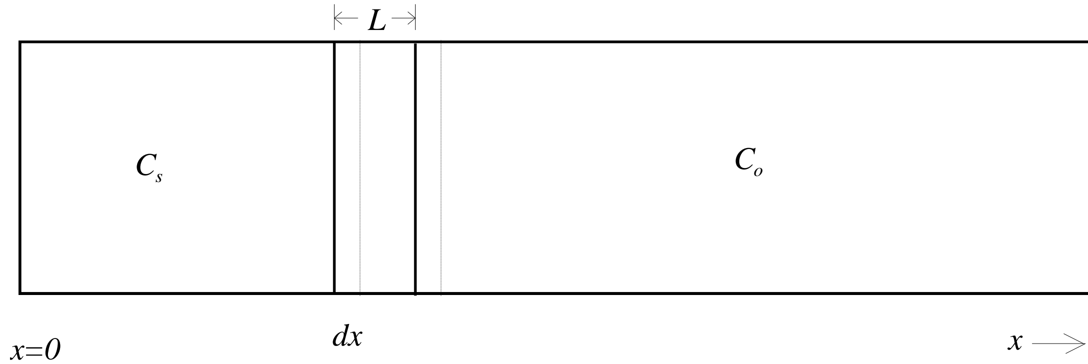


Fig. 16: Schematic diagram of the Float Zone (FZ) process

Here, L is the length of the molten zone, C_s is redefined as the impurity concentration in the refined section of the ingot, and C_o is the impurity concentration in the unrefined section. The variable, x, represents linear distance along the ingot (with the zone refining process initiated conventionally at $x=0$). If S is the number of impurity atoms in the molten zone and A is the cross sectional area of the ingot, then the differential change in S as the molten zone passes through the ingot is given by:

$$dS = C_o A dx - C_s A dx$$

This equation is just a formal expression of the difference in the rate that impurity atoms are added to the molten zone due to melting of unrefined material to the rate that they are lost to the melt at the freezing interface. Clearly, the impurity concentration in the liquid, C_l, is just S/AL. Hence, if one assumes that the distribution of impurity at the freezing interface is governed by the effective distribution coefficient, K_e, then it follows that:

$$\frac{dS}{dx} = C_o A - K_e \frac{S}{L}$$

(Obviously, the term, $C_o A$, is proportional to the rate that impurity atoms enter the liquid at the melting interface and, likewise, $K_e S/L$ is proportional to the rate that impurity atoms are removed from the melt at the freezing interface.) If one assumes that the initial impurity concentration, C_o, is uniform throughout the unrefined section of the ingot, then,

this differential equation is easily integrated by means of an exponential integrating factor. Accordingly, if one defines a new solute parameter, Q, as $S\exp(K_e/L)$, then one obtains:

$$\frac{dQ}{dx} = C_o A e^{K_e x / L}$$

Therefore, it follows trivially that:

$$Q(x) - Q(0) = \frac{C_o AL}{K_e}\left(e^{K_e x / L} - 1\right)$$

Recasting this expression in terms of S yields the result:

$$S(x)e^{K_e x / L} - S(0) = \frac{C_o AL}{K_e}\left(e^{K_e x / L} - 1\right)$$

The boundary condition for $S(0)$ must just be $C_o AL$ since melting an unrefined portion of the ingot of volume AL forms the initial molten zone. Upon substitution, one obtains:

$$S(x) = \frac{C_o AL}{K_e}\left(1 - (1 - K_e)e^{-K_e x / L}\right)$$

If one uses the definition of C_l and the distribution equilibrium, it follows that:

$$C_s(x) = C_o\left(1 - (1 - K_e)e^{-K_e x / L}\right)$$

This expression describes the concentration of initially uniformly distributed impurity after single pass zone refining.

Further purification can be achieved by additional zone refining. The equation describing the process is just the same as before, however, the initial impurity concentration is no longer uniform; hence, the original differential equation must be modified as follows:

$$\frac{dS}{dx} = C_o(x)A - K_e \frac{S}{L}$$

Here, $C_o(x)$ is an arbitrary (*i.e.*, non-uniform) initial impurity distribution. Again, this expression is recast in terms of Q:

$$\frac{dS}{dx} = \frac{dQ}{dx}e^{-K_e x / L} - K_e \frac{Q}{L}e^{-K_e x / L} = C_o(x)A - K_e \frac{Q}{L}e^{-K_e x / L}$$

Therefore, one obtains:

$$\frac{dQ}{dx} = C_o(x) A e^{K_e x / L}$$

Naturally, this differential equation is easily integrated formally to give:

$$Q(x) - Q(0) = A \int_0^x dx' C_o(x') e^{K_e x' / L}$$

Following the result for constant C_o, one identifies $Q(0)$ as $C_o(0)AL$ and recasts this expression in terms of S:

$$S = A e^{-K_e x / L} \int_0^x dx' C_o(x') e^{K_e x' / L} + C_o(0) A L e^{-K_e x / L}$$

It immediately follows from the distribution equilibrium at the freezing interface that the solute concentration in the refined ingot is given by the expression:

$$C_s(x) = \frac{K_e}{L} e^{-K_e x / L} \int_0^x dx' C_o(x') e^{K_e x' / L} + K_e C_o(0) e^{-K_e x / L}$$

Of course, to determine $C_s(x)$ numerically, prior knowledge of $C_o(x)$ is required. Naturally, $C_o(x)$ may be the result of a previous zone refining pass or may be a "grown-in" distribution due to a particular set of process parameters for a CZ growth process.

Within this context, if one substitutes the concentration profile obtained previously for single pass zone refining of an initially uniform ingot into the preceding expression, one obtains:

$$C_s(x) = \frac{K_e C_o}{L} e^{-K_e x / L} \int_0^x dx' \left(1 - (1 - K_e) e^{-K_e x' / L} \right) e^{K_e x' / L} + K_e^2 C_o e^{-K_e x / L}$$

Here, $C_s(x)$ is evidently the concentration profile of the ingot following two zone refining passes. Clearly, the integrals are all of elementary form and can be constructed explicitly to give the result:

$$C_s(x) = C_o \left(1 - \left(1 - K_e^2 + \frac{K_e}{L}(1 - K_e) x \right) e^{-K_e x / L} \right)$$

Of course, C_o retains the usual definition as initial uniform impurity concentration in the ingot. Naturally, one can compare this expression for two passes to the single pass result

and also to the initial uniform impurity concentration. Clearly, at the starting, *i.e.*, seed, end of the ingot, the impurity concentration was, of course, just C_o prior to zone refining. After a single pass, this is reduced to $K_e C_o$. After two passes this is further reduced to $K_e^2 C_o$. Clearly, since K_e is smaller than one, this shows that the seed end of the ingot is progressively purified. Concentration profiles for one and two zone refining passes with various effective distribution coefficients are shown in the following figure:

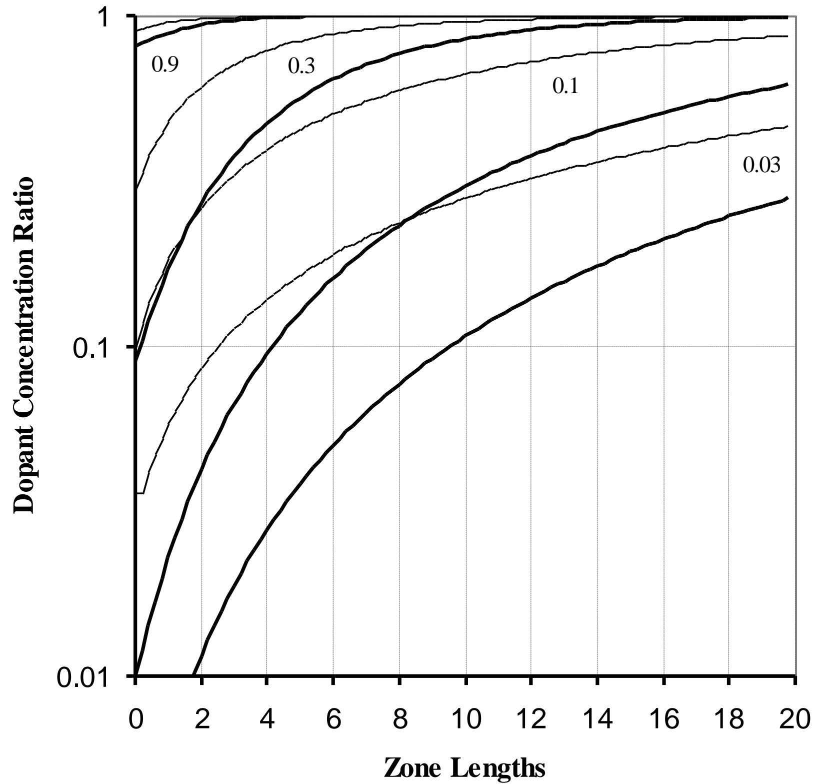


Fig. 17: Doping profile of an FZ crystal following one (*narrow line)* and two (wide line) zone refining passes

Of course, the dopant concentration ratio is, again, just C_s/C_o. The horizontal axis is the number of zone lengths refined. (Hence, by definition, the seed end of the ingot corresponds to zero zone lengths.) One observes that in contrast to the Czochralski process for which an effective distribution coefficient near unity is desired, in zone refining better results are obtained the smaller the value of K_e. (Of course, K_e can never become smaller than K itself.) Clearly, this implies that a slow rate of migration of the molten zone through the ingot is desirable.

Physically, zone refining literally sweeps impurity atoms toward one end of the ingot, *i.e.*, the end toward which the molten zone moves. Since, no impurity is physically removed from the ingot this end actually becomes more impure. However, zone refining process parameters can be set up in such a way that the impure end is a reasonably small part of the total volume of the ingot. Clearly, if zone refining is repeated many times, the bulk of an ingot can be refined to ultrahigh purity and the impure end can simply be

removed resulting in a large amount of highly refined material. Of particular importance is the production of wafers with low oxygen content since these cannot be produced using the CZ process. Of course, zone refining does not leave impurity uniformly distributed. If this is a requirement, then once the impure end has been removed, the ingot can be heat treated to redistribute impurity more uniformly. To reiterate, the main advantage of the float-zone process is the very low impurity concentration in the silicon crystal. In particular the oxygen and carbon concentrations are much lower as compared to the CZ process, since the melt does not come into contact with a quartz crucible, and no hot graphite container is used. As a practical matter, FZ ingots are produced from an initial polycrystalline silicon ingot, which is seeded with a crystal of desired orientation at the start of the process. Alternatively, a monocrystalline ingot can be further purified by zone refining. Even so, the FZ process is more expensive than then CZ process, and, at present, crystal diameter is limited to 200 mm.