JOURNAL OF MATERIALS SCIENCE 38 (2003) 3341 – 3429

Review
Applications of isotope effects in solids

V. G. PLEKHANOV
Fonoriton Science Lab., Garon Ltd., PO Box 2632, Tallinn, 13813, Estonia
E-mail: vgplekhanov@hotmail.com

This article reviews the current status of the employment of the isotope effect in solids. Diffusion, self-diffusion processes with different isotopes in pure materials and heterostructures (quantum wells), neutron transmutation doping of different semiconducting crystals, optical fiber as well as use isotope-mixed crystals (C, LiH) as the generator of the coherent radiation in the ultraviolet range of the spectrum are the main modern applications of isotope science and engineering. There are briefly discussed the other future applications including modern personal computer, isotope-based quantum computer as well as information storage. We hope to give sufficient references to published work so that the interested reader can easily find the primary literature sources to this rapidly expanding field of solid state physics. © *2003 Kluwer Academic Publishers*

Introduction

The availability of isotopically pure crystals with low carrier and impurity concentrations has allowed in the last three decades the investigation of isotope effects on lattice dynamical and electronic properties of solids [1]. The results of experimental and theoretical studies of the fundamental properties of the objects of research that earlier were simply in accessible (naturally with exception of LiH_xD_{1-x} crystals) briefly are presented in the reviews [2–4]. The use of such objects allows the investigation of not only the isotope effects in lattice dynamics (elastic, thermal and vibrational properties) but also the influence of such effects on the electronic states via electron-phonon coupling (the renormalization of the band-to-band transition energy E_g, the exciton binding energy E_B and the size of the longitudinal-transverse splitting Δ_{LT}). The thermal conductivity enhancement in the isotopically enriched materials amounts (C, Ge, Si) to almost 60% at room temperature and is close to a factor six at the thermal conductivity maximum around 20 K (Si-case) (see also [3, 4]). The change in the lattice constant is $\Delta a/a \sim 10^{-3} \div 10^{-4}$, while the change δc_{ik} in the elastic constants amounts to several percent. In addition, crystals of different isotopic compositions possess different Debye temperatures. This difference between a LiH crystal and its deuteride exceeds hundred degrees. Of the same order of magnitude is the difference between Debye temperatures for diamond crystals. Very pronounced and general effects of isotopic substitution are observed in phonon spectra. The Raman lines in isotopically mixed crystals are not only shifted (the shift of LO phonon lines exceeds 100 cm^{-1}) but are also broadened. This broadening is related to the isotopic disorder of a crystal lattice. It is shown that the degree of change in the scattering potential is different for different isotopic mixed crystals [1]. In the case of semiconducting crystals (C, Ge, Si, α-Sn etc.), phonon scattering is weak, which allows one to successfully apply the coherent potential approximation (CPA) for describing shift and broadening of scattering lines in Raman spectra [2, 3]. In the case of LiH, the change in the scattering potential is so strong that it results in phonon localization, which is directly observed in experiments [1, 4].

Substituting a light isotope with a heavy one increases the interband transition energy E_g (excluding Cu-salts) and the binding energy of the Wannier-Mott exciton E_B as well as the magnitude of the longitudinal-transverse splitting Δ_{LT} [5, 6]. The nonlinear variation of these quantites with the isotope concentration is due to the isotopic disordering of the crystal lattice and is consistent with the concentration dependence of line halfwidth in exciton reflection and luminescence spectra. A comparative study of the temperature and isotopic shift of the edge of fundamental absorption for a large number of different semiconducting and insulating crystals indicates that the main (but not the only) contribution to this shift comes from zero oscillations whose magnitude may be quite considerable and comparable with the energy of LO phonons. The theoretical description of the experimentally observed dependence of the binding energy of the Wannier-Mott exciton E_B on the nuclear mass requires the simultaneous consideration of the exchange of LO phonons between the electron and hole in the exciton, and the separate interactions of carriers with LO phonons (see also [1]). The experimental dependence $E_B \sim f(x)$ for LiH_xD_{1-x} crystals fits in well enough with the calculation according to the model of large-radius exciton in a disordered medium; hence it follows that the fluctuation smearing of the band edges is caused by isotopic disordering of the crystal lattice. Due to zero-point motion, the

atoms in a solid feel the anharmonicity [7] of the interatomic potential even at low temperatures. Therefore, the lattice parameters of two chemically identical crystals formed by different isotopes do not coincide heavier isotopes having smaller zero-point delocalization (as expected in a harmonic approximation) and smaller lattice parameters (an anharmonic effect). Moreover, phonon—related properties such as thermal conductivity, thermal expansion or melting temperature, are expected to depend on the isotope mass (details see [1]).

Our brief discussion, we start with a fact that phonon frequency are directly affected by changes of the average mass of the whole crystal or its sublattice (VCA-model), even if we look upon them as noninteracting particles, i.e., as harmonic oscillators. The direct influence of the isotope mass on the frequencies of coupled phonon modes may been used to determine their eigenvectors. Secondly, the mean square amplitude $\langle u^2 \rangle$ of phonons depend on the isotope masses only at low temperature, while they are determined by the temperature T only, once T becomes larger than Debye temperature. A refinement of these effects must take place when taking interactions among phonons into account. These interactions lead to finite phonon lifetimes and additional frequency renormalization. The underlying processes can be divided into two classes: (1) anharmonic interactions in which a zone center phonon decays into two phonons or more with wave-vector and energy conservation, and (2) elastic scattering in which a phonon scatters into phonons of similar energies but different wave-vectors. While the former processes arise from cubic and quartic terms in the expansion of lattice potential [7], the latter are due to the relaxed wave-vector conservation rule in samples that are isotopically disordered and thus not strictly translationally invariant. Since the vast majority of compounds derive from elements having more than one stable isotope, it is clear that both processes are present most of the time. Unfortunately, their absolute sizes and relative importance cannot be predicted easily. However, isotope enrichment allows one to suppress the elastic scattering induced by isotope disorder. In contrast, the anharmonic phonon-phonon interaction cannot be suppressed, so that isotope-disorder-induced effects can only be studied against a background contribution from anharmonic processes. However, if one assumes that the two processes are independent of each other one can measure the disorder-induced renormalization by comparison of phonon energies and linewidth of isotopically pure samples with those gained from disordered ones.

The isotopic composition affects the band-gaps through the electron-phonon coupling and through the change of volume with isotopic mass. Although the electronic properties of different isotopes of a given atom are, to a very good approximation, the same, isotope substitution in a crystal modifies the phonon spectrum which, in turn, modifies the electron energy bands through electron-phonon interaction. Measuring the energy gaps in samples with different isotopic composition then yields the difference in the changes of the valence- and conduction-band renormalization. The reason for the changes lies in the fundamental quantum-mechanical concept of zero-point motion—the vibrational energy that the atoms in the crystal have, even at low temperatures. If we excite an electron from one electronic state to another, we actually excite the whole crystal. In other words, we move the crystal from a ground state made up of low-energy electrons plus zero-point vibrations to an excited state in which there is one excited electron plus the zero-point vibrations of the crystal. The values of the zero-point energy in the two electronic states are slightly different because the vibrational frequencies depend on the chemical bonding, which is changed by exciting an electron. If the average mass of the vibrating atoms is increased, then the vibrational frequencies will be reduced. As a result, the difference in zero-point motions will be smaller, and the transition energy will therefore increase with increasing mass [1].

Present review is devoted to description of different applications of the isotope effect in solids. In the Chapter 1 we detail analyze the process of self-diffusion in isotope pure materials and heterostructure. Interest in diffusion in solids is as old as metallurgy or ceramics, but the scientific study of the phenomenon may probably be dated some sixth-seven decades ago. As is well-known, the measured diffusion coefficient depends on the chemistry and structure of the sample on which it is measured. We have organized the chapter around general principles that are applicable to all materials. We are briefly discussed the SIMS (secondary ion mass spectrometry) technique. In SIMS technique, the sample is bombarded by reactive ions, and the sputtered-off molecules are ionized in a plasma and fed into a mass-spectrometer. Self-diffusion is the migration of constituent atoms in materials. This process is mediated by native defects in solids and thus can be used to study the dynamics and kinetics of these defects. The knowledge obtained in these studies is pivotal for the understanding of many important mass transport processes such as impurity diffusion in solids.

Chapter 2 describes the new reactor technology—neutron transmutation doping (NTD). Capture of thermal neutrons by isotope nuclei followed by nuclear decay produces new elements, resulting in a very number of possibilities for isotope selective doping of solids. There are presented different facilities which use in this reactor technology. The feasibility of constructing reactors dedicated to the production of NTD silicon, germanium (and other compounds) was analyzed in terms of technical and economic viability and the practicality of such a proposal is examined. The importance of this technology for studies of the semiconductor doping as well as metal-insulator transitions and neutral impurity scattering process is underlined. The introduction of particle irradiation into processing of semiconductor materials and devices creates a new need for additional understanding of atomic-displacement-produced defects in semiconductors. It is shown that measurement of decay rates of induced radioactivity and the system of clearance and certification such as to allow the solids to be internationally transported as "Exemt Material", as defined in IAEA Regulations, are dealt with.

The use of the isotopes in a theory and technology of the optical fibers we considered in Chapter 3. This chapter is addressed to readers who wish to learn about fiber communications systems and, particular, about the properties of optical fibers. Very briefly in this chapter we describe the Maxwell equations as well as wave electromagnetic equation. In this chapters we describe not only the properties of optical fibres but also the materials for optical fiber and fiber technology. For the first time it was shown also the influence of the isotopes on properties of the optical fibers.

Chapter 4 is devoted the application of isotope effect in laser physics. There is short description of theory and practice of semiconductor lasers. The discovery of the linear luminescence of free excitons observed over a wide temperature range has placed lithium hydride, as well as crystals of diamond in line as prospective sources of coherent radiation in the UV spectral range. For LiH isotope tuning of the exciton emission has also been shown [1].

The last chapter of this review is devoted to description of the other unexplored applications of isotopic engineering. In the first place we considered the materials for information storage in modern personal computers as well as in biology. In this chapter is shown that isotopic substitution has made it possible to produce the objects of research that earlier were simply inaccessible (with exception of the LiH-LiD system). The use of such objects allows the investigation of not only the isotope effects in lattice dynamics (elastic, thermal and vibrational properties) but also the influence of such effects on the electronic states (the renormalization of the band-to-band transition energy E_g, the exciton binding energy E_B, and the size of the longitudinal-transverse splitting Δ_{LT}). Very perspective is isotope-based quantum computer. We should add here that the strength of the hyperfine interaction is proportional to the probability density of the electron wavefunction at the nucleus. In semiconductors, the electron wavefunction extends over large distances through the crystal lattice. Two nuclear spins can consequently interact with the same electron, leading to electron-mediated or indirect nuclear spin coupling. Because the electron is sensitive to externally applied electric fields, the hyperfine interaction and electron-mediated nuclear spin interaction can be controlled by voltages applied to metallic gates in a semiconductor device, enabling the external manipulation of nuclear spin dynamics that is necessary for quantum computation in quantum computers (details see [383, 384]).

A brief summary is presented in the conclusion. The difficult and unsolvable problems of isotope effects in solids are considered there. The main aim of this review is to familiarize readers with present and some future applications in isotope science and engineering.

Chapter 1. Process of self-diffusion in isotope pure materials and heterostructures

1.1. General remarks

Interest in diffusion is as old as metallurgy or ceramics. The first measurement of diffusion in the solid state was made by Roberts–Austen in 1896 [8]. Many measurements, especially of chemical diffusion in metals, were made in the 1930s; the field was reviewed by Mehl [9], Jost [10], and Seith [11]. Diffusion research increased after World War II; the increase was motivated by the connection among diffusion, defects, and radiation damage and helped by the availability of many artificial radiotracers. These researchers were the first to attempt to identify the basic underlying atomistic mechanisms responsible for mass transport through solids by a quantitative investigations and theoretical analysis of the activation energies required for diffusion by exchange, interstitial, and vacancy mechanisms in solids. Prior to this time, there had been little concern with treating diffusional phenomena on a microscopic basis, and most research was concerned with fairly crude observation of overall bulk transfer processes at junctions between regions with strong compositional differences. It was at this time that suggestions on how to carry out high-precision, highly reproducible diffusion experiments were first put forward (Slifkin *et al.* [12], Tomizuka [13]). The three major factors that determine the quality of a diffusion measurement are

1. the method used,
2. the care taken in the measurement, and
3. the extent to which the material is specified [14].

The most accurate method has, in general, been considered to be radiotracer sectioning (Tomizuka [13]), and most of this article is devoted to this method, especially to points for which special care must be taken; these are the measurement of temperature, the accuracy of sectioning, and the reproducibility of counting the radioactivity. The importance of specifying the material cannot be overstated. The measured diffusion coefficient depends on the chemistry and structure of the sample on which it is measured. Impurities, non-stoichiometry of compounds, grain boundaries, and dislocations can give apparent values of the diffusion coefficient that are different from, and usually larger than, the true value (see also [15, 16]). The objective of this chapter is to describe some experimental results as well as their theoretical analysis that are received in last decade. We have organized the chapter around general principles that are applicable to all materials, and then listed the particulars. The materials we consider are mainly inorganic solids, especially semiconductor and insulator materials. The effects of pressure on diffusion is omitted. Previous reviews covering mainly metals and inorganic materials have been given by Hoffman [17], Tomizuka [13], Cadek and Janda [18], Adda and Pholibert [19], Lundy [20], Beniere [21] and last two book of Academic Press [22, 23].

Radioactive tracers are essential to many of the experiments described in this chapter [24–27]. Radioactive tracers are hazardous materials, and the experimenter who uses them is under the strongest moral obligation to avoid exposure of his colleagues and contamination of his environment.

1.2. The relation of diffusion experiments to the mathematics of diffusion

For measurable diffusion to take place a gradient of some kind is necessary. Diffusion is a consequence of the hopping motion of atoms through a solid. The diffusion coefficient D is defined in Fick's first law (Fick [28]),

$$\vec{J} = -D\vec{\nabla}C + C\vec{V} \tag{1}$$

where $\vec{J}$ is the flux of atoms, C their concentration, and $\vec{V}$ the velocity of the center of mass, which moves due to the application of a force such as an electric field or a thermal gradient (see also [29]). A number of different diffusion coefficients exist, e.g., for the diffusion of a radioactive tracer in a chemically homogeneous solid in the absence of external forces,

$$\vec{J}^* = -D^*\vec{\nabla}C^* \tag{2a}$$

where the asterisk denotes the radioactive species. For diffusion in a chemical gradient,

$$\vec{J} = -\tilde{D}\vec{\nabla}C \tag{2b}$$

where $\tilde{D}$ is the interdiffusion or chemical diffusion coefficient. Any of these equations can be combined with the equation of continuity

$$\partial C/\partial t = -\vec{\nabla}\cdot\vec{J} \tag{3}$$

to yield Fick's second law

$$\partial C/\partial t = \vec{\nabla}\cdot(D\nabla C) \tag{4a}$$

where the mass flow term has been omitted. For a tracer in a homogeneous system,

$$\partial C^*/\partial t = -\vec{\nabla}^*\cdot\vec{\nabla}^2 C^*. \tag{4b}$$

Equations 4a and 4b describe the types of diffusion experiments discussed in this chapter.

The tracer diffusion coefficient is given also in the atomistic form

$$D^* = \gamma a^2 \Gamma f \tag{5}$$

where γ is a geometric factor, a the jump distance, Γ the atomic jump frequency, and f the correlation factor (see, e.g. [29]). It is thus possible, in principle, to measure D^* by mesuring Γ in a resonance experiment of some kind [30, 31].

We are concerned here with diffusion measurements where the diffusion coefficient is obtained via Fick's second law, i.e., from a solution of the diffusion equation (see, also [1]). Fick's second law is used rather than his first because concentrations are easier to measure than fluxes and because of D in the solid state are so small that the required steady state is seldom reached. In order to obtain a solution of the diffusion equation, the initial and boundary conditions (IC and BC) must be known. The IC correspond to the distribution of the diffusing substance in the sample before the diffusion anneal, and the BC describe what happens to the diffusing substance at the boundaries of the sample during the diffusion anneal. If the experimental IC and BC correspond to the mathematical conditions, the mathematical solution to the diffusion equation $C(x, y, z, t)$ will describe the distribution of the diffusing substance as a function of position in the sample and of annealing time. The diffusion coefficient is finally obtained by fitting the experimentally determined $C(x, y, z, t)$ to the appropriate solution of the diffusion equation with D as a parameter.

Most laboratory experiments are arranged so that diffusion takes place in one dimension. The solution of the diffusion equation is then $C(x, t)$. One most often determines $C(x)$ at constant t, i.e., the concentration distribution along the diffusion direction after a diffusion annealing time t. It is also possible to determine $C(t)$ at a constant x (e.g., the concentration at a surface) or $\iint C(x, t)\mathrm{d}x\mathrm{d}t$ (e.g., the weight gain of a sample as a function of time). The IC, BC, and solutions to the diffusion equation (for $D = \text{const}$) for some common geometries are described below. These, and solutions for other cases, are given by Crank [32] and Carslaw and Jaeger [33].

1. Thin Layer or Instantaneous Source Geometry (Fig. 1a). An infititesimally thin layer ($\ll(Dt)^{1/2}$) of diffusing substance is deposited on one surface of a semi-infinitive ($\gg(Dt)^{1/2}$) solid. The initial conditions is

$$C(x, 0) = M\delta(x) \tag{6}$$

where δ is the Dirac delta function and M the strength of the source in atoms per unit area. The boundary condition is

$$\partial C(0, t)/\partial t = 0 \tag{7}$$

i.e., there is no flux through the surface. The solution is

$$C(x, t) = (M/\sqrt{\pi Dt})\exp(-x^2/4Dt). \tag{8}$$

One determines $C(x)$ for constant t.

2. Thick Layer Geometry (see Fig. 1b). Similar to the above, except that the layer thickness h is of order of the diffusion distance:

$$\begin{aligned} \text{IC}: C(x,0) &= C_0, \quad h \geq x \geq 0 \\ C(x,0) &= 0, \quad x > h. \end{aligned} \tag{7a}$$

$$\text{BC}: \partial C(0,t)/\partial x = 0 \tag{8a}$$

solution:

$$C(x, t) = \frac{C_0}{2}\left[\mathrm{erf}\left(\frac{x+h}{2\sqrt{Dt}}\right) - \mathrm{erf}\left(\frac{x-h}{2\sqrt{Dt}}\right)\right], \tag{9}$$

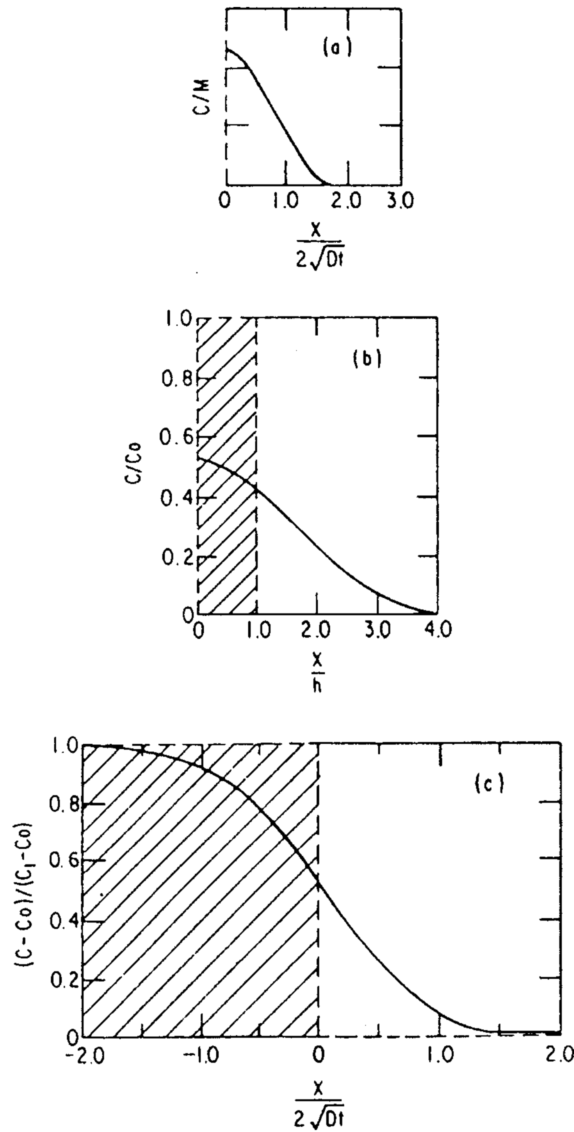


Figure 1 Concentration distributions for different initial conditions. Dotted line is for $t = 0$, solid line is for a finite t. (a) Thin layer geometry [case (1)]; (b) thick layer geometry [case (2)]; solid curve for $Dt = h^2$; (c) infinite couple [case (3)] (after Crank [32]).

where

$$\mathrm{erf}(\lambda) = \frac{2}{\sqrt{\pi}} \int_0^{\lambda} \exp(-\eta^2)\mathrm{d}\eta. \qquad (10a)$$

Measure $C(x)$ for constant t. Note:

$$\mathrm{erfc}(\lambda) \equiv 1 - \mathrm{erf}(\lambda). \qquad (10b)$$

3. Infinite Couple (see Fig. 1c). A sample of uniform concentration C_0 is welded to a sample of uniform concentration C_1. The weld plane is situated at $x = 0$. Couple containing a volatile

$$\text{IC: } C(x,0) = C_1, \quad x < 0$$
$$C(x,0) = C_0, \quad x > 0 \qquad (11)$$

Solution:

$$C'(x,t) \equiv \frac{C(x,t)-C_0}{C_1-C_0} = \left[1 - \mathrm{erf}\left(\frac{x}{2\sqrt{Dt}}\right)\right]. \qquad (12)$$

Measure $C(x)$ for constant t.

4. Vapor-Solid Coule. A semi-infinite couple containing a volatile component component is placed into a dynamic vacuum at $t = 0$:

$$\text{IC}: C(x,0) = C_0, \quad x > 0 \qquad (13a)$$

$$\text{BC}: C(0,t) = 0, \quad t > 0. \qquad (13b)$$

Solution:

$$C(x,t) = C_0 \mathrm{erf}(x/2\sqrt{Dt}). \qquad (14a)$$

Exposing a sample initially devoid of volatile component to a vapor of the volatile componenet at a pressure in equilibrium with C_0 gives the analogous mathematics:

$$\text{IC}: C(x,0) = 0, \quad x > 0 \qquad (14b)$$

$$\text{BC}: C(0,t) = C_0, \quad t > 0. \qquad (15a)$$

Solution:

$$C(x,t) = C_0[1 - \mathrm{erf}(x/2\sqrt{Dt}]$$
$$= C_0 \mathrm{erfc}(x/2\sqrt{Dt}). \qquad (15b)$$

The same equations apply to isotopic exchange between solid and vapor. Measure either $C(x)$ at constant t or integral weight gain (loss)

$$\int_0^{\infty} \int_0^{t} C(x,t)\mathrm{d}t\mathrm{d}x.$$

5. Grain Boundary Diffusion. The mathematics in this case are more complicated (see, e.g. [34]), owing to the coupled lattice diffusion, but one still measures $C(x)$ at constant t.

6. Exchange experiment [35]. This technique is used for materials for which a massive sample cannot be prepared. It involves diffusion exchange between an assembly of powder and a gas of limited volume, from which very small aliquots are drawn at different times.

In the first three sample configurations two bodies of widely different composition are brought into contact. The assumption implicit in the BC is that diffusing material passes across the resulting interface without hindrance, i.e., it is not held up by surface oxides, low solubility, chemical reactions, etc. Nonfulfillment of this condition leads to deviation of the experimental $C(x)$ from solution of the diffusion equation (details see [36]).

In the vapor-solid couple and the exchange experiment, the assumption implicit in the BC is that the surface of the solid equilibrates with the gas phase instantaneously. However, optical measurements of the change of the surface concentration at low temperature have indicated that the attainment of solid-gas equilibrium can be slow process (see also [37]).

In this connection we should add, that the thin geometry has several advantages. The thin layer can be

deposited without straining the sample, which is essential for single crystal samples. A thin layer also allows the use of high specific radioisotopes, and thus measurements of diffusion without a chemical gradient. Diffusion under large chemical gradient can lead to deformation of the sample and generation of defects (see, also [38–41]). For the above reasons, the thin layer geometry is most often used in experiments in which diffusion is measured in order to study the fundamentals of diffusion and defect behavior in solids. Such experiments usually concern diffusion as a function of temperature, pressure, or concentration, and small differences in D are import, in contrast to engineering experiments in which the magnitude of the penetration of one material into another is of interest.

It should be noted that all the solutions to the diffusion equation considered above are expressed in terms of the dimensionless variable $x/(2\sqrt{Dt})$. The length is a kind $2\sqrt{Dt}$ is a kind of mean penetration distance, and this has to be the same order of magnitude as the characteristic distance associated with an experiment. For sectioning experiment, the characteristic distance is the section thickness. For ion-beam depth profiling, it is the ion range, etc. [27].

In the ordinary thin-layer sectioning experiment, one wishes to measure diffusion over a drop in specific activity C of $\sim 10^3$; any effects due to diffusion along short-circuiting paths are likely to show up as curvature in the penetration plot over such a range, while they may not be visible if the measurement is only over factor of 6 in C [42]. Usually twenty sections suffice to define a penetration plot; from Equation 8, the section thickness required to get a drop of 10^3 in C over 20 sections is

$$\theta \approx \sqrt{Dt}/3.8. \tag{16}$$

A preliminary estimate of D is useful in planning an experiment.

If the isotope decays significantly during the time of the experiment, more radioisotope has to be deposited. Under the conditions of $\theta \approx \sqrt{Dt}/3.8$, the specific activity drops by a factor of ~ 2 per section at the twentieth section. These last points on the penetration plot have the greatest weight in determining D, so the counting statistics must be maintained and the penetration plot extended as far as possible. This implies use of an intense source of radioisotope, on the other hand, too much activity poses an unnecessary health hazard as well as increasing the dead-time correction. The radiotracer may rapidly reach the side surfaces of the sample by surface diffusion or evaporation, and then diffuse inward. To keep the diffusion one-dimensional, one removes $\approx 6\sqrt{Dt}$ from the sides of the sample before sectioning.

1.3. Self-diffusion process

As is well-known, in all diffusion mechanisms the atoms under consideration have to carry out jumps between different sites (see, e.g. [43]). If the extreme case of coherent tunneling [44] is left aside, the diffusional jumps are assisted by the thermal movement of the atoms. In the standard situation the jump rate is entirely determined by the temperature T (apart from the effects of hydrostatic pressure, which may be incorporated by formulating the theory in terms of enthalpy and Gibbs free energy). For the purposes of the present chapter, we may in the first approximation disregard quantum mechanical contribution to the diffusity (naturally excluding the self-diffusion in LiH), so that in cubic crystals the diffusion coefficient under standard conditions may be written as an Arrhenius expression

$$D_\alpha = D_{\alpha 0} \exp\big(-H_\alpha^{\mathrm{M}}/kT\big) \tag{17}$$

with the preexponential factor

$$D_{\alpha 0} = g_\alpha a_0^2 \nu_{\alpha 0} \exp\big(S_\alpha^{\mathrm{M}}/k\big) \tag{18}$$

Here H_α^{M} denotes the enthalpy and S_α^{M} the entropy of migration, a_0 the lattice constant, and $\nu_{\alpha 0}$ the attempt frequency, k has its usual meaning as Boltzmann's constant, and g_α is a factor that takes into account the geometry of the crystal structure and the atomistic details of the different process. The subscript α refers to the defect species controlling the diffusion process, i.e., in the case of the direct interstitial mechanism it indicates the chemical nature, geometrical configuration, etc., of the interstitial involved, whereas in the case of indirect diffusion it characterizes the intrinsic defects acting as diffusion vehicles. In the latter case, we should write β instead of α if we wish to indicate that these intrinsic defects are monovacancies or monointerstitials.

The tracer self-diffusion coefficient, i.e., the diffusity of radioactive self-atoms under thermal-equilibrium conditions, is given by (see also [43])

$$D^{\mathrm{T}} = \sum_{\beta=I,V} f_\beta D_\beta^{\mathrm{SD}} = \sum_{\beta=I,V} f_\beta D_\beta C_\beta^{\mathrm{eq}} \tag{19}$$

where

$$C_\beta^{\mathrm{eq}} = \exp\big(S_\beta^{\mathrm{F}}/k\big)\exp\big(-H_\beta^{\mathrm{F}}/kT\big) \tag{20}$$

are the concentrations of self-interstitials ($\beta = I$) and monovacancies ($\beta = V$) in thermal equilibrium. As it is clear, in Equation 19, contributions by clusters of I or V are neglected. The f_β denote correlation factors, $D_\beta^{\mathrm{SD}} \equiv D_\beta C_\beta^{\mathrm{eq}}$ contributions to the uncorrelated self-diffusion coefficient $\sum_{\beta=I,V} D_\beta^{\mathrm{SD}}$, and S_β^{F} and H_β^{F} entropies and enthalpies of formation, respectively. Insertion of Equations 17, 18 and 20 into Equation 19 yields

$$\begin{aligned} D^{\mathrm{T}} &= \sum_{\beta=I,V} D_\beta^{\mathrm{T}} = \sum_{\beta=I,V} f_\beta g_\beta a_0^2 \nu_{\beta 0} \exp\big(-G_\beta^{\mathrm{SD}}/kT\big) \\ &= \sum_{\beta=I,V} D_{\beta 0}^{\mathrm{T}} \exp\big(-H_\beta^{\mathrm{SD}}/kT\big) \end{aligned} \tag{21}$$

with the preexponential factors

$$D_{\beta 0}^{\mathrm{T}} = f_\beta g_\beta a_0^2 \nu_{\beta 0} \exp\big(-H_\beta^{\mathrm{SD}}/kT\big) \tag{22}$$

and the Gibbs free energy of self-diffusion

$$G_{\beta}^{SD} = H_{\beta}^{SD} - TS_{\beta}^{SD} \quad (23)$$

the self-diffusion enthalpy

$$H_{\beta}^{SD} = H_{\beta}^{F} + H_{\beta}^{M} \quad (24)$$

and the self-diffusion entropy

$$S_{\beta}^{SD} = S_{\beta}^{F} + S_{\beta}^{M} \quad (25)$$

The diffusion coefficient D^{s} of foreign substitutional atoms in thermal equilibrium may be derived from Equations 19 or 21 by inserting factor h_{β} under the summation signs. These factors account for the interaction between the intrinsic thermal equilibrium defects and the substitutional atoms. They depend on temperature and the atomic fraction of the substitutional atoms, unless this is small compared to unity. For more detailed and complete discussion in this field we refer the reader to reviews by Frank *et al.* [43](see also [27]).

Compared with metals, self-diffusion in semiconductors is very slow process. For the elemental semiconductors this is illustrated in Fig. 2, in which the self-diffusivities of the cubic semiconductors Si and Ge and of the trigonal semiconductors Te and Se are compared with those of typical metals such as Cu, Ag, and Au on a temperature scale normalized to the melting temperature T_{m}. Fig. 2 reveals the following differences between metals and semiconductors, already emphasized by Seeger and Chik [48].

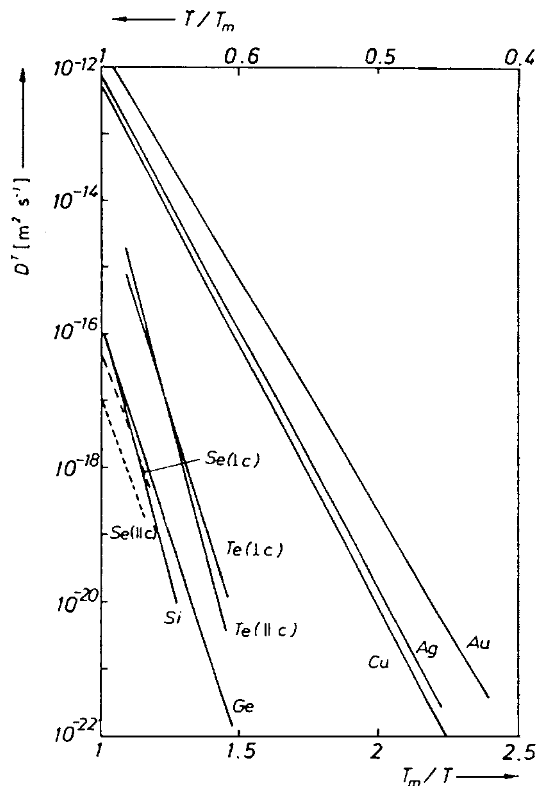


Figure 2 Comparison between the self-diffusivities of the cubic semiconductors Ge and Si (Frank *et al.* [43]), the trigonal semiconductors Te (Ghoshtagore [45], Werner *et al.* [46]) and Se (Brätter and Gobrecht [47]) and the typical metals Cu, Ag, Au (after [43]).

1. Near the melting temperatures the self-diffusion in semiconductors is several orders of magnitude slower than in typical metals.
2. At lower normalized temperatures the ratio of the self-diffusivities of metals and semiconductors becomes even larger.

Generally speaking, the origin of these differences lies in the homopolar bonding of the semiconductors (details see [43]).

As is well-know, the conventional and well-established techniques of determining the tracer self-diffusion coefficient D^{T} based on studying the redistribution of radioactive or stable tracers initially deposited on the specimen surface of serial sectioning methods. In the case of radioactive isotopes, the redistribution may be investigated with radiation detection methods; for stable isotopes—secondary ion mass spectroscopy (SIMS) may be used (see also below).

1.4. SIMS-technique

The most complete description of the experimental technique for study the diffusion processes in solids the readers may be found in the excellent review by Rothman [36]. Here we are briefly discussed the peculiarity of sputtering and SIMS techniques.

We define microsectioning [49] as the cutting of sections a few hundreds nm or less in thickness, so that the surface on which the tracer is deposited, the "front" surface, is not necessary flat on the scale of $\sqrt{Dt}$, and so the thickness of the individual sections are not determined separately. The isoconcentration contours then follow the contour of the front surface, and one must remove sections parallel to this nonflat surface, rather than parallel to a flat surface. If this condition is met and if the indulations in the front surface are gentle (radius of curvature $\rho \gg \sqrt{Dt}$), one can treat the sections as if they were flat (see, however [50]).

Simple chemical dissolution followed by counting the solution has been used for metals [51], alloys [52], and silicate glasses [53, 54]. One uses a constant volume of solvent in the counting vial, rinses the sample so that the rinse flows into the same vial, and then adjusts the total volume of solution so that it is the same for all sections.

In sputtering, material is removed by ion bombardment owing to the transfer of momentum from the bombarding ions to the atoms of the targets. A depth profile can be constructed by (1) analyzing the sputtered-off material in a mass spectrometer (SIMS), (2) collecting and analyzing the sputtered-off material, (3) determining the concentration of the diffusing material in the remaining surface by, e.g., Auger electron spectroscopy (see, e.g. [55–58]), or (4) counting the residual activity of the entire sample [56]. As a rule, noble gas ions, especially Ar^{+}, are accelerated to a few hundred eV or more, with current densities ≤ 1 mA cm^{-2}. This is called physical sputtering, in contrast to bombardment with reactive ions, which is called chemical sputtering. Typical removal rates are of the order of 10 nm min^{-1} for 1 mA cm^{-2} of 500-eV Ar^{+} ions. There are two

excellent reviews of the subject of sputtering (Behrish [60], Chapman [61]), and the reader is referred to these for an understanding of the process. All equipment for sputtering includes a vacuum chamber, pumping equipment, and a controlled gas leak such as a micrometer needle valve. A hohg-speed pumping system is needed as gas is passed continuously and there are bursts of desorbed gases to cope with. Cold-trapped diffusion pumps, cryopumps, or turbopumps have all been used. All sputtering equipment has a gaseous discharge in it. Common glow discharges are not suitable, as too high a gas pressure is required, with resulting low mean-free paths and back diffusion of the sputtered atoms. Therefore, either an ion gun an rf power source is used. Two types of ion sources have been used in sectioning experiments, the custom-modified duoplasmatron of Maier and Schule (see [43]), and a commercially available Kauffman-type gun [62, 63]. Almost any ion source used in ion milling should be usable. The main requirement is that the source put out ion currents ≥ 1 mA cm^{-2} at ~ 1 kV over ~ 4 cm^2 area in a reasonably uniform beam ($\pm 10\%$ except at the very edge), and that the current stay constant over period of several hours. The length of a run is limited by life of the filament. In addition to the ion source, chamber, pumps, and valving, one needs a collector and a sample holder. These are usually custom made. Designs have been given by Gupta and Tsui [64] and Atkinson and Taylor [65] for rf systems as well as Mundy and Rothman [60] for ion gun systems. The collector is either a carousel, with six Al planchets, which allow six sections to be taken before the chamber is opened [67], or a device like a cameras back, on which poleester film is rolled; the latter allows 32 sections to be taken.

In the SIMS technique, the sample is bombarded by reactive ions, and the sputtered-off molecules are ionized in a plasma and fed into a mass spectrometer. The mass spectrum is scanned and the ion current for tracer and host atoms can be recorded simultaneously. The beam is swept over the sample and, in effect, digs a crater, the bottom of which is more or less flat; an aperture prevents ions originating from the edges of the crater from reaching the mass spectrometer (see Fig. 3). The penetration plot is constructed from the plots of istaneous tracer/host atom ratio versus sputtering time and of distance sputtering time. The distance is obtained by using interferometric measurement of the total crater depth under the assumption that material is

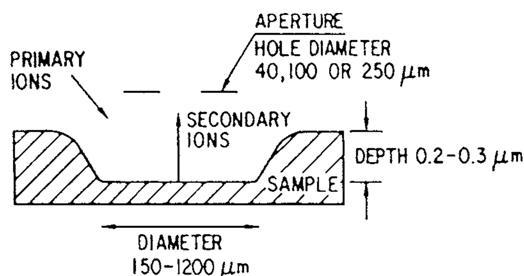


Figure 3 Schematic diagram of crater caused by sputtering in a SIMS apparatus (after Dorner *et al.* [66]).

removed uniformly as a function of time. Large changes of chemical composition along the diffusion direction can invalidate that assumption. The limitations of the SIMS technique have been discussed by Liebl [67] and Reuter and Baglin [68], and a detailed description of its application to diffusion has been given by Seran [69] and Macht and Naundorf [70]. A recent paper [66] shows the quality of results that can be obtained (see also [71]). In general, the resolution is no worse than that obtained by sputtering and counting the sections, and the sensitivity is no worse than that of counting techniques. The major disadvantage of SIMS is its cost. The SIMS apparatus is commercially made (see e.g. [36] and references therein) but represents a large capital investiment. Not withstanding the cost of the apparatus careful contros must be applied to the measurements and artifacts [72] must be avoided.

If the entire of the sample is section the depth of material removed is best determined by weighing on a microbalance. With care, a sample can be weighed to ± 3 μg, which corresponds to ± 150 nm for cross-sectional area of 0.1 cm^2 which is about the minimum useful area, and a density of 2 g cm^{-3}. For larger areas or densities, even better sensitivities are obtained, down to perhaps ± 10 nm (details see review [36] and references therein).

1.5. Self-diffusion of Li and H in LiH crystals

Self-diffusion is the migration of constituent atoms in materials. This process is mediated by native defects in solids and thus can be used to study the dynamics and kinetics of these defects. The knowledge obtained in these studies is pivotal for the understanding of many important mass transport processes such as impurity diffusion in materials. Self-diffusion of $D(H)$ and Li in LiH crystals is studied in papers [73, 74] and [75, 76] respectively. As was above shown the gas-solid isotope exchange method has been used for the measurement of self-diffusion coefficients in solids. Two papers [74, 75] have reported on a thermogravimetric study of the pressure and temperature dependence of diffusion coefficient of the deuteride ion in LiH crystals. As is well-known, in this method a crystal of the compound of interest is equilibrated in a furnace with a gas usually containing an isotopic species of the diffusant. The weight change of the crystal due to the permutation process from the gas to the solid is then monitored as a function of time. By assuming that the mass uptake is due to the isotopic exchange process with deuterium gas and subsequent diffusion of the deuteride ion into the crystal and that the rate of the exchange process is diffusion controlled, the mass gain of the crystal was calculated from solution of Fick's law. The best least squares fit of the data obtained in papers [74, 75] to one-dimensional and three-dimensional models was used to find the diffusion coefficients and the activation energy for the deuteride ion.

According to [74] solution of Fick's second law [29]

$$\partial C/\partial t = \partial^2 C/\partial x^2 \tag{26}$$

subject to the boundary conditions

$$C(x,0) = 0$$
$$C(0,t) = C_s, \tag{27}$$

where C_s is the surface concentration of the diffusing species and

$$\lim C(,t) \rightarrow 0, x \rightarrow 0$$

gives, for Q, the total amount of diffusing substance which has entered the solid at time t,

$$Q = 2A_sC_s(D/\pi)^{1/2}t^{1/2} + B \tag{28}$$

where A_s is the total surface area of the crystal, D is the diffusion coefficient, and B a constant that accounts for the initial condition that $Q \neq 0$ at $t = 0$. Q is expressed as the ratio of the number of moles of deuteride ion diffusing to total moles of LiH contained in the crystal.

Equation 28 may be written as

$$Q = C't^{1/2} + B, \tag{29}$$

where C' is defined as

$$C' = 2A_sC_s(D/\pi)^{1/2}. \tag{30}$$

With the assumption that the surface coverage C_s can be written in terms of an adsorption isotherm, θ, C' was rewritten as

$$C' = 0.8748(A_s\theta/\rho V)D^{1/2}, \tag{31}$$

where ρ is the density of LiH at a given temperature, V is the volume of the crystal, and the constant includes the necessary factors for consistency of units. Rearrangements gives

$$C = C'(A_s/\rho V)^{-1} = 0.8748\theta D^{1/2}. \tag{32}$$

The values for C' were found from the fit of Equation 29 to the data at each temperature (see, Fig. 4). All fits were exceptionally good (see also Table I). Values of C calculated from C' by using the relation given in Equation 26 are also given in Table I.

Equation 32 may be written in the form

$$C = 0.8748\theta D_0^{1/2}e^{-E_a/2RT} \tag{33}$$

by substituting for D,

$$D = D_0 e^{-E_a/2RT}, \tag{34}$$

where D_0 is a constant and E_a the activation energy.

According to Equation 33, the temperature variance of C is determined by the exponential term involving temperature and the temperature dependence of the fractional surface coverage θ. If θ is known as a function of temperature, C may be fit to Equation 33 to give D_0 and the activation energy. Lacking knowledge of the exact variation of θ with temperature, Equation 27 may be rewritten in the more convenient form

$$C/\theta = 0.8748D_0^{1/2}e^{-E_a/2RT} \tag{35}$$

and θ assumed to vary with temperature according to the expected behavior outlined above. The quantity C/θ was next calculated using $\theta = T/(b+T)$ and the fit of the data to Equation 29 found, giving the activation energy as 22.1 ± 3.0 kcal and D_0 as 4.01×10^{-3} cm^2 s^{-1}. The avctivation energy thus found is in good agreement with the value obtained by Funkee and Richtering [77] from NMR measurements, 22 ± 2 kcal.

In three-dimensional bulk diffusion model the solution to Fick's law [28] for a finite solid has the next relation

TABLE I Summary of one-dimensional semi-infinite solid data (after [75])

Sample no.	T°C	$C' \times 10^3$, min^{-1}	$C \times 10^4$ g cm^{-2} min$^{-1/2}$	E_a, kcal
8	550	4.74	3.96	22.7 ± 2.8[a]
9	524	4.37	3.49	22 ± 2[b]
10	500	2.80	2.52	
11	450	1.50	1.53	
12	410	1.37	0.96	
13	399	1.09	0.86	

[a] E_a as found from the best least squares fit of Equation 28 to C with $\theta = 0.80$. The error reported is the 95% confidence level fit to the data.
[b] E_a from Ref. [77].

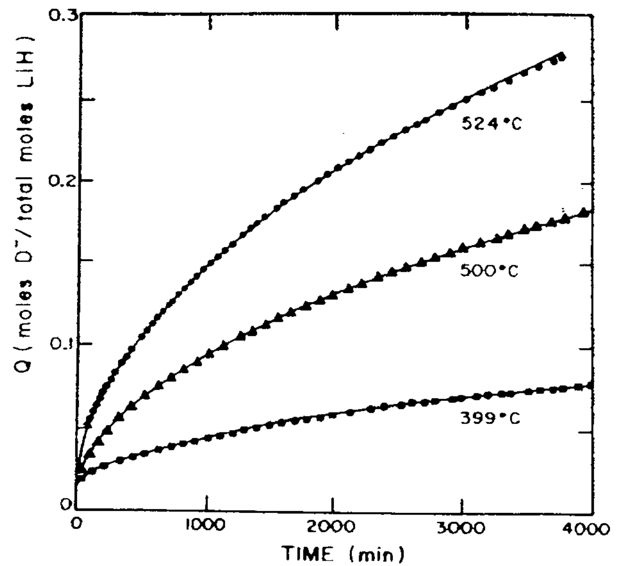


Figure 4 Q as a function of time for data of samples 9, 10 and 13. The solid curves are the best least squares fit of Equation 29 to the experimental data (after [75]).

$$Q = \theta\left\{1 - \left[\sum_{u=0}^{\infty}\frac{8}{(2u+1)^2\pi^2}\exp\left(\frac{-(2u+1)^2\pi^2Dt}{4a^2}\right)\right] \times \left[\sum_{v=0}^{\infty}\frac{8}{(2u+1)^2\pi^2}\exp\left(\frac{-(2v+1)^2\pi^2Dt}{4b^2}\right)\right] \times \left\{\sum_{w=0}^{\infty}\frac{8}{(2w+1)^2\pi^2}\exp\left(\frac{-(2w+1)^2\pi^2Dt}{4c^2}\right)\right\} + B\right\}, \tag{36}$$

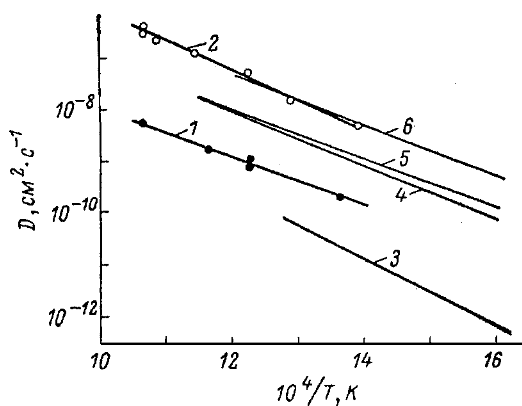


Figure 5 Temperature dependence of self-diffusion coefficients Li (1, 3) and H (2, 4–6) in LiH single crystals; 1, 2 [76]; 3, 6—[77]; 4,5—[75] (after [76]).

where the symbols are as previouly defined and 2a, 2b and 2c are the dimensions of the crystal in the x, y and z directions. The best fit of the experimental data of papers [74, 75] gives for the diffusion coefficient as a function of temperature, $D = 2.41 \times 10^{-2} e^{-24.3\times 10^3/RT}$ $cm^2\,s^{-1}$. The diffusion coefficient for deuteride ion in lithium hydride at 465°C reported in paper [74] to have been found from three dimensional model was $(1.9 \pm 0.6) \times 10^{-9}$ $cm^2\,s^{-1}$ (see also Fig. 5). The diffusion coefficient calculated from Equation 36 was 1.6×10^{-9} $cm^2\,s^{-1}$. A comparison of tracer diffusion coefficients as calculated from one- and three dimensional models is given in Table II, and a plot of $-\log D$ versus $10^3/T$ is given in Fig. 6.

The activation energies for various migrating species have been theoretically calculated by Dellin *et al.* [78]. They find activation energies for interstitial H^- diffusion to lie in the range 11.5 to 23 kcal, while activation energy for H^- vacancy migration is calculated to be 2.3 kcal. From these calculations interstitial H^- migration would seem possible based on the activation energy of about 24 kcal found by Spencer *et al.* [75]. However, Dellin *et al.* [78] in agreement with Pretzel *et al.* [79] find that interstitial H^- is an unstable species in LiH and thus could not be the difffusing species.

To the conclusion of this part, we should indicate once more that the activation energy found by Spencer *et al.* [75] ~24 kcal, is in excellent agreement with the 22 kcal determined for H^- sel-diffusion in LiH by NMR [77]. This agreement, plus the consistencies of

TABLE II Comparison of tracer diffusion coefficients as found from one- and three-dimensional models (after [75])

Sample no.	$T°$, C	1-D[a] $\times 10^9$ $cm^2\,s^{-1}$	3-D[b] $\times 10^9$ $cm^2\,s^{-1}$	3-D[c]
8	550	5.8	8.7	0.2 ± 0.1
9	524	3.7	5.3	
10	500	2.3	3.4	
11	450	0.84	1.1	
12	410	0.34	0.41	
13	399	0.26	0.31	

[a]Calculated from Equation 28.
[b]Calculated from Equation 30.
[c]3-D is the data from [76].

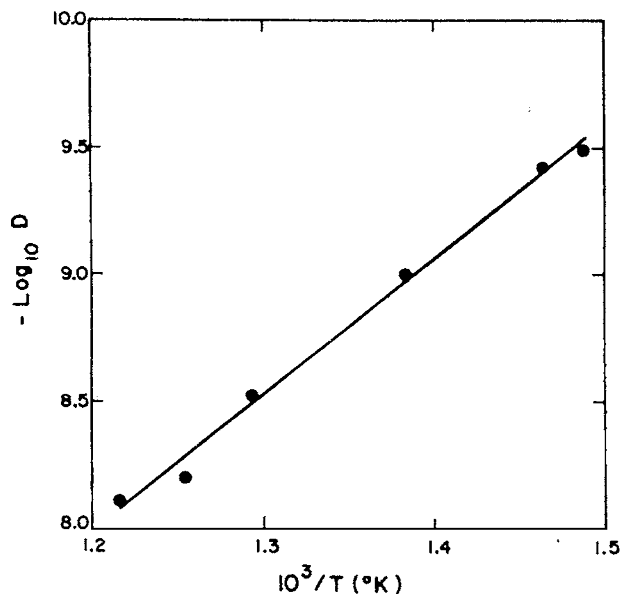


Figure 6 Arrhenius plot of the diffusity of D^- in LiH in the temperature region 400–550°C for $\theta = 0.8$. The activation energy as determined from this plot is 24.3 ± 2.6 kcal. (after [75]).

previous work [74], make D^- vacancy migration still the most likely species and mode of migration (see also [76]).

1.6. Self-diffusion in intrinsic Ge

In intrinsic germanium the temperature dependence of the tracer self-diffusion coefficient of the radioactive isotope ^{71}Ge has been measured by several groups [80–85] by means of different techniques (see Fig. 7). With the exception of the latest experiments, precision grinding techniques were used to remove sections with thickness of the order of 1 μm from the diffusion zone of the annealed specimens. As a consequence, the temperature range covered by the earlier experiments is rather limited. By means of a sputtering technique for serial sectioning [84, 85] have been able to extend the range of self-diffusion studies in Ge to diffusivities as low as 10^{-22} $m^2\,s^{-1}$.

The overall agreement between Ge self-diffusion data of different authors is good. In the region of overlap a small difference between the data of Vogel *et al.* [84] and those of the earlier workers may be seen. We tend to attribute this to problems in determining small diffusion coefficients during the earlier work. Widmer and Gunther-Mohr [82] used Gruzin's [86] or Steigmann's [87] methods, both of which are known to be less reliable than the layer-counting method since these methods require a precise knowledge of the absorption coefficient of the radiation involved. In the work of Valenta and Ramasastry [81], the condition $\delta \ll \sqrt{D^T t}$ (δ — thickness of the deposited tracer layer) was not always fulfilled. Since, nevertheless, these authors used the thin-film solution of the diffusion equation to deduce tracer diffusion coefficients, the obtained values are likely to be somewhat larger than the true D^T values. As may be seen in Fig. 7, the temperature dependence of the D^T data of Ge is well described by an Arrhenius law (the preexponential factors D_0^T and the self-diffusion enthalpies H^{SD} obtained from measurements

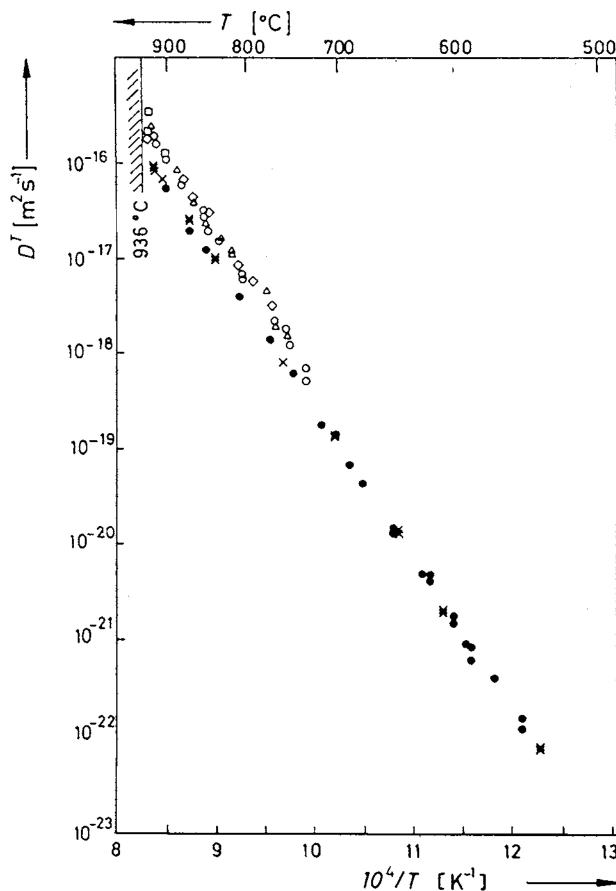


Figure 7 Tracer self-diffusion coefficient of Ge as a function of temperature: ◇—[80]; △—[81]; ○—[82]; □—[83]; ●—[84]; *x*—[85] (after [43]).

of different authors are compiled in Table III). Seeger and Chik [88] argued that this result may be accounted for in terms of an indirect self-diffusion mechanism involving one type of intrinsic defect. Guided by further observations, they suggest that it is the vacancy mechanism (details see [43]) that controls self-diffusion in Ge. Table III shows that the preexponential factor D_0^T of Ge is considerably larger than the D_0^T values typical for metals (10^{-6} m^2 s^{-1} $\lesssim D_0^T \lesssim 10^{-4}$ m^2 s^{-1}) [43]. Arguing that for an ordinary mechanism the product $f_v g_v a_0^2 \nu_{v0}$ in Equation 16 for D_{v0}^T ($\equiv D_0^T$) should be of the same order of magnitude for Ge and metals, Seeger and Chik [88] interpreted the large D_0^T value of Ge in terms of a large self-diffusion entropy of the vacancy in Ge, $S_v^{SD} \approx 10$ k. They suggested that this large S_v^{SD} value arises from a spreading out of the vacancy over several atomic volumes.

TABLE III Self-diffusion data for germanium and silicon (after [43])

Element	D_0^T (10^4 m^2 s^{-1})	H^{SD} (eV)	Temperature range (K)	Technique	References
Ge	7.7	2.95	1039–1201	SG	[80]
	32	3.1	1023–1143	SG	[81]
	44	3.12	1004–1188	SM + GM	[82]
	10.8	2.99			
	24.8	3.14	822–1164	SS	[84]
	13.6	3.09	808–1177	SS	[85]
	1.2×10^{-3}	3.05	543–690	SIMS	[97]
Si	1800	4.77	1473–1673	HL	[389]
	1200	4.72	1451–1573	CS	[90]
	9000	5.13	1373–1573	ES	[91], [92]
	1460	5.02	1320–1660	SS	[93]
	8	4.1	1173–1373	R	[94]
	154	4.65	1128–1448	SIMS, ^{30}Si	[95]
	20	4.4	1103–1473	R	[96]

SG = sectioning by grinding; SAM = Steigmann's method; GM = Gruzin's method; SS = sectioning by sputtering; HL = hand lapping; CS = chemical sectioning, n activation of ^{30}Si; ES = electrochemical sectioning; R = (p,γ) resonance of ^{30}Si.

As we can see from Table III the published value of fundamental quantities such as the diffusion coefficient vary by several orders of magnitude for various authors (see also [43, 88]). Such a spread in the experimental data makes it difficult to determine conclusively the underlying physical processes. Reliable diffusion data are therefore crucial to clarify the diffusion mechanisms and to accurately determine the corresponding material parameters. The conventional technique (see, e.g., Table III) to determine the self-diffusion coefficient D^{SD} in semiconductors is to deposit thin layer of radioactive tracer on the surface of the crystal (e.g., ^{71}Ge, ^{31}Si). In a subsequent annealing step the tracers diffuse into the crystal. The depth profile of the tracer atoms is then determined by serial sectioning and measurements of the corresponding radioactivity. There are several experimental difficulties arising from this method (see also [97]).

1. Traditionally, lapping and grinding was used for the serial sectioning. This requires that the mean penetration distance $(D^{SD}t)^{1/2}$ of the tracer atoms during the time t of a diffusion anneal has to be in the μm range. Especially in silicon, the large distance and the short half-life (2.6 h for ^{31}Si) limit this method to be applicable only to higher temperatures (larger D^{SD}). Germanium is more convenient in this respect (the half-life period of ^{71}Ge is 11.2 days), but it was not until microsectioning technique (e.g., sputtering) were invented that the measurements could be extended to lower temperatures in recent years (see [285]).
2. Surface effects such as oxidation, contamination, strain, etc. might influence the tracer diffusion substantially (e.g., through the formation of intrinsic defects).

Fuchs *et al.* [97], recently reported results of a very accurate method to measure the self-diffusion coefficient of Ge which circumvents many of the experimental problems encountered in the conventional methods. These authors used germanium isotopic heterostructures (stable isotopes), grown by molecular-beam epitaxy (MBE) (details see [98–100]). As is well-known, isotope heterostructures consist of layers of pure (e.g., ^{70}Ge, ^{74}Ge) or deliberately mixed isotopes of a chemical element. Fig. 8 shows the schematic of the particular samples used by Fuchs *et al.* [97]. At the interface only the atomic mass is changing, while (to first order) all the other physical properties stay the same. In the as grown samples, this interface is atomically flat with layer thickness fluctuations of about two atomic ML (details see [99]). Upon annealing, the isotopes diffuse into each other (self-diffusion) with a rate which depends strongly on temperature. The concentration

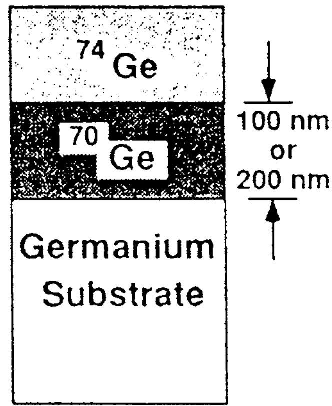


Figure 8 Schematic of the isotope heterostructure used by Fuchs *et al.* (after [97]).

profiles in paper [97] were measured with SIMS, after pieces of the same samples have been separately annealed at different temperatures. This allows an accurate determination of the self-diffusion enthalpy as well as the corresponding entropy. The isotopic heterostructures are unique for the self-diffusion studies in several aspects (see also [1]).

1. The interdiffusion of germanium isotopes takes place at the isotopic interface inside the crystal, unaffected by possible surface effects (e.g., oxidation, impurities and strain) encountered in the conventional technique.

2. One sample annealed at one temperature provides five more or less independent measurements (Ge consists of five stable isotopes). Their initial respective concentrations vary for the different layers of the as-grown isotope heterostructure. After annealing, the concentration profile of each of the five isotopes can be analyzed separately to obtain five data points for each annealing temperature. The samples were cut into several pieces. One piece was kept in paper [97] for reference (as-grown), the were separately annealed at five different temperatures (543, 586, 605, 636, and 690°C). The temperature controller permitted a variation of the temperature of 1–2°C. The recording of the concentration depth profiles of all five stable Ge isotopes was performed with SIMS. The oxygen primary beam had an impact energy of 8 keV per incident ion. The beam was rastered over a square area of about 200 μm in size and the detected secondary ions extracted from the central 30 μm diameter region of the crater. The precision of the SIMS data was estimated to be within ±5%. The depth resolution of the system was determined from profiles taken from the as-grown samples with an atomically flat interface. What theoretically should be a step function in the concentration profile appeared as a slope of about 4 nm per decade of the measured atomic fraction at the leading edge of a layer, and about 16 nm per decade at the falling edge (details see [97]).

As is well-known, diffusion in the crystals occurs through jumping thermally activated between different sites in the lattice [15, 16]. In principle, there are many possibilities for such jumps (substitutional or interstitial sites, vacancies, etc. (details see [43])). In Ge crystals, however, it is known that the only process of significance for the migration of germanium atoms is through the vacancy mechanism (see also [88]) In this case the self-diffusion coefficient D^{SD} can be written as as Arrhenius expression [97] (see Equation 22)

$$D^{\mathrm{SD}} = gfa^2\nu_0 \exp\left[\frac{-G^{\mathrm{SD}}}{kT}\right] = D_0 \exp\left[\frac{-H^{\mathrm{SD}}}{kT}\right], \quad (37)$$

wher G^{SD} is the Gibbs free energy of self-diffusion,

$$G^{\mathrm{SD}} = H^{\mathrm{SD}} = TS^{\mathrm{SD}}, \quad (38)$$

H^{SD} is the self-diffusion enthalpy, and the preexponemntial factor

$$D_0 = gfa^2\nu_0 \exp\left[\frac{S^{\mathrm{SD}}}{k}\right] \quad (39)$$

contains the self-diffusion entropy S^{SD}, the correlation factor f ($f = 1/2$ for the vacancy mechanism in the diamond lattice [101], the attempt frequency ν_0, the geometric factor g ($g = 1/8$ for vacancies in Ge and the lattice constant a, k is Boltzmann's constant (see also part 1.3) The enthalpy H^{SD} and the entropy S^{SD} depend on the formation (subscript F) as well as the migration (subscript M) of the vacancy:

$$H^{\mathrm{SD}} = H_{\mathrm{F}}^{\mathrm{SD}} + H_{\mathrm{M}}^{\mathrm{SD}} \quad \text{and} \quad S^{\mathrm{SD}} = S_{\mathrm{F}}^{\mathrm{SD}} + S_{\mathrm{M}}^{\mathrm{SD}}. \quad (40)$$

The quantity which we can extract from the data of paper [97] is primarily the self-diffusion coefficient D^{SD} as a function of annealing temperature T. This was done in citing paper by fitting of experimental depth profiles to theory, with D^{SD} being the only fitting parameter. Equation 31 then allows to determine the self-diffusion enthalpy H^{SD}, and the self-diffusion entropy S^{SD} is deduced using Equation 40. Solving Fick's diffusion equation for the specific geometry of samples used in indicated paper (see Fig. 8), these authors obtain the atomic fraction c_{i} of a given germanium isotope i in terms of error functions (erf) (see Equation 36):

$$c_{\mathrm{i}}(x) = \left\{\frac{c_{\mathrm{i}}^{0,\mathrm{I}} - c_{\mathrm{i}}^{0,\mathrm{II}}}{2}\mathrm{erf}\left[\frac{h/2 + x}{2\sqrt{D^{\mathrm{SD}}t}}\right] + c_{\mathrm{i}}^{0,\mathrm{I}}\right\} + \left\{\frac{c_{\mathrm{i}}^{0,\mathrm{II}} - c_{\mathrm{i}}^{0,\mathrm{III}}}{2}\mathrm{erf}\left[\frac{h/2 - x}{2\sqrt{D^{\mathrm{SD}}t}}\right] + c_{\mathrm{i}}^{0,\mathrm{III}}\right\}, \quad (41)$$

where h is the layer thickness (110 or 200 in Ge samples, see Fig. 8), and $c_{\mathrm{i}}^{0,\mathrm{I}}$, $c_{\mathrm{i}}^{0,\mathrm{II}}$, and $c_{\mathrm{i}}^{0,\mathrm{III}}$ are the initial concentrations of the isotope i in the enriched ^{74}Ge layer, in the enriched ^{70}Ge layer, and in the substrate, respectively. Fig. 9 shows the profiles of all five isotopes of an annealed sample (586°C for 55, 55 h), together with a fit of the data to Equation 41. For clarity only the fit to the ^{70}Ge profile is shown, but other profiles can be independently fitted as well. The excellent quality of the fit over four orders of magnitude displays

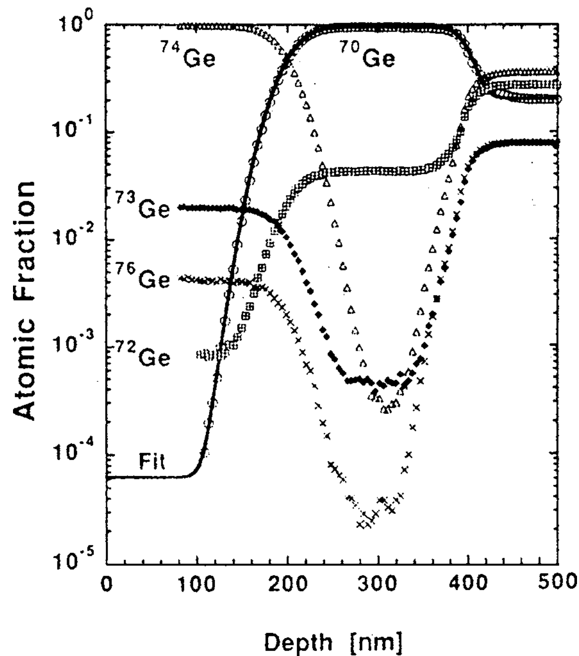


Figure 9 Experimental depth profile of the atomic fraction of ^{70}Ge, ^{72}Ge, ^{73}Ge, ^{74}Ge and ^{76}Ge (symbols) of a diffusion annealed sample (annealed at 586°C for 55.55 h). The solid line is a fit of the ^{70}Ge data of Equation 41. For clarity, only the fit to the ^{70}Ge data is shown (after [97]).

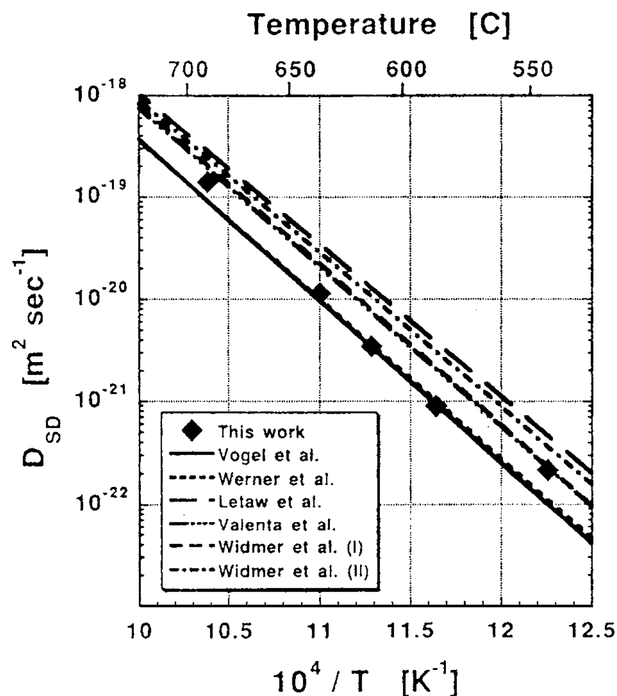


Figure 11 Arrhenius plot of the self-diffusion coefficient as a function of temperature. Data of Fuchs *et al.* [99] agrees favorably well with the most recent data ([84, 85]). The older data [80–82] might be less accurate (after [97]).

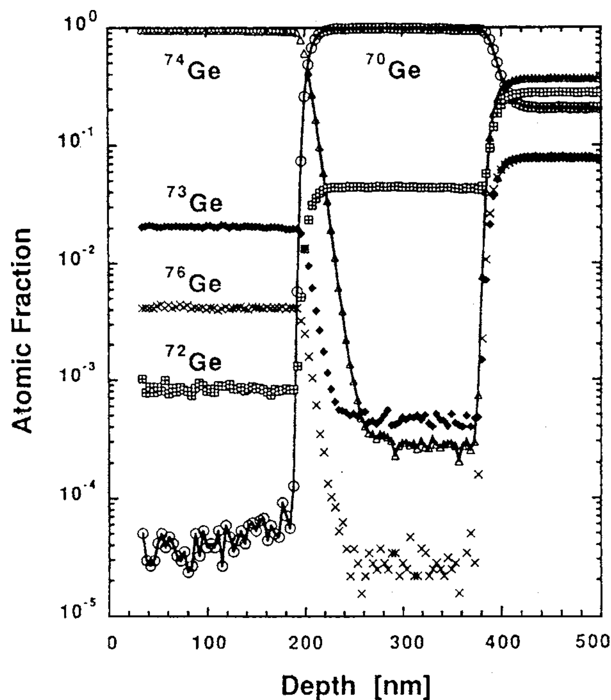


Figure 10 Experimental depth profiles of the same sample as Fig. 9, but before annealing (after [97]).

the remarkable accuracy of the method used by Fuchs *et al.* As a reference, the corresponding concentration profiles for as-grown sample are displayed in Fig. 10. The annealing time was purposefully chosen such that the plateaus in the annealed samples (around 300 and 100 nm) correspond to the original concentrations in the isotopically enriched layers.

The values for the self-diffusion coefficient D^{SD} obtained at 543, 586, 605, 636 and 690°C are presented in an Arrhenius plot in Fig. 11. The lines in Fig. 11 represent the results of previous authors [43]. The variation in D^{SD} obtained from different groups is comparable with the scatter of the data within the work of each of the publications. Fitting the experimental values of D^{SD} to Equation 37 Fuchs *et al.* obtain the self-diffusion enthalpy H^{SD} equals 3.0(5) eV. As can see from Table III this in excellent agreement with previously published values of 2.95–3.14 eV. The value of experimental preexponential factor D_0 is 1.2×10^{-3} m^2 s^{-1}. This compares to previously published values of $(0.78–4.4) \times 10^{-3}$ m^2 s^{-1}. Converting D_0 into the self-diffusion entropy S^{SD} through Equation 39 they obtain $S^{SD} \approx 9$ k (using $\nu_0 = 8 \times 10^{12}$ s^{-1} and $a = 0.565$ nm). The self-diffusion entropy for Ge is larger than for metals (2–4) k. As an explanation, Seeger and Chik [88] invoked the idea of extended (spread-out) defects, and Bourgoin and Lanoo [102] have proposed that vacancy in Ge is strongly relaxed.

Finally, we want to mention the effect of the isotopic mass on the self-diffusion coefficient (see also [43]). The many-body treatment of atomic jump processes leads to an expression for the strength of the isotope effect in terms of the correlation factor f of Equation 19 and the fraction ΔK of the kinetic energy which is associated with the motion in the jump direction [103, 104].

$$\frac{\left[\frac{D_{\mathrm{I}}}{D_{\mathrm{II}}}\right]-1}{\left[\frac{m_{\mathrm{II}}}{m_{\mathrm{I}}}\right]^{1/2}-1} = f\Delta K. \tag{42}$$

In previous Ge self-diffusion experiments, Campbell [83] found $f\Delta K$ values between 0.26 and 0.30, which translates into a ratio of $D_{\mathrm{SD}}^{70\mathrm{Ge}}/D_{\mathrm{SD}}^{74\mathrm{Ge}}$ between 1.007

and 1.008 [105]. This small difference, however is below the precision of the Fuchs *et al.* work. When fitting the experimental depth profilies to Equation 41, they could indeed not detect any appreciable difference between the different isotopes. In addition, such small deviations would be insignificant in the Arrhenius plot (logarithmic scale of D^{SD} in Fig. 11) for the determination of the self-diffusion enthalpy H^{SD} and entropy S^{SD}.

1.7. Self- and interdiffusion of Ga and Al in isotope pure and doped heterostructures

Self-diffusion is the most fundamental matter transport process in solids. Understanding this process is pivotal to understanding all diffusion phenomena in solids, including those for native defects and impurities. As was noted above, compared to metals [106], self-diffusion processes in semiconductors are significantly more complex (see also [10, 11, 29, 107, 108]). This is due to the much richer spectrum of native defects and to the much larger effects of small concentrations of defects on the Fermi level position and other properties [43]. In III–V compounds, experiments are more difficult to perform because of the high partial vapor pressure of the group-V elements and the dependence of native defect species and concentrations on stoichiometry [109].

Over the past thirty years, there have been only a few attempts to directly study Ga self-diffusion in GaAs using isotopes [110–112]. Goldstein [110] and Palfrey *et al.* [111] diffused radioactive ^{72}Ga into bulk GaAs at elevated temperatures and obtained depth profiles of ^{72}Ga by mechanical sectioning and radioactive assaying. With a rather limited temperature range investigated, they reported activation enthalpies of 5.6 and 2.6 eV, respectively. Tan *et al.* [112] studied the disordering of ^{69}GaAs/^{71}GaAs isotope superlattice structure and found an activation enthalpy of 4 eV. However, arguing that the heavily Si-doped substrates in their samples affected the result, these authors discarded this value in favor of their earlier estimate of a 6 eV activation enthalpy [109]. In view of this controversy and the fact that our knowledge on self-diffusion in GaAs is primarily derived from studies of interdiffusion (see also below) of Ga and Al in $Ga_xAl_{1-x}As$ systems [113, 115], it is of great interest to study diffusion of Ga isotopes with some new approaches.

Wang *et al.* [114] reported results from Ga self-diffusion studies in GaAs. They measured concentration profiles of ^{69}Ga and ^{71}Ga using SIMS and determined the activation enthalpy and entropy by analyzing the diffusion coefficients obtained between 800 and 1225°C as well as examined effects of substrate doping. ^{71}GaAs and ^{69}GaAs layers of 200 nm each were grown using molecular beam epitaxy (MBE) at 580°C on GaAs substrates of natural isotope composition (^{69}Ga: ^{71}Ga $= 60.2{:}38.8$) The nominal isotope purity in the epilayers is 99.6%.

Fig. 12 shows the SIMS depth profiles of ^{69}Ga and ^{71}Ga after annealing at $T = 974$°C for 3321 s. Excellent fits over 2.5 orders of magnitude in concentration are obtained for this and all the other depth profiles taken from samples with smooth surfaces.

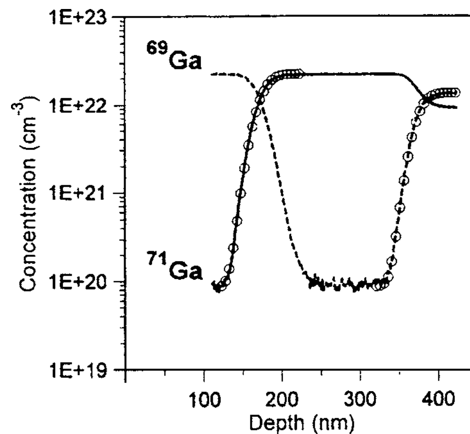


Figure 12 SIMS depth profiles of ^{69}Ga and ^{71}Ga in GaAs isotope epilayers annealed at 974°C for 3321 s. The circles are theoretical fits (after [114]).

The characteristic diffusion length R is defined as

$$R = 2\sqrt{Dt}, \qquad (43)$$

where D (as usually) is the Ga self-diffusion coefficient and t is the annealing time. The Ga self-diffusion coefficient D can be derived from Equation 37. An Arrhenius plot for D is presented in Fig. 14 (see below). The D values span 6 orders of magnitude in the temperature range from 800 to 1225°C.

Expanding the research on Ga self-diffusion beyond GaAs to other III–V compound semiconductors can be quite instructive in elucidateng the microscopic mechanism. This was the main reason an investigation [116] on Ga in gallium phosphide (GaP) using the same SIMS technique. For the experiment, ^{71}GaP and ^{69}GaP epitaxial layers 200 nm thick were grown by solid source molecular beam epitaxy (MBE) at 700°C on undoped GaP substrates. The isotope composition was the same as in GaAs. Fig. 13 shows the SIMS profiles (solid

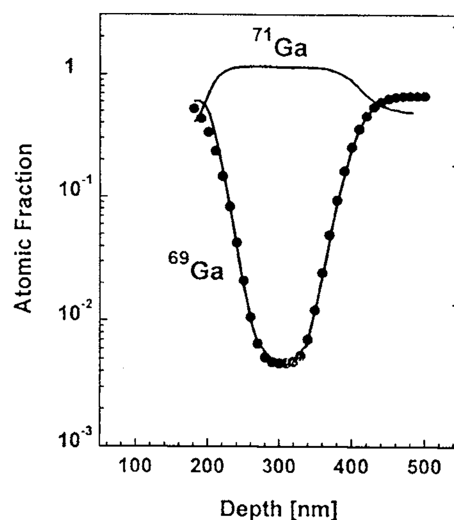


Figure 13 SIMS depth profiles of ^{69}Ga and ^{71}Ga in GaP isotope epilayers annealed at 1111°C for 231 min. The filled circles represent the calculated ^{69}Ga concentration profile (after [116]).

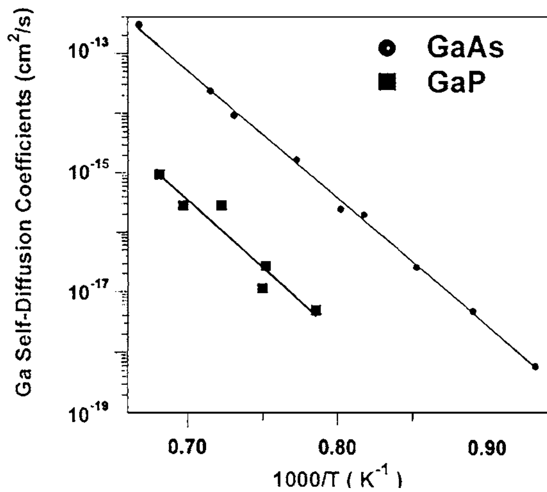


Figure 14 Arrhenius plots of Ga self-diffusion coefficients in GaAs (filled circles) and GaP (filled squares) (after [116]).

lines) and the calculated $C(x)$ of ^{69}Ga (circles) and ^{71}Ga (continuous line) in a sample annealed at 1111°C for 3 h and 51 min. Excellent agreement is obtained between the measured and the calculated profiles over two and a half orders of magnitude in concentration. This agreement strongly supports the assumptions made in Equation 12. The fitting procedure leads to an accurate determination of D through Equation 43.

In Fig. 14, D is plotted versus temperature T. The D values span over two orders of magnitude in the temperature range from 1000 to 1190°C. They can represented by Equation 17. From this the cited authors are determined the activation enthalpy H^{SD} and the preexponential factor D_0 to be 4.5 eV and 2.0 cm^2 s^{-1}, respectively. The self-diffusion entropy can be obtained used the Equation 39. In GaP $g \sim 1$, $a = 5.45$ Å, and $\nu_0 = 1.2 \times 10^{13}$ Hz. Using these values Wang *et al.* [116] obtained $S^{\mathrm{SD}} = 4$ k. The Ga self-diffusion coefficients in GaAs from [114] are also shown in Fig. 14 for comparison. The activation enthalpy and entropy for GaAs are 4.24 eV and 7.5 k, respectively.

As in the case of GaAs, the Ga self-diffusion coefficients in GaP follow an Arrhenius relation described by Equation 37, indicating that a single type of native defect is most likely responsible for mediating the Ga self-diffusion in GaP over studied temperature range. In intrinsic GaAs, the defect mediating Ga self-diffusion has been ascertained to be the triply negatively charged gallium vacancy acceptor, V_{Ga}^{3-} [109]. Such assignment may still be premature for GaP. From a recent positron annihilation study in GaP, Krause-Rehberg *et al.* [117, 118] reported that positron trapping by vacancies behaves similarly in GaP as in GaAs. Vacancies are detected by these authors only in n-type GaP at room temperature, with a detection limit of 2×10^{15} cm^{-3}. These findings neither support nor exclude the possibility that in intrinsic GaP it is also the acceptor-like V_{Ga} that mediates Ga self-diffusion.

Fig. 14 shows that the Ga self-diffusion coefficient in GaP is about two orders of magnitude lower than that in GaAs. The decreased cation diffusion in III–V compounds, with phosphorus replacing arsenic as anion, has been previously observed [117, 118]. Interdiffusion of Ga and Al was determined to be two orders of magnitude slower in the AlGaInP/GaInP superlattice system than in the AlGaAs/GaAs system (details see below). The change has been attributed to the stronger Ga—P bond compared to the Ga—As bond [118]. Although Wang *et al.* [116] measured a higher activation enthalpy in GaP (4.5 eV) than in GaAs (4.24 eV), the difference is not large enough to be outside the experimental uncertainty. More reliable is the difference in the preexponential factor D_0, or the entropy term S^{SD}, between GaP (4 k) and GaAs (7.5 k). As was shown above, this entropy term is the sum of the formation entropy S^{F} and migration entropy S^{M} for the native defect mediating the self-diffusion. S^{F} or S^{M} represents the number of equivalent formation configurations or migration jumps. The significant difference in S^{SD} indicates profound variations in the way that the mediating native defects are formed or migrate in GaP as compared to GaAs (see also [1]). The small value of S supports a simple native defect species as the major diffusion vehicle (details see [114]).

Bracht *et al.* [115] were used three undoped isotope heterostructures of $\mathrm{Al}_x^{71}\mathrm{Ga}_{1-x}\mathrm{As}/\mathrm{Al}_y^{69}\mathrm{Ga}_{1-y}\mathrm{As}/{}^{71}\mathrm{GaAs}$ with $(X, Y) = (0.41, 0.62)$ (a); (0.62, 0.85) (b); and (0.68, 0.88) (c) and in addition an AlAs/^{71}GaAs (d) layer structures for Ga self- and Al-Ga interdiffusion experiments. The thickness of the layers lie between 100 and 200 nm. The structure were grown by MBE at about 600°C on a 200 nm thick undoped natural GaAs buffer layer which was deposited on (100) oriented GaAs substrate wafers. *A* natural GaAs capping layer, about 200 nm thick, was grown om top of the structure to protect the AlGaAs layer against oxidation in air. Concentration profiles of Al, ^{69}Ga, ^{71}Ga, and As in the annealed samples were measured with SIMS using a Cs$^+$ ion beam with an energy of 5.5 keV. The depth of the craters left from the analysis were determined with a Tencor P-10 surface profilometer. The measured secondary ion counts were converted into concentrations taking into account the count rates obtained on an $\mathrm{Al}_{0.56}\mathrm{Ga}_{0.44}\mathrm{As}$ standard.

Concentration profiles of Al, ^{69}Ga, and ^{71}Ga of the as-grown structure b are shown in Fig. 15a. Fig. 15b illustrates the corresponding distribution after annealing of sample b at 1050°C for 1800 s. Concentration profiles of ^{69}Ga (see also [115]) which lie within the Al^{71}GaAs and ^{71}GaAs layers, are accurately described by the solution of Fick's law (see above) for self-diffusion across an interface taking into account a concentration-independent diffusion coefficient. The measured ^{69}Ga profiles within the ^{71}GaAs layer of samples a to d all yield the same Ga self-diffusion coefficient within experimental accuracy even though the Al concentration in the ^{71}GaAs layer varies from 10^{18} cm^{-3} (detection limit) in e.g., sample d up to 10^{20} cm^{-3} in sample b due to in-diffusion of Al from the adjacent AlGaAs layer. The temperature dependence of Ga self-diffusion in $\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{As}$ with $X = 0$, 0.41, 0.62, and 0.68 is shown in Fig. 16. The activation enthalpy H^{SD} of Ga self-diffusion in $\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{As}$ and the corresponding preexponential factor D_0 are summarizes in Table IV.

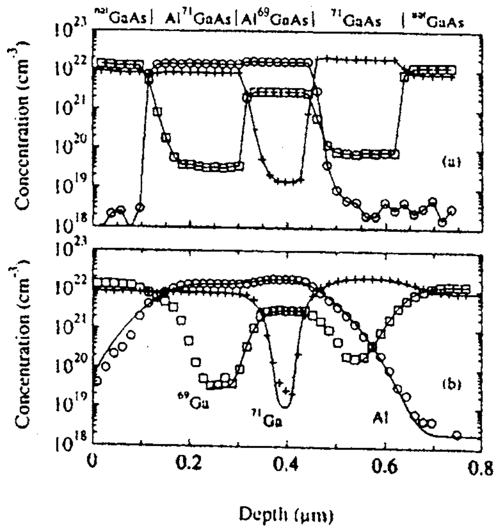


Figure 15 SIMS depth profiles of Al(○), ^{69}Ga (□), and ^{71}Ga (+) in the as-grown Al^{71}GaAs/Al^{69}GaAs/^{71}GaAs heterostructure b [see (a)] and after annealing at 1050°C for 1800s [see (b)]. Solid lines in (a) connect the data to guide the eye. Solid lines in (b) show best fits to the experimental profiles. For clarity, only every fourth data point is plotted in (a) (b) (after [115]).

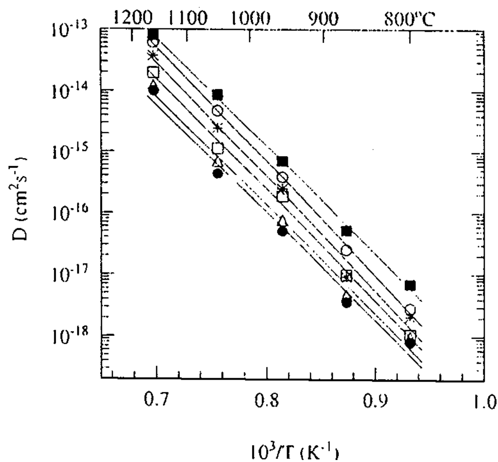


Figure 16 Temperature dependence of the diffusion coefficient D of Ga in Al$_x$Ga$_{1-x}$ As with $x = 0$ (○); 0.41 (*); 0.62 (□); 0.68 (△) and 1.0 (●) and of Al in GaAs (■) (after [115]).

Recently Wang *et al.* [114] determined an activation enthalpy of (4.24 ± 0.06) eV for Ga self-diffusion in GaAs which deviates from the present result of (3.71 ± 0.07) eV. The authors [114] favor the activation enthalpy of 3.71 eV for Ga self-diffusion in GaAs, because simultaneous annealing of the former ^{69}GaAs/^{71}GaAs heterostructure with sample a has revealed that the Ga profile near the interface of the ^{71}GaAs layer and the GaAs substrate, which was considered by Wang *et al.* for the self-diffusion study, deviates from the expected error function solution. This may be caused by surface contamination of the GaAs substrate wafer. The temperature dependence of Ga self-diffusion in AlGaAs reveals that Ga diffusion decreases with increasing Al content whereas H^{SD} appears to be constant within the experimental error (see Fig. 16 and Table IV).

The Al and Ga profiles near the Al^{69}GaAs/^{71}GaAs interface result from Al-Ga interdiffusion. The Al-Ga interdifussion coefficient $\tilde{D}$ can be expressed as

TABLE IV Activation enthalpy H^{SD} and natural logarithm of the preexponential factor D_0 for Al and Ga diffusion in Al$_x$Ga$_{1-x}$ As for intrinsic and As-rich ($p_{As} \sim 1$ atm) conditions (after [115])

Element	X of Al$_x$Ga$_{1-x}$As	H^{SD}	$\ln (D_0/\text{cm}^2\,\text{s}^{-1})$
Al	0.0	3.50 ± 0.08	-1.77 ± 0.80
Ga	0.0	3.71 ± 0.07	-0.45 ± 0.67
Ga	0.41	3.70 ± 0.21	-1.05 ± 1.98
Ga	0.62	3.60 ± 0.13	-2.51 ± 1.24
Ga	0.68	3.51 ± 0.19	-3.96 ± 1.78
Ga	1.0	3.48 ± 0.23	-4.51 ± 2.16

(see [107])

$$\tilde{D} = (X_{Al} D_{Ga} + X_{Ga} D_{Al}) \Phi S, \tag{44}$$

where X_{Al} and X_{Ga} are the mole fractions of Al and Ga. D_{Ga} and D_{Al} represent the Ga and Al diffusion coefficients in AlAs and GaAs, respectively. Φ is the thermodynamic favtor and S the vacancy wind factor which takes vacancy interaction and correlation effects into account. For modeling Al-Ga interdiffusion authors [115] assume the simplest possible values of $\Phi = 1$ (ideal solution) and of $S = 1$. On this basis, Fick's second law was solved numerically. The measured Al and ^{71}Ga profiles shown in Fig. 15b can both be described with the composition dependent interdiffusion coefficient $\tilde{D}$ according to Equation 44 which takes into account the actual mole fraction of Al and Ga as a function of depth. According Bracht *et al.* [115] , all interdiffusion profiles of samples a to d, which were annealed at the same temperature, are accurately described with data for D_{Ga} and D_{Al} which are consistent within 40%. The temperature dependence of Al diffusion in GaAs and of Ga diffusion in AlAs is shown in Fig. 16. Fitting Equation 37 to these results yields data for $H^{SD}$and D_0 which are listed in Table IV.

The values for H^{SD} shown in Table IV all lie in the range of (3.6 ± 0.1) eV. Recently Wee *et al.* [119] reported an activation enthalpy of (3.6 ± 0.2) eV and a pre-exponential factor of 0.2 cm^2 s^{-1} for interdiffusion of Al$_{0.2}$Ga$_{0.8}$As/GaAs at temperatures between 750 and 1150°C. Their data are consistent with data of paper [115] on Ga diffusion in AlGaAs and its dependence on the Al content. Tan *et al.* (see [112] and references therein) have proposed 6 eV for Ga self-diffusion. This result is based on a compilation of Ga self-diffusion and Al-Ga interdiffusion data obtained under various experimental conditions which includes AlGaAs/GaAs heterostructures with Al contents up to 100%. The activation enthalpy of 6 eV appears now to be questionable since Bracht *et al.* results unambiguously shows that Al-Ga interdiffusion does not represent Ga self-diffusion.

The single activation enthalpy found for Ga self- and Al-Ga interdiffusion suggests that the diffusion is mediated by the same native defect. Vacancies on the sublattice of the group-III atoms are assumed to mediate the Ga self- and Al-Ga interdiffusion under intrinsic conditions [120]. In this case the self-diffusion coefficient is given by (see also (37))

$$D^{SD} = f_V C_V^{eq} D_V = f_V g a^2 \nu \exp\left(-G_V^{SD}/kT\right), \tag{45}$$

where f_V is a correlation factor and C_V^{eq} and D_V the thermal equilibrium concentration of vacancies and their diffusivity. $C_V^{eq} D_V$ can be expressed by the geometry factor g, the jump distance a, the jump attempt frequency ν, and the Gibbs free energy G_V^{SD} of self-diffusion via vacancies. ν is proportional to $1/\sqrt{m}$, where m represents the atomic mass of the jumping atom. Different jump frequencies of ^{27}Al and ^{71}Ga caused by the difference in their masses are proposed to be the cause for the experimentally observed higher Al diffusion in GaAs compared to Ga self-diffusion. The experimentally determined ratio between the Al and the Ga diffusion coefficient in GaAs is consistent with $(m_{Ga}/m_{Al})^{0.5}$. The decreasing Ga diffusivity with increasing Al content in AlGaAs can be interpreted with a different location of the intrinsic Fermi energy level with respect to the vacancy charge transition state [121]. This causes different thermal equilibrium concentrations of vacancies in AlGaAs for different Al compositions.

In the next we follow Bracht *et al.* papers consider the experimental ratio between Ga diffusion in GaAs and AlAs. This equals the ratio between $C_V^{eq} D_V$ for GaAs and AlAs if the binding energy between Ga and a vacancy in AlAs is negligible. Assuming only a single negatively charged vacancy, the total concentration of vacancies in thermal equilibrium is given by [122]

$$C_V^{eq} = C_{V^0}^{eq}\left[1 + \exp\left(E_F^{in} - E_{V-/0}/kT\right)\right], \quad (46)$$

where $C_{V^0}^{eq}$ is the equilibrium concentration of a neutral vacancy, E_F^{in} the Fermi level under intrinsic condition, and $E_{V-/0}$ the acceptor energy level of a singly charged vacancy. The Fermi-level position is given by

$$E_F^{in} = 0.5E_g + 0.75kT\ln(m_V^*/m_C^*), \quad (47)$$

where E_g is the band gap energy and m_V^* and m_C^* are the effective density of state masses for holes and electrons in GaAs and AlAs, respectively [108]. Using Equations 46 and 47, the ratio between C_V^{eq} in GaAs and AlAs is estimated to be 6.5 $\times$ exp $\cdot$ [0.077 eV/(kT)] assuming that $C_{V^0}^{eq}$ in GaAs and AlAs are similar with respect to the valence-band position of AlAs as energy reference. Bracht *et al.* obtain a ratio of 12.8 at, e.g., 1050°C, which is consistent with their experimental result of 11.1.

Dopant enhanced as well as reduced layer disordering of semiconductor heterostructures are phenomena that have been reported frequently in the literature. Superlattice structure doped with Si during growth [123, 124] or by implantation [125, 126] reveal an enhanced Al-Ga interdiffusion after annealing compared to intrinsic conditions. Conversely, Be doping if Si-doped AlAs/GaAs has been found to suppress the superlattice disordering when the Be doping level exceeds that of Si [127, 128]. The dopant induced change in the position of the Fermi level, which strongly alters the concentrations of charged native point defects and hence the self- and interdiffusion, is generally considered to be possible for this diffusion phenomenon [129]. So far, the relative contributions of the various charged native point defects to the self- and interdiffusion, which are required to predict the disordering of GaAs based superlattice structures for different doping levels, are not accurately known. Total energy calculations provide one way to determine the nature of native defects in group III–V compound semiconductors. These calculations predict that the triply negatively charged vacancy on the gallium sublattice (V_{Ga}^{3-}) is the dominant native defect in GaAs, both for intrinsic and n-type doping conditions under an As-rich ambient [130]. On the other hand, the charge state of native defects can be deduced exp erimentally from the doping dependence of Ga self-diffusion in GaAs. Generally, Al-Ga interdiffusion in AlAs/GaAs superlattices is considered to simulate the self-diffusion process of the group-III atom. Tan and Gosele [131] analyzed data of Al-Ga interdiffusion in Si doped AlAs/GaAs superlattices [132] and found, in agreement with the theoretical calculations of paper [130], that the self-diffusion on the Ga sublattice is mediated by V_{Ga}^{3-}. In contrast to these seemingly consistent theoretical and experimental studies show that singly negatively charged Ga vacancies V_{Ga}^{-} mediate the Ga self-diffusion and the Al-Ga interdiffusion under n-type doping conditions [132, 134]. Some of the data of Mei *et al.* [132] , which have been used to support the triply charged defect hypothesis [131], can also be described assuming a singly charged vacancy [134]. Furthermore, Ga self-diffusion in GaAs has been proposed to be governed by triply charged defects at high Si doping levels, whereas at low doping concentrations a Fermi level independent mechanism was assumed [135, 136]. These discrepancies in the doping dependence of group-III atom diffusion in the AlGaAs material system forced Bracht *et al.* [137] to investigate with SIMS technique the Ga self-diffusion in undoped, Si- and Be-doped GaAs by using ^{71}GaAs/^{nat}GaAs isotope heterostructures.

Three Al^{71}GaAs/Al^{69}GaAs/^{71}GaAs/^{nat}GaAs isotope heterostructure with different Al content and one AlAs/^{71}GaAs/^{nat} GaAs structure were used for the diffusion experiments [137]. One set of samples was kept undoped and another set was doped either with Si or Be, i.e., altogether 12 different isotope structures were used. Each of these samples contains a ^{71}GaAs/^{nat}GaAs interface on which was focused on in the present self-diffusion study.

SIMS measurements on the undoped set of isotope heterostructures have revealed a change in the SIMS sputter rate with increasing Al concentration [138]. Since all samples used for the Ga self-diffusion experiments contain AlGaAs layers, a decrease in the sputter rate with increasing Al content results an apparently thicker AlGaAs layers and thinner GaAs layers. Therefore a point by point correction of the penetration depth of the SIMS profile was performed taking into account the measured sputter-rate dependence. According to cited authors the thickness of the AlGaAs and GaAs layers corrected in this way are consistent within 10% or better with the thickness determined directly with transmission electron microscopy.

Concentration profiles of ^{69}Ga near the ^{71}GaAs/^{nat}GaAs interface measured with SIMS after annealing

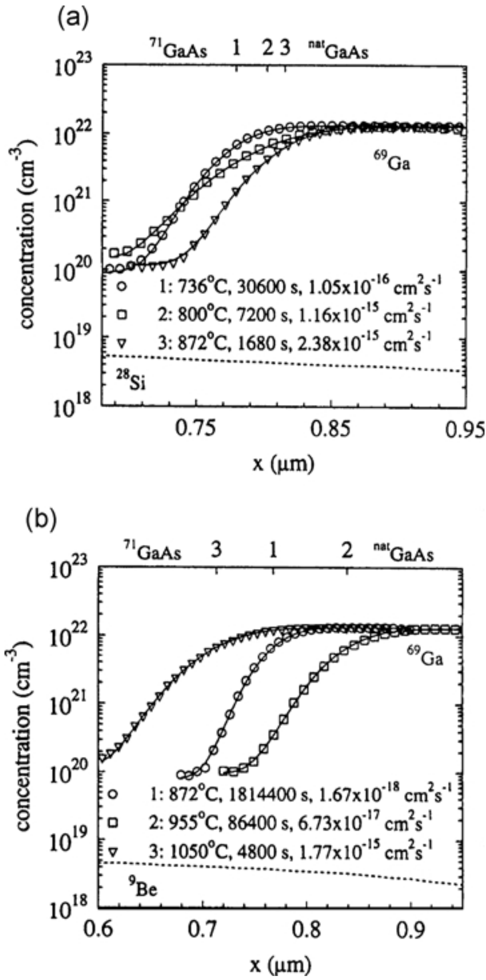


Figure 17 SIMS depth profiles of ^{69}Ga, ^{28}Si and ^{9}Be after annealing of (a) Si-doped and (b) Be doped ^{71}GaAs/^{nat}GaAs isotope structures at various temperatures and times as indicated. The solid lines show the best fits to experimental data. Ga self-diffusion coefficients deduced from each profile are listed in the figure. The numbers 1, 2, and 3 at the top X-axis indicate the position of the ^{71}GaAs/^{nat}GaAs interface of the particular sample. For clarity only the dopant profile (dashed line) from the Ga profile 3 is shown (after [137]).

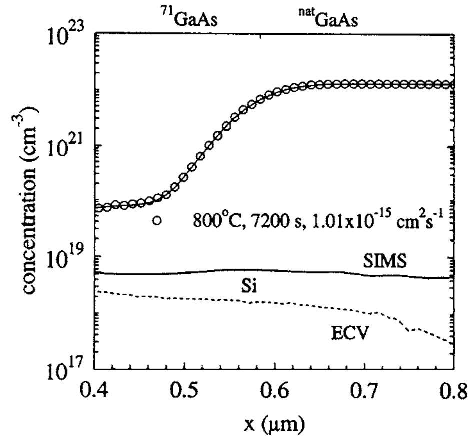


Figure 18 SIMS depth profiles of ^{69}Ga and ^{28}Si of a Si doped ^{71}GaAs/^{nat}GaAs structure annealed at 800°C for 2 h. The solid line shows the best fit to the experimental data which yields the Ga self-diffusion coefficient listed in the figure. The room temperature electron concentration, which was measured by C-V profiling, is shown as dashed line (after [137]).

of Si- and Be doped samples are shown in Figs 17a,b and 18. The Si- and Be profiles reveal a nearly constant dopant concentration across the ^{71}GaAs/^{nat}GaAs interface. A few sufficiently large Si-doped samples were also analyzed with C-V profiling to determine the free carrier concentration after annealing. The carrier concentration profile of one sample is shown in Fig. 18 together with the corresponding SIMS profiles of SI and Ga. C-V profiling revealed a Si donor concentration, C_{Si}, of about 2×10^{18} cm^{-3} along the ^{69}Ga profile. Similar results were obtained from the analysis of other samples. The electron concentration n after annealing is smaller than C_{Si} of the as-grown structures. This decrease is also apparent in the observed effect of doping on Ga self-diffusion. The hole concentration p of Be-doped samples could not be accurately determined by C-V profiling after annealing. In all cases the craters left behind by the electrochemical etching were very non-uniform. Consequently, no reliable data could be obtained for p. SIMS analysis shows that the Be concentration along the Ga self-diffusion profile equals the concentration of the as-grown structure. Therefore, it is assumed that the free carrier concentration p due to Be doping is not significantly affected by the thermal annealing. This is supportted by the doping dependence of Ga self-diffusion, which is accurately described taking into account the Be doping level of the as-grown samples (details see [137]).

Ge self-diffusion profiles shown in Figs 17 and 18 are accurately reproduced by the solution of Fick's law for self-diffusion across an interface (see above). All experimental ^{69}Ga profiles were described on the basis of Equation 41, in order to take into account the diffusion of ^{69}Ga from the adjacent Al^{69}GaAs layer into the ^{71}GaAs layer. The thickness $d = x_1 - x_2$ determined in this way for the ^{71}GaAs layer is consistent within 10% with the thickness in the as-grown structure. The uncertainty results mainly from the accuracy of the crater depth measurements.

The self-diffusion coefficients D_{Ga} extracted from the analysis of all Ga diffusion profiles in undoped, Si- and Be-doped isotope samples are listed in Table V. Each value of D_{Ga} is the average of at least four diffusion coefficients where each of them has been deduced from a different sample. The accuracy given for D_{Ga} represents the standard deviation of all data which belong to the same temperature and doping level. The temperature dependence of D_{Ga} for intrinsic, Si- and Be-doped GaAs is depicted in Fig. 64. Experimental results [138] for intrinsic conditions are accurately described by (see also [115])

$$D_{\mathrm{Ga}} = 0.64 \exp\left(-\frac{3.71\,\mathrm{eV}}{kT}\right) \mathrm{cm}^2\,\mathrm{s}^{-1}. \qquad (48)$$

Solid lines in Fig. 19 are best fits to the experimental data which accurately reproduce the temperature dependence of Ga sel-diffusion under the different doping conditions.

According to Bracht *et al.* [137], the Ga self-diffusion coefficient is given by the sum of the transport coefficients of vacancies in various charge states

$$D_{\mathrm{Ga}} = \frac{1}{C_0} \sum_{\mathrm{r}=0}^{3} f_{\mathrm{r}} C^{\mathrm{eq}}_{V^{\mathrm{r}-}_{\mathrm{Ga}}} D_{V^{\mathrm{r}-}_{\mathrm{Ga}}} \qquad (49)$$

TABLE V Ga self-diffusion coefficients D_{Ga} in undoped, Si- and Be-doped ^{71}GaAs/^{nat}GaAs isotope heterostructures. The numbers in parenthesis represent the number of samples which have been annealed at the particular temperature T and time t. A free electron and hole concentration of both $(3.0 \pm 0.5) \times 10^{18}$ cm^{-3} was determined by Hall effect measurements for the as-grown Si- and Be-doped samples. n_i, E_F^{in} and n/n_i represent the intrinsic carrier concentration [140], the position of the Fermi level under intrinsic conditions [140], and the ratio of the electron concentration in the doped and undoped samples, respectively (after [137])

$T°$, C	$t(s)$	D_{Ga}(cm^2 s^{-1})	Type of doping	n_i(cm^{-3})	E_F^{in}(eV)	n/n_i^a
1160	270 (4)	$(6.12 \pm 0.17) \times 10^{-14}$	Undoped	1.56×10^{18}	0.512	1.0
1050	1800 (4)	$(4.76 \pm 0.80) \times 10^{-15}$	Undoped	7.41×10^{17}	0.545	1.0
1050	3600(1); 4800(4)	$(1.46 \pm 0.20) \times 10^{-15}$	Be-doped	7.41×10^{17}	0.545	0.233
955	12600(1); 15120 (3)	$(3.84 \pm 0.61) \times 10^{-16}$	Undoped	3.54×10^{17}	0.572	1.0
955	10800 (1); 86400(4)	$(8.28 \pm 2.50) \times 10^{-17}$	Be-doped	3.54×10^{17}	0.572	0.116
872	260400(4); 432000(2)	$(2.47 \pm 0.17) \times 10^{-17}$	Undoped	1.70×10^{17}	0.595	1.0
872	496800(1); 1814400(4)	$(2.37 \pm 0.59) \times 10^{-18}$	Be-doped	1.70×10^{17}	0.595	0.056
872	1680(5)	$(3.67 \pm 0.93) \times 10^{-15}$	Si-doped	1.70×10^{17}	0.595	14.8
800	1209600(2); 1370700(2)	$(2.84 \pm 0.36) \times 10^{-18}$	Undoped	8.32×10^{16}	0.615	1.0
800	7200(4); 10800(1)	$(9.86 \pm 1.65) \times 10^{-16}$	Si-doped	8.32×10^{16}	0.615	26.6
736	–	1.88×10^{-19}(b)	Undoped	4.09×10^{16}	0.632	1.0
736	25200(1); 30600(4)	$(1.45 \pm 0.21) \times 10^{-16}$	Si-doped	4.09×10^{16}	0.632	45.9

[a]Data correspond to fit #3 (see Table VI).
[b]Calculated with followed equation $D_{Ga} = 0.64 \exp(-\frac{2.71 eV}{k_B T})$ cm^2 s^{-1}.

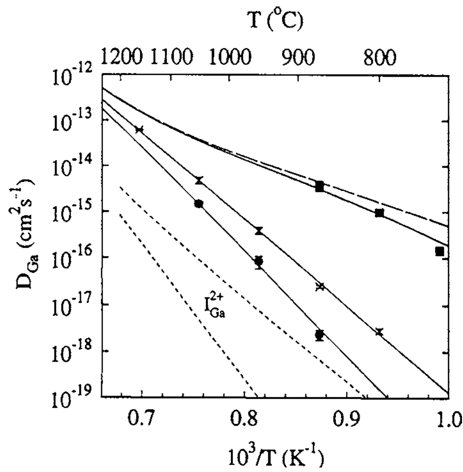


Figure 19 Temperature dependence of the Ga self-diffusion coefficient D_{Ga} in undoped (x), Si-doped (■), and Be-doped (●) GaAs for $P_{As_4} = 1$ atm. The influence of doping on D_{Ga} is best reproduced (see solid lines) if the effect of the Fermi level together with a compensation of Si donors by negatively charged vacancies is taken into account. The long-dashed line is expected if the compensation by vacancies is ignored. Short-dashed lines represent the contribution of the doubly positive charged Ga self-interstitial I_{Ga}^{2+} to Ga self-diffusion. Lower dashed line: I_{Ga}^{2+} contribution for intrinsic conditions and $P_{As_4} = 1$ atm; upper dashed line: I_{Ga}^{2+} contribution calcula ted for a hole concentration of 3×10^{18} cm^{-3} (after [137]).

where C_0, f$_r$, $C_{V_{Ga}^{r-}}^{eq}$ represent the Ga atom density in GaAs ($C_0 = 2.215 \times 10^{22}$ cm^{-3}), the diffusion correlation factor, and the thermal equilibrium concentration and the diffusion coefficient of the vacancy $D_{V_{Ga}^{r-}}$ with the charge $r \in [0, 1, 2, 3]$, respectively. The correlation factor contains information about the microscopic jump mechanism [29]. The different vacancy configurations can introduce energy levels within the energy band-gap of GaAs. Occupation of these energy states depends on the position of the Fermi level. Under extrinsic conditions, i.e., when the hole or the electron concentration introduced by doping exceeds the intrinsic carrier concentration, the Fermi level deviates from its intrinsic position. As a consequence, the ratio of the charged to neutral vacancy concentrations is changed [122]. For vacancies which can introduce single (V_{Ga}^{-}), double (V_{Ga}^{2-}), and triple (V_{Ga}^{3-}) acceptor states with energy levels at $E_{V_{Ga}^{-}}$, $E_{V_{Ga}^{2-}}$, $E_{V_{Ga}^{3-}}$ above the valence band edge E_V, the rations are given by

$$\frac{C_{V_{Ga}^{-}}^{eq}}{C_{V_{Ga}^{0}}^{eq}} = g_{V_{Ga}^{-}} \exp\left(\frac{E_F - E_{V_{Ga}^{-}}}{kT}\right) \tag{50}$$

$$\frac{C_{V_{Ga}^{2-}}^{eq}}{C_{V_{Ga}^{0}}^{eq}} = g_{V_{Ga}^{2-}} \exp\left(\frac{E_F - E_{V_{Ga}^{2-}} - E_{V_{Ga}^{-}}}{kT}\right) \tag{51}$$

$$\frac{C_{V_{Ga}^{3-}}^{eq}}{C_{V_{Ga}^{0}}^{eq}} = g_{V_{Ga}^{3-}} \exp\left(\frac{E_F - E_{V_{Ga}^{3-}} - E_{V_{Ga}^{2-}} - E_{V_{Ga}^{-}}}{kT}\right) \tag{52}$$

The thermal equilibrium concentration of the neutral Ga vacancy $C_{V_{Ga}^{0}}^{eq}$ is independent of the position of the Fermi level E_F. However, if E_F lies above $E_{V_{Ga}^{r-}}$, the formation of negatively charged vacancies is energetically favored compared to the formation of neutral defects. In Equations 49–51 the degeneracy factors $g_{V_{Ga}^{r-}}$, which take into account a spin degeneracy of the defect and a degeneracy of the GaAs valence band, have all been set to one.

The total concentration of V_{Ga} in thermal equilibrium, $C_{V_{Ga}^{eq}}$, is given by the sum of the corresponding concentrations of the various charge states of gallium vacancies

$$C_{V_{Ga}}^{eq} = \sum_{r=0}^{3} C_{V_{Ga}^{r-}}^{eq} = C_{V_{Ga}^{0}}^{eq} \sum_{r=0}^{3} \frac{C_{V_{Ga}^{r-}}^{eq}}{C_{V_{Ga}^{0}}^{eq}}. \tag{53}$$

Based on Equations 49–51 it becomes clear, that $C_{V_{Ga}}^{eq}$ changes when the position of the Fermi level changes. Naturally, the Ga self-diffusion coefficient given by Equation 49 may be written as

$$D_{Ga} = \frac{1}{C_0} C_{V_{Ga}^{0}}^{eq} \sum_{r=0}^{3} \frac{C_{V_{Ga}^{r-}}^{eq}}{C_{V_{Ga}^{0}}^{eq}} D_{V_{Ga}^{r-}}. \tag{54}$$

Recent published molecular dynamic calculations [139] show that the migration enthalpy of a Ga is nearly independent on its charge state. For accuracy migration vis second-nearest neighbor hopping, an enthalpy of 1.7 eV was determined for V_{Ga}^{0} compared to 1.9 eV for V_{Ga}^{3-} [139]. In addition, in paper [138] was assumed that the entropy change associated with the V_{Ga}^{r-} migration and the correlation factor f_r are similar for all charge states, i.e., $D_{V_{Ga}^{r-}} = D_{V_{Ga}}$ and $f_r \approx f$. With these assumptions, it follows that [138]

$$D_{Ga} = \frac{1}{C_0} f C_{V_{Ga}^{0}}^{eq} D_{V_{Ga}} \sum_{r=0}^{3} \frac{C_{V_{Ga}^{r-}}^{eq}}{C_{V_{Ga}^{0}}^{eq}}. \tag{55}$$

Taking into account Equations 50–52, and the well-known expression

$$\frac{n}{n_i} = \exp\left(\frac{E_F - E_F^i}{kT}\right), \tag{56}$$

the following relationship between the Ga self-diffusion coefficient under intrinsic and extrinsic conditions have next form

$$\frac{D_{Ga}(n)}{D_{Ga}(n_i)} = \frac{1 + \sum_{m=1}^{3} \left(\frac{n}{n_i}\right)^m \exp\frac{mE_F^i - \sum_{m=1}^{3} E_{V_{Ga}^{m-}}}{kT}}{1 + \sum_{m=1}^{3} \exp\frac{mE_F^i - \sum_{m=1}^{3} E_{V_{Ga}^{m-}}}{kT}}. \tag{57}$$

In Equations 56 and 57, n_i. E_F^i, n, and E_F define the free electron concentration and the Fermi level position under intrinsic and extrinsic conditions, respectively. If only one charge state r dominates Ga self-diffusion, Equation 57 is reduced to

$$\frac{D_{Ga}(n)}{D_{Ga}(n_i)} = \left(\frac{n}{n_i}\right)^r \tag{58}$$

with $r \in [0, 1, 2, 3]$. This simplified relationship has been generally used to analyze the doping dependence of Ga self-diffusion and Al-Ga interdiffusion [131]. Bracht *et al.* [137] prefer more general Equation 57, since this equation takes into account that the charge state of the vacancy mediating Ga self-diffusion may change with doping and temperature.

Fitting of Equation 57 to experimental results required data for E_F^i and n_i. For these quantities the cited authors used the data reported by Blakemore [140] which are listed in Table V. The electron concentration n of the Si- and Be-doped GaAs samples were calculated via the charge balance equation yielding [137]

$$n = \frac{1}{2}\left(C_{Si} - \sum_{m=0}^{3} mC_{V_{Ga}^{m-}}^{eq}\right) + \sqrt{n_i^2 + \frac{1}{4}\left(C_{Si} - \sum_{m=0}^{3} mC_{V_{Ga}^{m-}}^{eq}\right)} \tag{59}$$

and

$$\frac{n_i^2}{n} = p = \frac{1}{2}\left(C_{Be} - \sum_{m=0}^{3} mC_{V_{Ga}^{m-}}^{eq}\right) + \sqrt{n_i^2 + \frac{1}{4}\left(C_{Be} - \sum_{m=0}^{3} mC_{V_{Ga}^{m-}}^{eq}\right)}, \tag{60}$$

respectively. C_{Si}—represents the Si donor concentration and C_{Be}—the acceptor concentration due to Be doping. Both concentrations are equal the free carrier concentration of 3×10^{18} cm^{-3}, measured at room temperature. Negatively charged vacancies affect the free carrier concentration are taken into account in both Equations 59 and 60.

The free carrier concentration given by Equations 58 and 60 also depends via Equations 49–51 and 56 on the vacancy-related energy levels. Therefore Equations 57, 59 and 60 were solved by Bracht *et al.* simultaneously. $C_{V_{Ga}^{m-}}^{eq}$ ($r \in [0, 1, 2, 3]$) were calculated with Equations 44–46 using a thermal equilibrium concentration of neutral vacancies which is given by [141, 142].

$$C_{V_{Ga}^{0}}^{eq} = C^* P_{As_4}^{1/4} T^{-5/8} \exp\left(\frac{H_{V_{Ga}^{0}}^{f}}{kT}\right), \tag{61}$$

where C^* is a preexponential factor and $H_{V_{Ga}^{0}}^{f}$ the formation enthalpy of the neutral vacancy. Equation 55 includes the influence of the As_4 vapor phase on $C_{V_{Ga}^{0}}^{eq}$ [141, 142]. So far equilibrium concentrations of V_{Ga}^{0} have been estimated with an uncertainty factor of at least 10 [141]. In order to accurately describe the doping dependence of Ga self-diffusion on the basis of Equation 51, not only $E_{V_{Ga}^{m-}}$ ($m \in [0, 1, 2, 3]$) but also C^* were used as fit parameters. The formation enthalpy of neutral vacancies was set to a value of (1.9 ± 0.2) eV. This value equals the difference between the activation enthalpy of (3.71 ± 0.07) eV for Ga self-diffusion and the vacancy migration enthalpy of (1.8 ± 0.2) eV which was deduced in paper [136] from Al-Ga interdiffusion in nonstoichiometric AlAs/GaAs quantum wells (see also Equation 48). Authors [137] concluded that $H_{V_{Ga}^{0}}^{f} = (1.9 \pm 0.2)$ eV is believed to be fairly reliable.

The temperature and doping dependences of Ga self-diffusion in intrinsic, n-type and p-type GaAs which were calculated in paper [137] on the basis of the parameters given under fit #3 in Table VI, are shown by the solid lines in Fig. 19. The long-dashed line in Fig. 19 displays the temperature dependence of D_{Ga} which is expected if compensation via negatively charged vacancies is ignored in the n-type GaAs with $C_{Si} = 3 \times 10^{18}$ cm^{-3}.

A representation of D_{Ga} data [137] versus the ratio n/n_i is given in Fig. 20. Solid lines have been calculated with Equation 57 taking into account the n/n_i values listed in Table V and the results for $E_{V_{Ga}^{m-}}$ of fit #3 (see Table VI). The doping dependence of D_{Ga} at 872°C, which was calculated on the basis of the theoretical results for $E_{V_{Ga}^{m-}}$ ($m \in [0, 1, 2, 3]$) reported by Baraff

TABLE VI Parameter values obtained from fitting Equation 51 to the experimental results. The Ga self-diffusion is best described by neutral, singly and doubly negatively charged vacancies with relative contributions of the various charge states which change with temperature and doping. Additionally, a compensation of the Si donor concentration by negatively charged Ga vacancies was taken into account; it becomes especially significant at low temperatures (after [137])

Fit parameter[a]	Fit #1[b]	Fit #2[c]	Fit #3	Fit #4	Fit #5
$E_{V_{Ga}^{-}} - E_V$ (eV)	0.38 ± 0.11	0.42 ± 0.04	0.42	0.42	0.42
$E_{V_{Ga}^{2-}} - E_V$ (eV)	0.72 ± 0.19	0.60 ± 0.04	0.60	0.60	0.60
$E_{V_{Ga}^{3-}} - E_V$ (eV)	1.33 ± 4.33	–	–	–	–
$H^f_{V_{Ga}^0}$ (eV)	1.9	1.9	1.9	1.7	2.1
$C^*(\mathrm{cm}^{-3})$	0.293	0.293	162 ± 54	18.6 ± 5.8	1385 ± 487

[a] E_V denotes the valence band edge.
[b] C^* and $H^f_{V_{Ga}^0}$ according to Tan [141].
[c] V_{Ga}^{3-} contribution ignored, C^* and $H^f_{V_{Ga}^0}$ according to Tan [141].

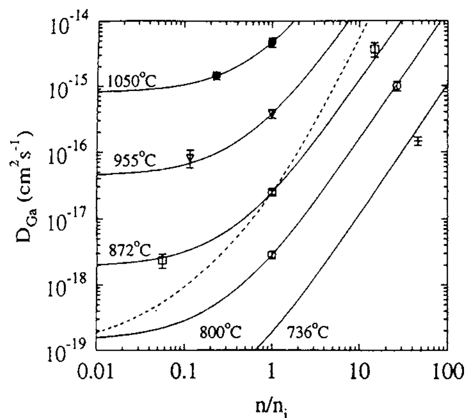


Figure 20 Ga self-diffusion coefficients D_{Ga} versus the ration n/n_i between the free carrier concentration under extrinsic and intrinsic doping conditions $P_{As_4} = 1$ atm. and different temperatures as indicated. Solid lines were calcuated via Equation 57 taking into account $E_{V_{Ga}^m}$ given by fit #3 of Table XII. The dashed line shows the doping dependence of D_{Ga} for 872°C which was calculated for $E_{V_{Ga}^-} - E_V \approx 0.20$ eV, $E_{V_{Ga}^{2-}} - E_V \approx 0.52$ eV, and $E_{V_{Ga}^{3-}} - E_V \approx 0.72$ eV according to Baraff and Schlüter [130] (after [137]).

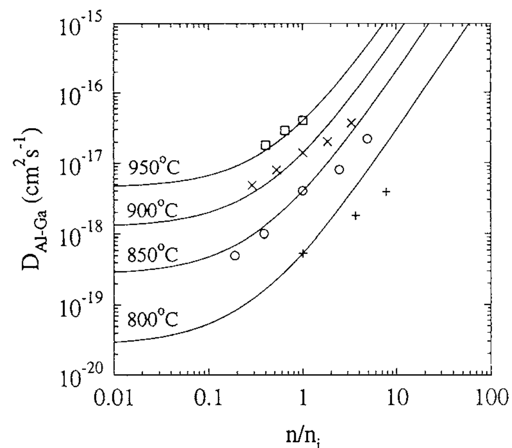


Figure 21 Al-Ga interdiffusion coefficient D_{Al-Ga} reported by Muraki and Horikoshi [135] versus n/n_i. Solid lines were calvulated with Equation 51 taking into account $D_{Ga}(n_i)$ of Muraki and Horikoshi and results of Bracht *et al.* [137] for $E_{V_{Ga}^-}$ (after [137]).

and Schluter [138], is shown in Fig. 20 as a dashed line. The strong deviation from data of paper [137] clearly indicates that in contrast to the theoretical calculations the triply charged Ga vacancy does not mediate the Ga self-diffusion under intrinsic and n-type doping conditions.

As was noted above, Muraki and Horikoshi [135] have studied, with the help of photoluminescence spectroscopy, the Al-Ga interdiffusion of Si- and Be-doped $Al_{0.34}Ga_{0.66}As/GaAs$ superlattice structures. They have proposed that Al-Ga interdiffusion is mediated by singly negatively charged Ga vacancies in both n- and p-type material. The Al-Ga interdiffusion in these structures with an Al content of 34 at.% is expected to simulate the Ga self-diffusion in GaAs. This is supported by recent results obtained in paper [115] on the Al composition dependence of Ga self-diffusion in AlGaAs (see also above) which show that for Al concentrations of 41 at.% D_{Ga} in AlGaAs is only a factor of two smaller than D_{Ga} in GaAs. The interdiffusion coefficients given by Muraki and Horikoshi for different diffusion temperatures and doping levels are illustrated in Fig. 21 as a function of the ratio n/n_i. A compensation of the Si donors by charged vacancies was also considered by taking into account the equilibrium concentrations of vacancies which, consistent with results of paper [137], were obtained for the particular Si doping levels. Solid lines shown in Fig. 21 were computed via Equation 51 with D_{Ga} $(n/n_i = 1)$ of Muraki and Horikoshi and the results of Bracht *et al.* [137] for $E_{V_{Ga}^{m-}}$ (see also fit #3, Table VI). No adjustable parameter were used. The dependence on As_4 pressure cancels out because all diffusion experiments were performed under identical conditions. The deviation between the experimental data and the corresponding solid line for 800°C may indicate that the Si donor concentration is additionally reduced by the formation of Si_{Ga}—acceptor pairs such as Si_{Ga}–Si_{As} complexes or other compensation centers. This is supported by the fact that after annealing Muraki and Horikoshi have observed broad photoluminescence signals in the Si-doped samples which they attributed to deep centers.

Self-diffusion coefficients for 900°C reported by Muraki and Horikoshi are displayed in Fig. 22 as a function of n/n_i. The solid line in Fig. 22 was calculated on the basis of the results of Bracht *et al.* and accurately describes the Ga self-diffusion data for $n/n_i > 1$. Small differences $(n/n_i > 1)$ between the as-grown and annealed structures can also be caused by

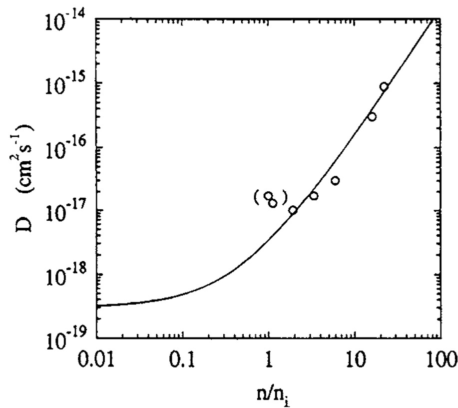


Figure 22 Ga self-diffusion coefficient D_{Ga} reported by Muraki and Horikoshi [135] versus n/n_i. The data in parenthesis are considered to be not reliable. The solid line was calculated via Equation 51 taking into account results of Bracht *et al.* for $E_{V_{Ga}^-}$ and a value $D_{Ga}(n_i)$ which is a factor of 5 smaller than the corresponding data given by Muraki and Horikoshi (after [137]).

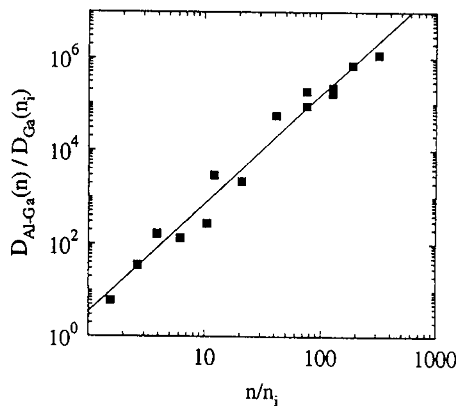


Figure 23 Al-Ga interdiffusion coefficients $D_{Al\text{-}Ga}$ (n) of Mei *et al.* [133] normalized by $D_{Ga}(n_i)$ [115] as a function of n/n_i. The solid line represents the best fit which is reproduced by $D_{Al\text{-}Ga}$ $(n)/D_{Ga}(n_i) = 3.4$ $(n/n_i)^{2.3}$ (after [137]).

SIMS broadening effects associated with a degradation of the surface quality during annealing under As-poor conditions.

Fig. 23 illustrates the ratio between the reduced interdiffusion coefficients $D_{Al\text{-}Ga}(n,\ P_{As_4} = 1$ atm) and $D_{Ga}(n,\ P_{As_4} = 1$ atm) as a function of n/n_i. In this double logarithmic representation, the slope of the experimental data equals the exponent r. The best fit [137] yields $r = 2.3 \pm 0.1$ showing that V_{Ga}^{2-} rather than V_{Ga}^{3-} mediate the Al-Ga interdiffusion in agreement with the results of Bracht *et al.* Based on this reanalysis, the V_{Ga}^{3-} mediated self- and interdiffusion in n-type GaAs, and the activation enthalpy of 6 eV for Ga self-diffusion under intrinsic conditions proposed by Tan and Gosele [131], are found to be incorrect (details see [137, 143]).

In summary, the doping dependence of group-III atom diffusion in the AlGaAs material system ([135, 136, 141, 142] and references therein) can be consistently explained with the result presented by Bracht *et al.* [137]. Neutral, singly and doubly charged Ga vacancies all contribute to the self-diffusion in undoped, p-type and n-type material with relative contributions which depend on temperature and doping. The lower power dependence for the doping effects of self-diffusion in samples containing group-VI donors is proposed to be caused by the formation of next nearest neighbor complexes between the dopant and the vacancy.

Chapter 2. Neutron transmutation doping

2.1. The NTD process—a new reactor technology

The neutron transmutation doping (NTD) process involves the cooperation of semiconductor materials specialists, device producers, radiation damage and defect specialists and reactor personnel. Of all possible interactions among these groups, those with the reactor community have traditionally been the weakest. Reactor personnel have, therefore, had the greatest learning curves to overcome. It is to the credit of both the reactor community and the semiconductor industry that these difficulties have been overcome so readily in the few years since 1975 when NTD silicon first appeared on the market. The transmutation doping process simply involves irradiation of an undoped semiconductor with a thermal neutron flux. The major advantages of the NTD process are illustrated schematically in Fig. 24. The homogeneity in NTD-Si is a result of a homogeneous distribution of silicon isotopes in the target material and the long range of neutrons in silicon. Doping accuracy is a result of careful neutron flux integration. The material improvements offered by the NTD process form the basis for semiconductor device improvement (details see [145]).

As is well-known, research reactor facilities provide the best source of thermal neutrons for this purpose at the present time (see e.g. [146–149]). These reactors are ideally suited for such projects because they have usually been constructed with sample irradiation as one of the prime design requirements. Although these reactor

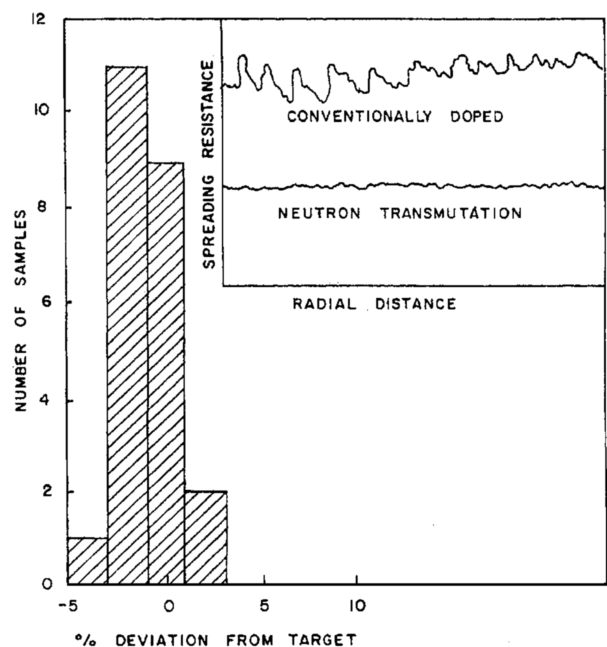


Figure 24 Advantages of NTD process. Histogram of irradiation *t* arg *et* accuracy obtained for commercial sample lot at NURR. Insert is a schematic representation of spreading resistance traces across a wafer diameter for conventionally doped and NTD Si (after [144]).

facilities provide a source of thermal ($E \sim 0.025$ eV) neutrons, this thermal flux is always accompanied by a fast neutron component which is not useful in providing doping transmutations, but does produce radiation damage (displacements of atoms from their normal lattice sites) which must be repaired by annealing, the process of heating the irradiated material to temperatures sufficiently high that the irradiation produced defects become mobile and can removed.

To understand the process further, we must be concerned with the interactions of neutrons, both thermal and fast, with the target material to be doped. Because neutrons are neutral particles, their range of penetration in most materials is usually very long. They interact only very weakly with atomic electrons through their magnetic movements. Being neutral, neutrons see no Coulombic barrier at the target nuclei and, therefore even very slow neutrons may reach into the nucleus without difficulty. In fact, the slower the neutron velocity, the greater is the time of interaction between the neutron and the target nucleus. We, therefore, expect the probability of neutron capture by the target nuclei to be enhanced at low neutron energies.

This interaction is described in terms of a capture cross-section, σ_c, where the number of captures per unit volume, N, is given by

$$N = N_T \sigma_c \Phi, \tag{62}$$

where N_T is the number of target nuclei per unit volume, σ_c the capture cross-section and $\Phi = \phi t$ is the influence (flux times time) given in n/cm^2. Fig. 25 shows the capture cross-section as a function of neutron energy for

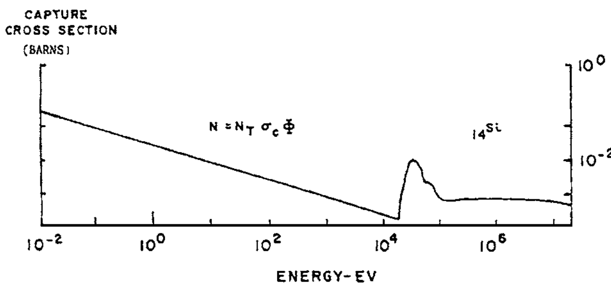


Figure 25 Neutron capture cross-section as a function of neutron energy for natural silicon (after [144]).

silicon as averaged over all three stable silicon isotopes [150]. Similar behavior is found individually for each silicon isotope. It can be seen in Fig. 25 that for low energies:

$$\sigma_c \sim E^{-1/2} \sim 1/V. \tag{63}$$

For a given nuclear radius, $(1/V)$ is proportional to the interaction time. Therefore, the cross-section represents a probability of interaction between the nucleus and neutron.

After neutron capture, the target nucleus differs from the initial nucleus by the addition of the nucleon and is a new isotope in an excited state which must relax by the emission of energy in some form. This emission is usually in the form of electromagnetic radiation (photons) of high energy usually called gammas (see e.g. [151] and references therein). The time for decay of this excess energy by gamma emission can be very short (prompt gammas) or can take an appreciable time in which case a half-life a factor of two, can be measured. The gamma emission spectrum is characteristic of the nuclear energy levels of the transmuted target nuclei and can be used as a powerful trace substance technique called neutron activation analysis (NAA), to detect quantitatively impurity levels as low as 10^9 atoms/cm^3 [152]. A typical trace substance NAA gamma spectrum is shown in Fig. 26. Each emission line is characteristic of a particular nuclear transition of a particular isotope. The absorption of a neutron and the emission of gammas is represented by the notation:

$$A_{X(n,\gamma)} A + 1_X, \tag{64}$$

where (n, γ) represents (absorption, emission), A is the initial number of nucleons in the target element X before neutron absorption while $A + 1$ is the number after absorption. It is possible for the product isotope $A + 1_X$ to be naturally occurring and stable. In many cases, however, the product isotope is unstable. Unstable isotopes further decay by various modes involving the emission of electrons (β-decay), protons, α-particles, K-shell electron capture or internal conversion until a stable isotopic state is reached (details see, e.g. [153]). These decays produce radioactivity and can be characterized by their half-lives $T_{1/2}$. In the case of

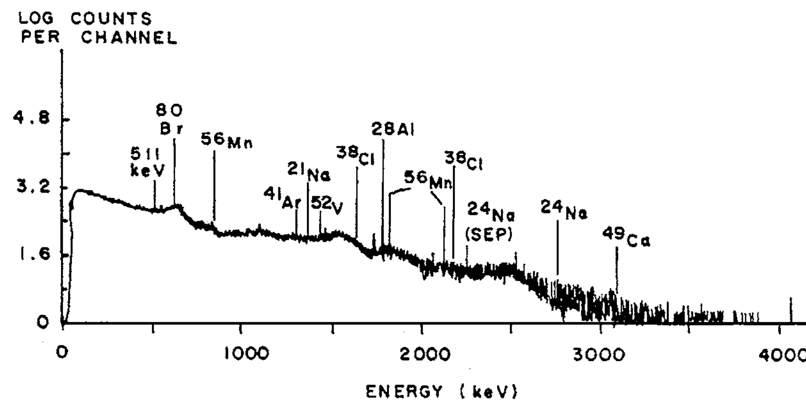
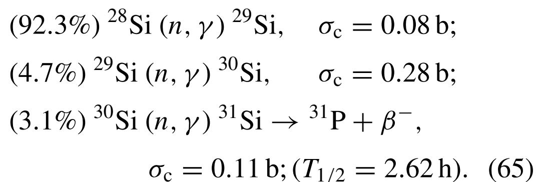


Figure 26 Typical neutron activation analysis (NAA) gamma-ray spectrum to search for trace substances deposited on an air filter after 1 min. irradiation at MURR (after [152]).

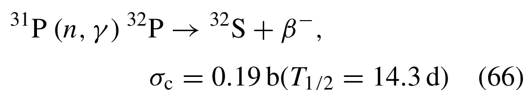
silicon, three stable target isotopes are transformed by (n, γ) reactions [154] as follows:



$$(92.3\%)\ {}^{28}\mathrm{Si}\ (n, \gamma)\ {}^{29}\mathrm{Si}, \quad \sigma_c = 0.08\,\mathrm{b};$$
$$(4.7\%)\ {}^{29}\mathrm{Si}\ (n, \gamma)\ {}^{30}\mathrm{Si}, \quad \sigma_c = 0.28\,\mathrm{b};$$
$$(3.1\%)\ {}^{30}\mathrm{Si}\ (n, \gamma)\ {}^{31}\mathrm{Si} \rightarrow {}^{31}\mathrm{P} + \beta^-,$$
$$\sigma_c = 0.11\,\mathrm{b}; (T_{1/2} = 2.62\,\mathrm{h}). \quad (65)$$

The relarive abundance of each stable silicon isotope is shown in parenthesis (see also [151]). The cross-sections are expressed in barns (1 barn = 10^{-24} cm^2). The first two reactions produce no dopants and only redistribute the relative abundances slightly. The third reaction produces ^{31}P, the desired donor dopant [155], at a rate of about 3.355 ppb per 10^{18} n_{th} cm^{-2} [144]. This production is calculated using (63), the ^{30}Si ($N_T \cong 5 \times 10^{22}$ Si $\cdot$ cm^{-3} $\times$ 0.031).

In addition to the desired phosphorus production reaction and its relatively short half-life for β^- decay, the reaction



$${}^{31}\mathrm{P}\ (n, \gamma)\ {}^{32}\mathrm{P} \rightarrow {}^{32}\mathrm{S} + \beta^-,$$
$$\sigma_c = 0.19\,\mathrm{b} (T_{1/2} = 14.3\,\mathrm{d}) \quad (66)$$

occurs as a secondary undesirable effect. The decay of ^{32}P is the primary source of radioactivity in NTD float fine Si. Of course, any undesirable trace impurities in the silicon starting material can lead to abnormally long half-life activities which may require that material be held out of production until exempt limits are reached. These factors have stimulated on the subject of radiation protection. Once the dopant phosphorus has been added silicon ingot by transmutation of the ^{30}Si isotope, the problems remains to make this radiation damaged and highly disordered material useful from an electronic device point of view. Several radiation damage mechanisms contribute to the displacement of the silicon atoms from their normal lattice position (details see below). These are:

1. Fast neutron knock-on displacements,
2. Fission gamma induced damage,
3. Gamma recoil damage,
4. Beta recoil damage,
5. Charged particle knock-on from (n, p); (n, α) etc reactions (see details [151]).

Estimates can be made of the rate at which Si atom displacements are produced by these various mechanisms, once a detailed neutron energy spectrum of the irradiation is known, and these rates compared to the rate at which phosphorus is produced.

The number of displaced atom per unit volume per second, N_D, is estimated from the equation

$$\mathrm{d}N_D/\mathrm{d}t = N_T \sigma \phi \nu, \quad (67)$$

where N_T is the number of target atoms per unit volume, ϕ is the flux of damaging particles and ν is the number of displacements per incident damaging particle. The cross-section for gamma induced displacements in silicon is small while the cross-sections for (n, p); (n, α), etc., are of the order of millibarns and have thresholds in the MeV range. The fast neutron knock-on displacements can be calculated from the elastic neutron scattering cross-section once the reactor neutron energy spectrum is known. Estimates of fission spectra and graphite moderated fission spectra can be found in the literature [156].

Even if the fast neutron damage could be completely eliminated, the recoil damage mechanisms, which are caused by thermal neutron capture, still would produce massive numbers of displacements compared to the number of phosphorus atoms produced. In the case of gamma recoil, a gamma of energy $\hbar\omega$ carries a momentum $\hbar\omega/c$ which must equal the Si isotope recoil momentum MV. The recoil energy

$$E_R = \frac{1}{2} M V^2 = \frac{1}{2} \frac{(\hbar\omega)^2}{MC^2} \quad (68)$$

is, therefore, departed to the silicon atom of mass M for each gamma emitted. An average over all possible silicon isotope gamma emission and cross-sections yields an average recoil energy of 780 eV [157] which is significantly higher than the Si displacement energy. A similar effect is encountered for ^{31}Si β^- decay. The β^- carries a momentum

$$p = \frac{1}{c}\sqrt{E_\beta^2 - (m_0 c^2)^2} \equiv MV. \quad (69)$$

Therefore,

$$E_R = \frac{1}{2} M V^2 = \frac{1}{2}\left[E_\beta^2 - (m_0 c^2)^2\right] / (Mc^2). \quad (70)$$

For a β^- emitted with an energy of 1.5 MeV, E_R = 33.2 eV or roughly twice the displacement threshold.

From the above considerations, a very crude estimate of the numbers of displacements per phosphorus atom produced can be made. The results of the estimation are shown in Table VII. While the absolute numbers of displacements should not be taken literally, the relative magnitudes of the amounts of damage produced by these various mechanisms are probably order of magnitude correct. An inspection of Table VII indicates that the gamma recoil mechanism is significant relative to the quantity of phosphorous produced even in highly moderated reactors. We are led to the inescapable conclusion that transmutation doping will always produce significant amounts of radiation damage which must be

TABLE VII Number of displaced silicon atoms per phosphorus produced for various damage mechanisms shown for an in-core fission spectrum and a graphite moderated spectrum (after [144])

Damage particle/position	In core	In pool
Fast neutron	4.06×10^6	1.38×10^4
Fission gamma	3.64×10^3	36.4
Gamma recoil	1.29×10^3	1.29×10^3
Beta recoil	2.76	2.76
Total DISP/(P)	4.06×10^6	1.51×10^4

repaired in some way. These defects introduce defect levels into the band gap which cause free carrier removal and a reduction in carrier mobility and minority carrier lifetime (see, e.g. [157]).

The defects produced by neutron irradiation are removed by thermal annealing as discussed previously. It is at this point in the process where disagreement as to the best procedure is likely to be the greatest. The spectrum of possible defect structures and their energetics is impressively large and incompletely understood. Therefore, annealing procedures according Meese [144] will be based on art rather than exact science. They will also tend to become proprietary for this reason.

This is unfortunate since it is precisely in this area that fundamental knowledge is needed to produce the best possible product. Although carrier concentration and mobility recovery are easily obtainable by various annealing procedures, minority carrier life-time recovery is very elusive at present.

So, neutron transmutation offers both advantages and disadvantages over conventionally doped silicon (details see [151]) .

Advantages

1. Precision target doping (= 1% or better).
2. Better axial and radial uniformity.
3. No microresistivity structure.

Disadvantages

1. Irradiation costs.
2. Reduction in minority carrier life-time.
3. Radioactive safeguards considerations.

The steady growth of the NTD-silicon (and others NTD-semiconductors) market suggests (see also below) that the advantages are outweighing the disadvantages.

2.2. Reactor facilities for transmutation doping

Irradiation of silicon for the purpose of phosphorus doping has been carried out in the Harwell (England) research reactors since 1975 [158]. At Harwell silicon is irradiated in the twin, heavy-water, materials-testing reactors DIDO and PLUTO. Both reactors operate continuously throughout the year and each actieves an availability of greater than 86% of calender time. As the reactors are D_2O moderated and cooled the irradiation conditions are particularly good for silicon doping. The ratio of thermal to fast neutrons is in excess of 1000:1, which minimizes the damage which has to be removed by annealing. In considering the accuracy which can be achieved in the neutron doping process it is necessary to consider the neutron flux profiles and gradients. Fig. 27 shows a typical unperturbed flux profile for an irradiation position over a length of 50 cm spaced about the maximum flux value. It will be noted that the maximum and minimum flux values differ by 8–12% of the maximum, and the gradient at the lower end is particularly steep. To reduce this variation and to smooth the profile, flux flatteners or neutron screens, in the form of stainless steel tubes are fitted to the facility liners with the result shown as "modified profile" in Fig. 27. As was shown by Smith, the severe gradient arising from the 8–12% variation has been reduced to 2–3% and the overall variation reduced to 5%.

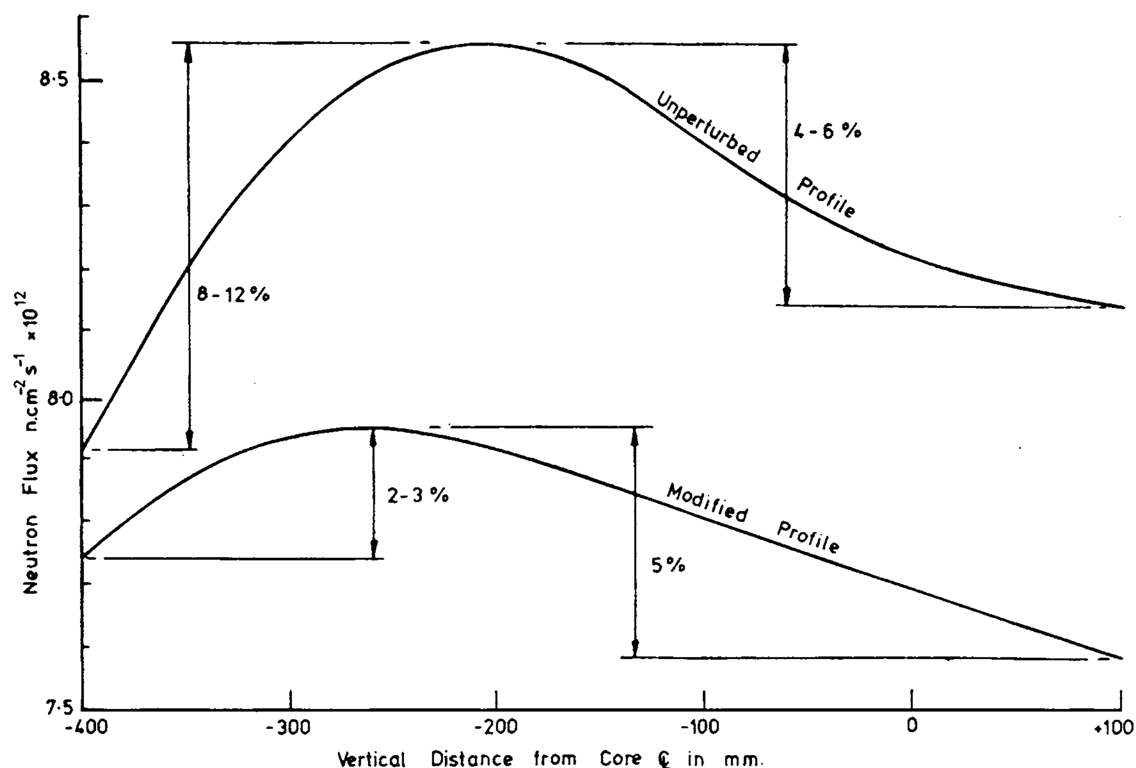


Figure 27 Typical neutron flux profiles (after [158]).

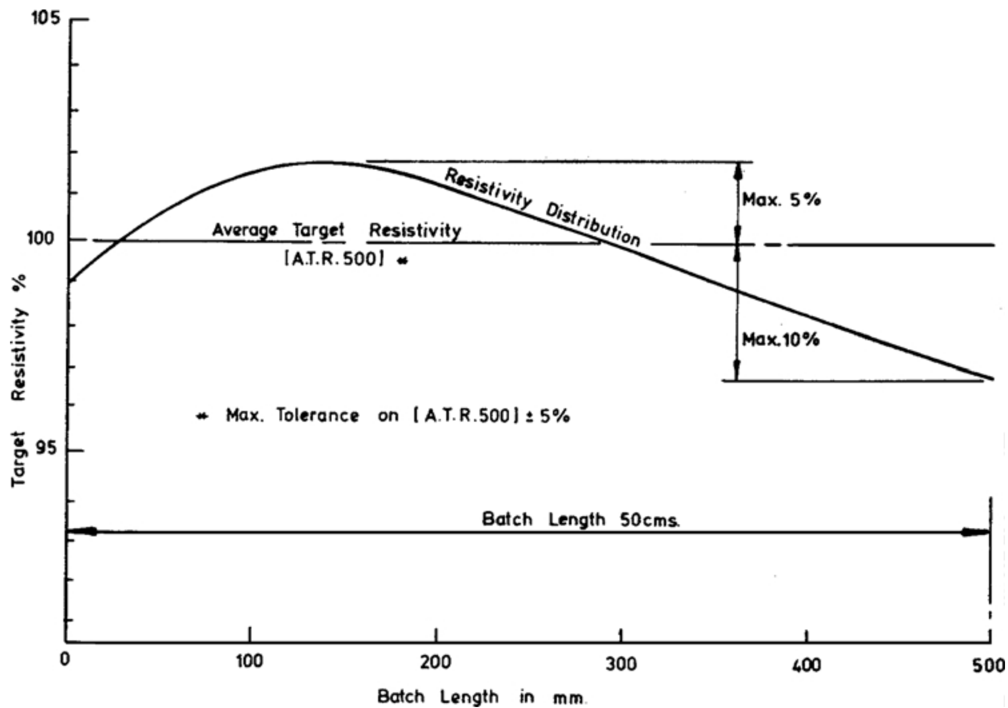


Figure 28 Resistivity variations (after [158]).

The "modified profile" can now be examined in more detail, and this is shown in Fig. 28. The profile can be considered in terms of "resistivity distribution" which is of interest to the customer and can form the basis for the technical specification of the product. The "Average Target Resistivity" or A.T.R. 50 [158] is, as the name suggests, the average of the resistivity over a length of 50 cm. Achievement of this value is subject to variations arising from the irradiation timing, the measurement of the mean flux, and the distribution shape; therefore indicating authors apply a tolerance of ±5% to the A.T.R.50 value. The exact shape of the distribution is also subject to variations due to disturbances in the reactor such as control-absorber movements and other irradiations and experiments; according to [158] tolerances are therefore also applied to the "Resistivity Distribution" of 5% maximum greater, and 10% maximum less, than A.T.R.50.

Researchers of Harwell offer irradiation of volume. Fig. 29 illustrates this and shows the volume which for convenience Smith describes as a "batch". It is cylindrical volume of 90 mm diameter and 500 mm in length. Although radial gradients are small, crystals are rotated during irradiation, and a maximum variation of ±1% on a diameter of 10 cm (4 in.) is guaranteed. In practice, according to Smith, variations are less than can be measured within the accuracy of a conventional four-point probe. The average annual dose is typically 6.5×10^{17}n cm^{-2}, which corresponds to a resistivity of 35 Ω-cm. In many of the literature references on the neutron doping of silicon one finds the statement that "three days, or at most a week, after irradiation, silicon is save to transport and to handle". This is, of course, a relative statement and it is necessary to define what is meant by "safe". In the I.A.E.A. publication "Regulations for the Safe Transport of Radioactive Materials, 1977" it states that to qualify as "Exempt" or safe material the

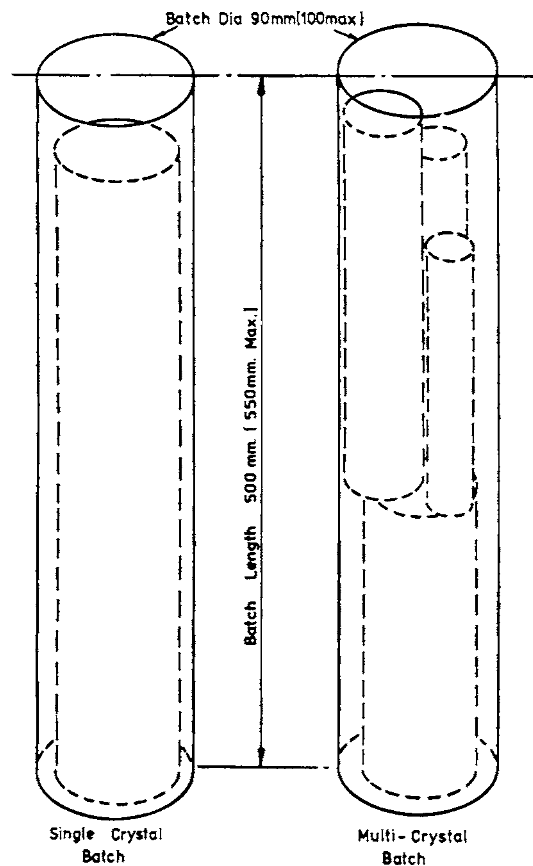


Figure 29 Batch dimensions (after [158]).

following conditions must be met:

1. The radiation level at any point on the external surface of the package shall not exceed 0. mRem/h; and

2. The non-fixed radioactive contamination of any external surface shall not exceed $10^{-4}\mu$Ci cm^{-2}. This

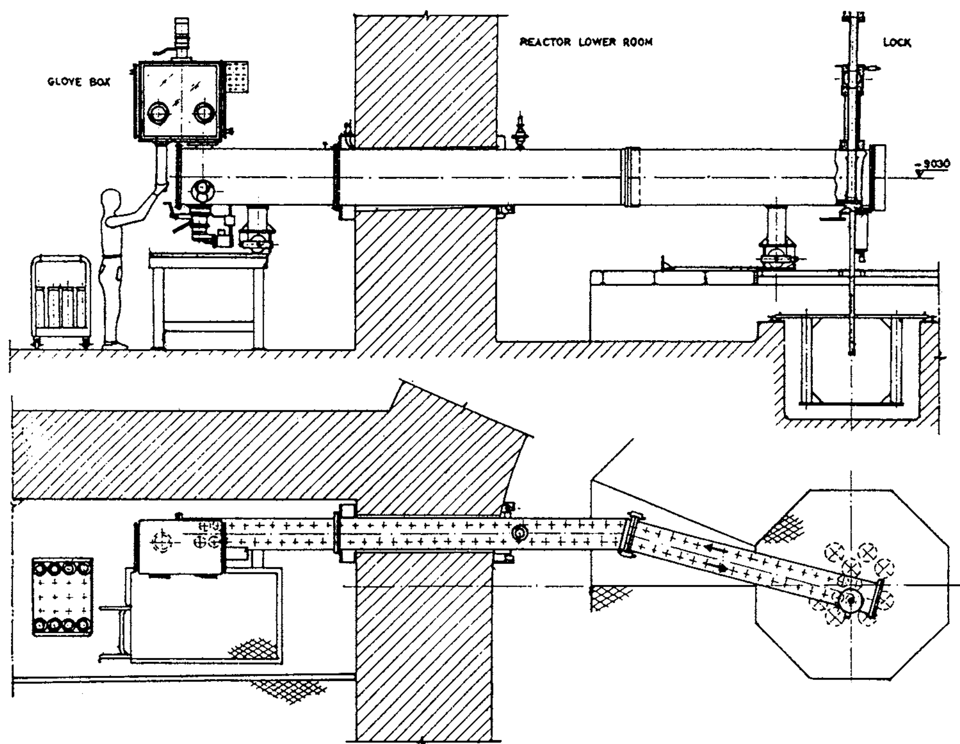


Figure 30 Section and layout of the irradiation of the irradiation facility (after [160]).

level is permissible when averaged over any area of 300 cm^2 of any part of the surface.

These are internationally accepted standards for the transport of packages and Harwell complies them. To meet this obligation Smith write that each crystal, or piece, of silicon complies with the following criteria before certification for dispatch:

1. Radiation is less than 0.1 mRem/h, a factor of 5 less than that necessary for the package according to I.A.E.A. Regulations. and
2. Contamination is less than $5 \times 10^{-5} \mu$Ci cm^{-2}, a factor of 2 less than that necessary for the package.

As a result of the application of these low levels which we consider to be essential only three days delay prior to shipment os not always possible [404], particularly for material irradiated down to low resistivities (see also [159]).

The tendency of the NTD silicon producers to increase their capacity and the extending range of devices in which NTD silicon is being used, call for special reactor irradiation facilities. The JRC heavy water moderated ESSOR reactor (Ispra, Italy) is described by Bourdon and Restelli [160]. In this article, attention has been especialy devoted to obtain an automated operation of the facility, and to optimize the characteristics of the irradiation volume. The height of 50 cm, with respect to a core vertical dimension of 150 cm, has been selected in order to obtain a minimum axial spread of the neutron flux ($\pm$4%). A thermal neutron flux of $(2.7 \div 3) \times 10^{14}$ n cm^{-2} s^{-1} is available at the irradiation position with a thermal to fast ($>$100 keV) neutron flux ratio larger than 400.

The silicon crystals are loaded into a transport unit which can locate up to 100 ingots 77 mm in diameter by 500 mm length (Fig. 30). The ingots are loaded protected by a bored plastic capsule which defines the irradiation volume indicated above. The capsules are then loaded into the reactor through a lock between the transport unit and the channel for transfer from the air to the heavy water circuit. The heavy water circulation assures efficient cooling of the silicon ingot during irradiation (the maximum crystal temperature should not exceed 70°C) and induces, by use of a suitable shape of the capsule (see Fig. 31), a slow rotation of the ingot for minimizing radial dispersion of the neutron fluency. At the end of the irradiation, determined by the control system, the capsule is transferred into the transport unit where it stays for at least 4 days before being disharged thus providing for decay time between the irradiation and the extraction time. The plastic material for the capsule must be chosen as a function of its qualities of mechanical behavior, radiation resistance, heat resistance and low radioactivation. On the basis of the three first requirements, some commercial plastics have been selected and samples of each irradiated in the HFR reactor of Petten (Netherland) or in ESSOR, in order to choose the best material (see also Fig. 32). None of the tested materials is completely satisfactory so that Bourdon and Restelli have foreseen the use of Noryl 731 for the preparation of capsules to be used for short irradiation times, and polysterene or PPO for long irradiation times.

The high thermal to fast neutron flux ratio assures fewer lattice defects as demonstrated by the fact that a thermal annealing of 5 min. at 750–800°C has been found sufficient to achieve complete recovery of the final resistivity (see Fig. 33).

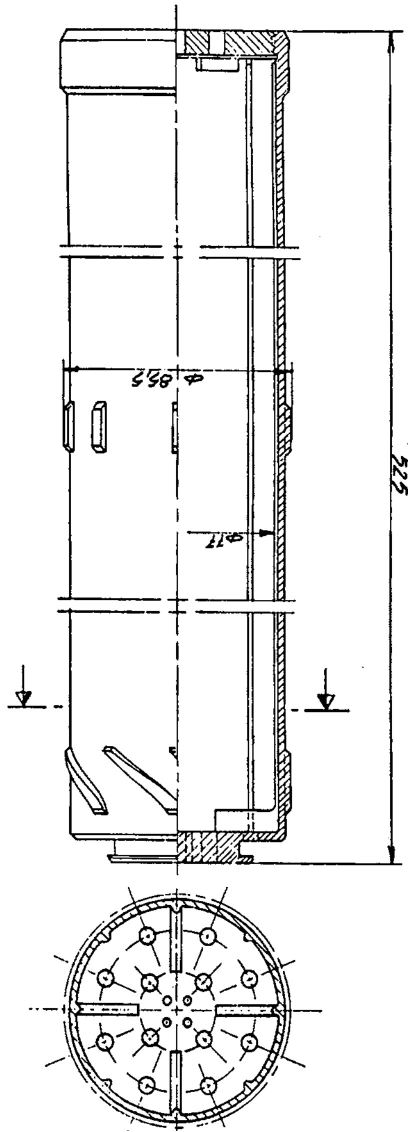


Figure 31 Irradiation capsule (after [160]).

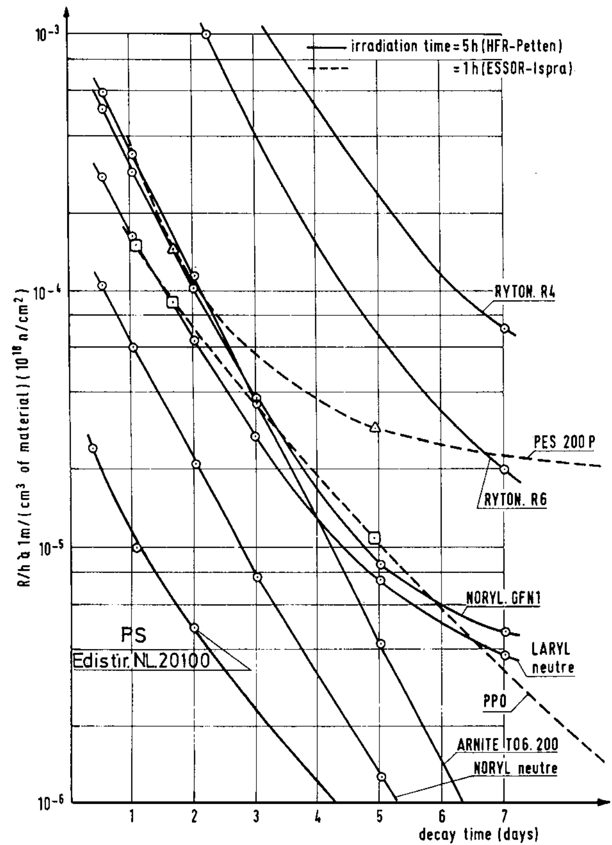


Figure 32 Radioactivation of some plastic materials (after [160]).

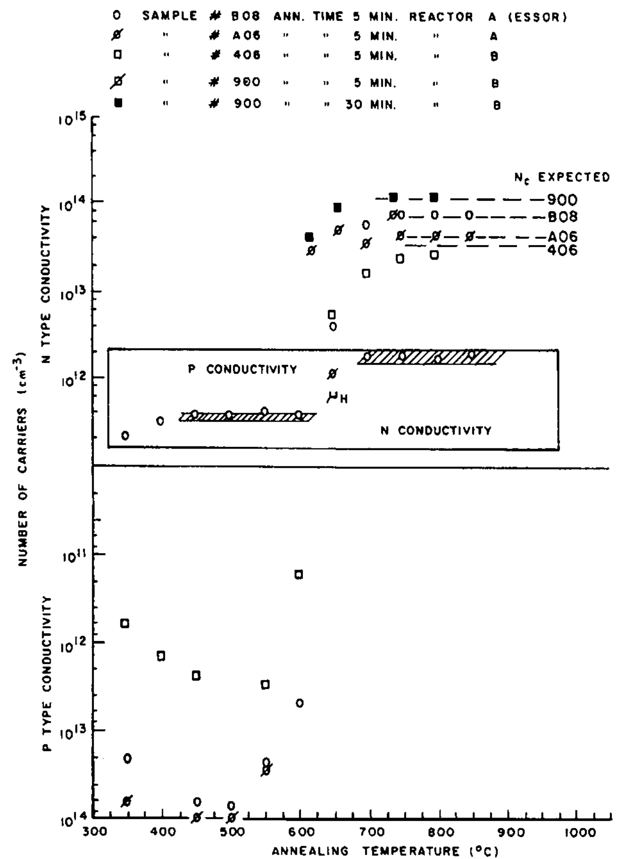


Figure 33 Number of carriers (and Hall mobility) versus annealing temperature (isochronal anneals) (after [160]).

Concluding this paragraph we should remember once more the advantages offered by the ESSOR reactor. These advantages are a high thermal neutron flux density, very uniform and easily controllable over a large irradiation volume, a good thermal to fast neutron flux ratio in conjunction with the possibility to irradiate the capsule immersed in D_2O which assures an efficient cooling of the Si crystal.

The General Electric Test Reactor (GETR) (Schenectady, U.S.A.) was designed and constructed to provide large irradiation volumes outside the reactor pressure vessel in *a* surrounding water pool. The thermal neutron flux available for silicon irradiation apans four decades, 10^{11} to $>10^{14}$ nv (see Fig. 34). The large irradiation volume permits the inclusion of flux flattening and spectral softening devices if desired [161]. As is well-known, NTD silicon offers significant technical advantages over chemically doped silicon (see also [162]). In particular, NTD silicon has a more uniform phosphorus concentration across the radius of an ingot or wafer than chemically doped material. The uniformity could approach 1% for a

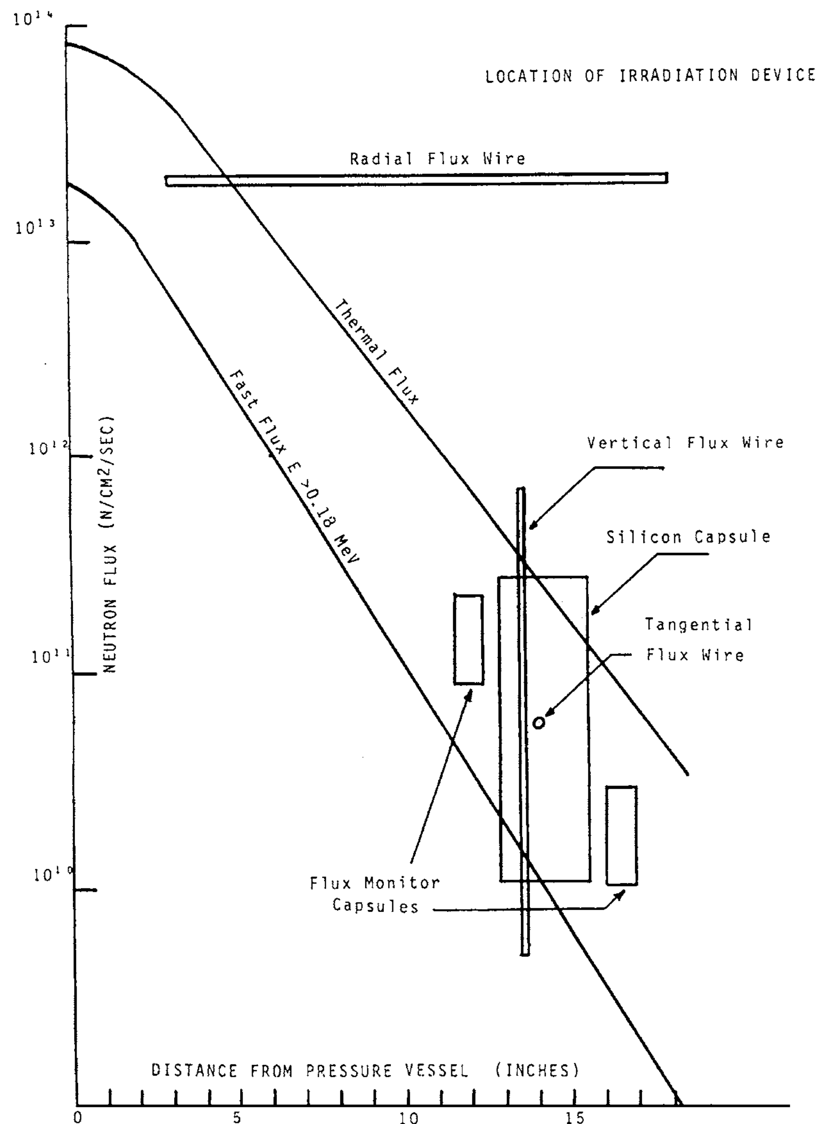
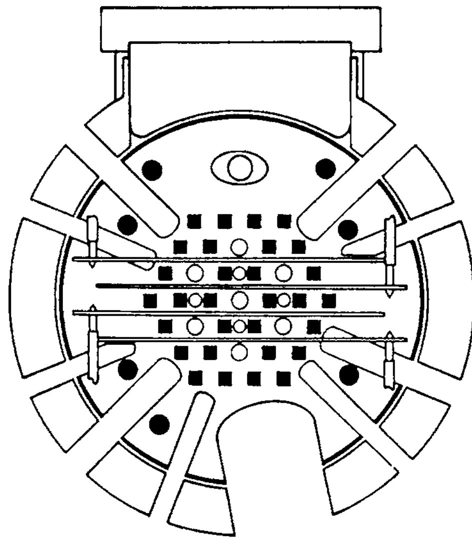


Figure 34 GETR 50 MW neutron flux pool position Z-6 (after [161]).

three inch diameter wafer activated in the 50 MW GETR. Availability of such uniform NTD silicon would make it possible to manufacture higher power density thyristors for high voltage applications (details see below).

A program of silicon irradiation is being carried out at the National Bureu of Standards (NBS) (Washington, DC, U.S.A.) 10 MW, heavy water moderated reactor. A plan view of the NDBS reactor core showing several of the irradiation locations is shown in Fig. 35. A set of five pneumatic rabbit tubes, useful for irradiating silicon chips to analyze for impurities or to study irradiation damage, provide a range of thermal fluxes from 2×10^{11} n cm^{-2} s^{-1} (copper-cadmium ratio of 3400) to 6×10^{13} (copper-cadmium ratio of 46) [164]. Those researchers interested in long term silicon doping irradiations can currently use two vertical facilities designed G2 and G4 [163]. Both facilities are D_2O filled and are completely isolated from the reactor coolant. Since they are isolated, encapsulation of the silicon is unnecessary and only an aluminium hardness is needed to hold the sample. The G2 tube will accept samples up to 1.6 inches in diameter and has a neutron flux at the core midplane of 1.1×10^{14} (copper-cadmium ratio of 55) [163]. A vertical flux profile of this facility is depicted in Fig. 36. Irradiation of samples in G2 for periods of one day to six weeks has been done for Oak Ridge National Laboratory (ORNL). The predicted phosphorus doping rate of 7.5×10^{13} atoms cm^{-3} hr^{-1} yielded a concentration in excellent agreement with that measured by ORNL. The G4 tube is located at the center of the reactor core and will accept samples up to 3 inches in diameter. Its neutron flux has roughly the same shape as that in G2 but is about 28% greater. A one kilogram silicon sample has been irradiated in G4 to a measured phosphorus concentration of 1.4×10^{17} atoms cm^{-3} (details see [163]).

The absolute differential neutron-energy spectrum for the low temperature fast-neutron irradiation facility in the CP-5 reactor by means of a 20-foil activation technique was determined by Kirk and Greenwood [165]. Fig. 37 shows a simplified schematic of VT53, the cryogenic fast-neutron irradiation facility at CP-5 (a more detailed description of this equipment see in paper [165] and references cited therein). The elements silicon, nickel, niobium, and gold were selected to illustrate



■ Fuel Element
● 4 inch Experimental Thimble
O 3½ inch Experimental Thimble
o 2½ inch Experimental Thimble

Figure 35 Plan view of NBSW reactor core and irradiation facilities (after [163]).

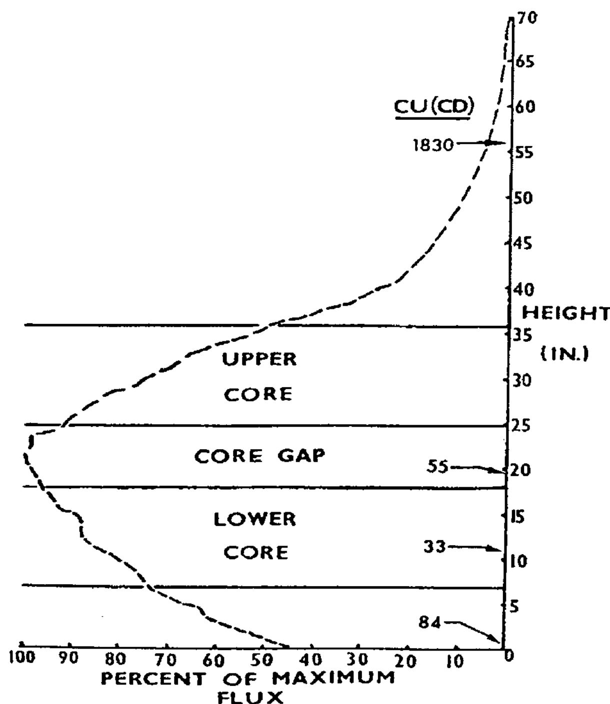


Figure 36 Vertical flux profile of position G2 (after [163]).

the results of these calculations. Table VIII gives the group distributions (partially collapsed from the original, more detailed group structure) of primary recoils and the total primary recoil cross-sections^{+} for these four elements. Also displayed in this table (for gold only) is an error of one standard deviation for recoils in each group, based on the covariant error matrix generated during the neutron spectrum error analysis by the SANDANL code. To facilitate comparison among the four elements, Fig. 38 shows the integral distribution of primary recoils (see also [166]). Using the Robinson

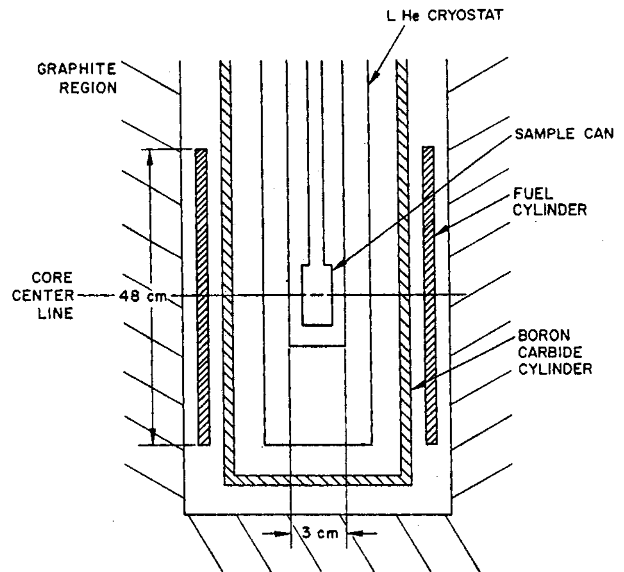


Figure 37 Simplified schematic of the cryogenic fast-neutron facility in CP-5 (VT53) (after [165]).

analytical approximation [167] to the Linfhard *et al.* [168] theory of electronic energy losses, it is also possible to calculate the damage energy distribution and spectrum averaged total damage energy cross-sections [169, 170]. The number of Frenkel defects (interstitial and vacancy pairs) (see also [157]) produced by a primary recoil of energy T is generally proportional to the damage energy available from this recoil, which is just the total recoil energy, T, minus the electronic energy losses at this recoil energy. The distribution of damage energy over the recoil energy groups thus gives a good indication of how the Frenkel defects are distributed with primary recoil events. As an example, Table IX gives the distributions of damage energy in recoil energy groups and the spectrum averaged damage energy cross-sections for the same four elements shown in Table IX. Fig. 39 graphically illustrates the corresponding integral damage energy distributions.

The integrated neutron flux determined by Kirk and Greenwood (2.2×10^{16} n/m^{2}s $\pm 13\%$, for $E_n > 0.1$ MeV) for the low temperature fast neutron facility in CP-5 is 70% greater than that determined less accurately early [171]. The amount of ^{235}U burn up in the fuel cylinder over this period of time is not known with certainty. However, based on a comparison of resistivity damage rate measurements in cooper made over a comparable period of time [172, 173], the burn-up is about 5% for an 8 year time period. Thus, it is reasonable to assume that the flux (corrected for burnup) in this facility has remained constant in time within the uncertainty of the present measurement. It is indeed found that the major cause of the difference between the flux measurements is the improvement in accuracy of the cross-section data on which these flux determinations are based. Therefore, any use of data from previous experiments in this facility will employ the presently determined neutron spectrum and integrated flux values. One caution that is perhaps obvious should be noted. In comparing experimental data from different neutron irradiation facilities, one must be careful when using

TABLE VIII Group distributions of recoils and spectrum averaged cross-sections in gold, niobium, nickel and silicon irradiated in the VT 53 fast neutron energy spectrum (after [165])

Primary recoil/ energy group	Si PRD (%)	Ni PRD (%)	Nb PRD (%)	Au PRD (%)	+/−
0.5 eV ↓	2.4	4.0	2.6	11.8	1.4
5–10	0.4	1.7	1.2	5.1	0.9
10–20	0.6	2.9	1.9	5.6	1.0
20–50	1.5	6.2	4.3	6.7	0.6
50–100	1.9	6.8	3.9	5.4	0.3
100–200	2.4	8.7	4.9	7.2	0.6
200–400	2.7	10.1	6.1	9.4	0.9
400–600	1.8	8.0	4.7	6.4	0.7
600–800	1.5	6.8	4.1	4.7	0.6
800–1000	1.3	4.4	3.5	3.7	0.6
1–1.5 keV ↓	2.8	5.4	7.2	6.6	1.4
1.5–2	2.4	3.9	5.5	4.2	1.1
2–3	3.9	5.3	8.4	5.4	1.5
3–5	6.7	5.8	11.5	5.8	1.3
5–7	5.5	3.6	7.2	3.0	0.5
7–10	6.9	3.9	6.7	2.6	0.4
10–20	18.8	5.7	9.0	3.9	0.4
20–40	18.1	3.3	4.7	2.1	0.1
40–60	6.7	1.3	1.5	0.3	0.02
60–80	3.5	0.8	0.6	0.08	0.004

PRD = Primary recoil distributions.
Spectrum averaged cross-sections (barns)

σ_{elastic}	2.86	8.82	6.77	9.35	±12%
$\sigma_{\text{inelastic}}$	0.05	0.12	0.40	0.80	±9%
$\sigma_{\text{n,2n}}$	0.0001	0.001	0.001	0.002	21%
$\sigma_{\text{n,p}}$	0.002	0.02	0.0001	–	–
$\sigma_{\text{n},\alpha}$	0.001	0.001			
σ_{total}	2.91	8.96	7.17	10.15	±11%

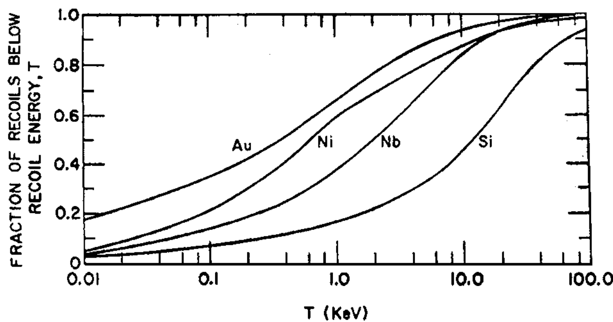


Figure 38 Integral distributions of primary recoils in Au, Nb, Ni and Si irradiated in VT53 (after [165]).

integrated flux values that have been determined at different times, since cross-sections have changed with time (details see [165, 174]).

2.3. Nuclear reaction under influence of the charged particles

According to the modern concept (see, e.g. [153]), the nuclear reactions leading to the formation doped impurities, may be proceed under the influence of charged particles (protons, deutrons, α-particles, etc) and neutrons and γ-quantes. In such a case the energy of bombardering on the nuclear particle must be sufficiently enough for overcoming mutual Coulomb repulsion of particle and nuclear. And in case of following particle flying out from excited compound nuclear the other charged particle should also obtain in compound

TABLE IX Group distributions os damage energy and spectrum averaged damage energy cross-sections ($\langle\sigma T_d\rangle$) in gold, niobium, nickel, and silicon irradiated in the VT 53 fast neutron energy spectrum (after [165])

Primary recoil/ energy group	Si DED (%)	Ni DED (%)	Nb DED (%)	Au DED (%)
0–5 eV ↓	0.00	0.001	0.001	0.01
5–10	0.0002	0.003	0.002	0.02
10–20	0.0005	0.01	0.006	0.04
20–50	0.003	0.05	0.03	0.11
50–100	0.008	0.1	0.05	0.2
100–200	0.02	0.3	0.14	0.5
200–400	0.04	0.6	0.3	1.2
400–600	0.05	0.8	0.4	1.4
600–800	0.06	1.0	0.5	1.4
800–1000	0.06	0.8	0.6	1.5
1–1.5 keV ↓	0.2	1.4	1.6	3.6
1.5–2	0.2	1.4	1.7	3.2
2–3	0.5	2.6	3.6	5.7
3–5	1.3	4.3	7.6	9.5
5–7	1.6	4.1	7.1	7.4
7–10	2.8	6.2	9.2	8.9
10–20	12.7	14.7	20.7	22.6
20–40	12.6	10.7	10.9	6.2
60–80 8.9	9.5	6.3	2.0	

DED = Damage energy distribution (%)
Spectrum Averaged damage energy cross-sections (keV-barns)
$\langle\sigma T_d\rangle$ 42.8 36.0 33.9 19.5(±8.7%)

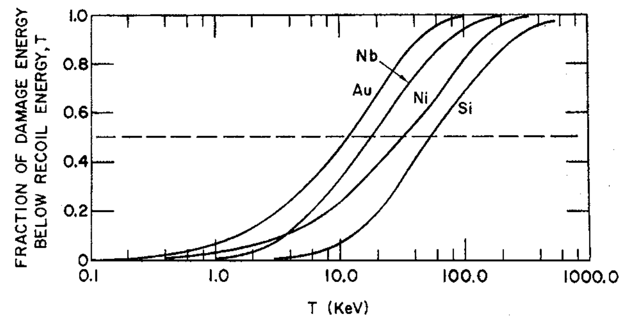


Figure 39 Integral damage energy distributions for Au, Nb, Ni and Si irradiated in VT53 (after [165]).

nuclear the energy sufficient for overcoming this Coulomb barrier. If r_n is a nuclear radii and e is the electron charge, in such a case the barrier height [151] has the following relation

$$B_b = \frac{Z_x Z_a e^2}{r_n} \simeq \frac{Z_x Z_a}{A^{1/3}}\ \text{MeV} \tag{71}$$

and achieves of 5–10 MeV for light nuclears, 10–20 MeV for middle nuclears and 20–30 MeV for heavy nuclears [151, 153].

Historically the first sourced of charged particles (α-particles) were the radioactive elements of Ra, Rn, Po, Pu and others, which in result of radioactive decay emit in 1 s on 1 gr emitter until 10^{10}–10^{11} α-particles [175] with energies of 4–8 MeV [153, 175]. The main question is how the beam of α-particles passes on its own energy to the atoms of irradiated material. It is appeared that owing to Coulomb interaction of the particles with material the kinetic energy of charged particles is spent on the ionization and excitation the atoms of irradiated

substance. The estimations show [153] that the cross-section of the ionization process ($\sim 10^{-21}$ cm^2) approximately $\sim$1000 times more than the cross-section of the nuclear interaction ($\sim 10^{-24}$ cm^2). Strictly speaking for monoenergetic beam of particles R_0 (the distance of particle run) has the meaning of the middle run distance relatively to which it has the Gauss distribution. During one act of ionization the particle energy is diminished approximately 3.5 eV [153] and probability of nuclear interaction depends on the particle energy at the moment of the particle and nuclear collision. The number of nuclear interactions in thin layer dx at the depth x from the surface of the target has the form

$$d\nu = \eta(x) N \sigma(x) dx \simeq \eta_0 N \sigma(x) dx, \qquad (72)$$

where N is concentration of target nucleus, η_0 and η are the particles beam on the surface and in depth x, respectively; $\sigma(x)$ is the cross-section of interaction. The full number of interactions in the depth of substance, the thickness of which is more than R_0, is equal

$$\nu = \eta_0 N \int_0^{R_0} \sigma(x) dx = \eta_0 N \int_0^{E_0} \frac{\sigma(E) dE}{\left|\frac{dE}{dx}\right|}, \qquad (73)$$

where E_0 is the start energy of particle. Yield of nuclear reaction which is determined by the part of particles undergoes the nuclear interaction.

$$V(E) = \frac{\nu}{\eta_0} = N \int_0^{E_0} \frac{\sigma(E) dE}{\left|\frac{dE}{dx}\right|}. \qquad (74)$$

From the last formula, it is followed that the yield of nuclear reaction at the energy of charged particle E is determined by the cross-section σ (E) and specific ionization of particle dE/dx. And, vice versa, it is known the functions dE/dx and dV/dE from (74) allow to find the cross-section of interaction

$$\sigma(E) = \frac{1}{N} \frac{dV}{dE} \left| \frac{dE}{dx} \right|. \qquad (75)$$

It is known (see e.g. [153]) that for the charged particle the magnitude dE/dx is proportional to the square root of particle charge, concentration of electrons in substance (n_e) and some function of the velocity $f(v) \sim 1/v$ and doesn't depend on particle mass:

$$\frac{dE}{dx} \sim Z^2 n_e f(v). \qquad (76)$$

The dependence dE/dx is permitted to recount the data on the motion of one concrete particle in concrete substance on the motion of another particles in another substances (see also [152]). So far dE/d$x = f(E)$, then in such a case taking the integral, we can obtain the full run of particle.

$$R = \int_0^{E_0} \frac{dE}{f(E)}. \qquad (77)$$

For example, α-particles ($R_{\alpha,x}$) in the substance X_Z^A can be defined on the run in air ($R_{\alpha,air}$) with an assistance of empirical formula [153]

$$R_{\alpha,x}(E) = 0.56 R_{\alpha,air}(E) A^{1/3}, \qquad (78)$$

where $R_{\alpha,air}$ in cm and relates in the air at the temperature 15°C and ambient pressure and $R_{\alpha,x}$ will be obtained in mg/cm^2.

The run of protons is connected with the run of α-particle of another formula, which works at $E \geq$ 0.5 MeV:

$$R_p = 1.007 R_\alpha(3.972) E - 0.2, \qquad (79)$$

where $R_\alpha(3.972)$ is a run of α particles with the energy of 3.972. The run of other charged particles with mass M_x (exclude electron) is connected with the run of protons in the next relation

$$R_x(E) = \frac{M_x}{M_p} R_p \left(\frac{M_p}{M_x} E \right). \qquad (80)$$

Owing to the fast retardation of charged particles, they can dope the layer of small thickness with not uniformly distribution of doped impurities on the depth. The possible reactions obtained with charged particles are described by Smirnov [151], where it was indicated the half-time decay. To conclusion of this part we should note that regular experiences in this field at the present time is absent.

2.4. Nuclear reaction under action of the γ-rays

The reactions of (γ, n)-, (γ, p)- and (γ, α) belong to reactions of splitting nuclears irradiated by γ-rays. These reactions are endoenergetical and have some energetic threshold. In common case the probability of (γ, n) reaction is more than the probability of (γ, p)- and (γ, α) reactions. The energy of γ-rays $\sim$10 MeV is called the (γ, n) reactions. At the E_γ $\sim$100 MeV the reactions with the creation of several particles (γ, 2n; γ, pn, etc.) are possible [151]. Inasmuch as the impurities can be produced in practically any substances with the assistance of photonuclear reactions, therefore they are represent the interest for semiconductor doping. The main reason of this circumstance is large penetrating capability compared to the charged particles [151]. This circumstance can guarantee the uniform doping of the large volume substance for some impurity atoms. The perspective of this direction at the present time is also supported by the possibility of receiving of the γ-quantes with any energy at retarding electron emission. For that the monoenergetical electrons with the energy 25–60 MeV received by means of the accelerator, is directed on the target from heavy metals (Pb, Bi, W, U etc.). As a result the retarding electrons create the continuous spectrum of γ-emission. The maximum energy of this continuum is equal to the kinetic energy of electrons E_e and intensity of γ-emission approximately inversely proportional to the γ-quantes energy

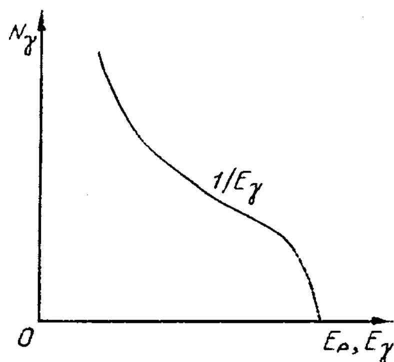


Figure 40 The spectrum of γ-emission is created under electrons retarding with E_e on the target of the accelerator (after [151]).

(see Fig. 40). The increase of γ-emission use for irradiation Si and other semiconductors has been noted [176].

2.5. Nuclear reactions under the influence of neutron

The main hope in the field of nuclear doping of semiconductors at the present time is placed on the neutrons (see also [177]) inasmuch as these particles are neutral particles and they possess a large penetrating capability. Furthermore neutron can interact practically with all nuclears. At the preset time there are the different sources of neutrons with different intensity accepted and approached (see above). The history of techniques by means of which was received the first neutrons beams can be found in paper [151]. The main source of the neutron beam is still nuclear reactors of different types (see also above). Typical neutron spectrum of the research water-water reactor WWR is given in review [178]. In this spectrum there are the neutrons with energy $0.01 \leq E \leq 10^7$ eV. At the irradiation of neutron beam the cross-section of compound nuclear σ_x essence depends on the degree of resonance to intrinsic magnitude of energy levels of compound nuclear.

For neutrons with the energy $E_n \sim 10^{-2} \div 10^4$ eV [178] we can write $\sigma(n, \gamma)$ for cross-section of nuclear reaction

$$\sigma(n,\gamma) = \sigma_0 (E_0/E_n)^{1/2} = \sigma_0(v_0/v), \qquad (81)$$

where σ_0, v_0 and E_0 are some constants of cross-section of the nuclear reaction, velocity and neutron energy taken as a starting point. As is well-known for NTD thermal neutrons are of greater interest. On this interval there are almost all number of neutrons in the spectrum of nuclear reactor which is described by Maxwell distribution on the velocities:

$$n(v) = \frac{4}{\sqrt{\pi}} \left(\frac{M_n}{2kT}\right)^{3/2} v^2 \exp(-M_n v^2/2kT), \qquad (82)$$

where as usually k is the Boltzman constant and T is temperature.

The average cross-section on the spectrum $n(v)$ provides

$$\bar{\sigma} = \frac{\int \frac{\sigma_0 v_0}{v} n(v) v \mathrm{d}v}{\int n(v) v \mathrm{d}v} = \frac{\sigma_0 v_0}{\bar{v}}, \qquad (83)$$

where $\bar{v} = \sqrt{8kT/\pi M_n}$ is the average magnitude of velocity. It is conveniently to use the next formula of more probable velocity of neutrons

$$v_{mp} = \sqrt{\frac{2kT}{M_n}} = \bar{v}\frac{\sqrt{\pi}}{2} = \bar{v}/1.228. \qquad (84)$$

This correspond to the maximum of distribution of $n(v)$. Taking into account that $v_0 = v_{mp}$ and $\sigma_0 = \sigma_{mp}$ and at ambient temperature ($T = 293$ K) v_{mp} is equal $v_{mp} = 2200$ m/s. In such a case we obtain the relation connected with the middle cross-section (n, γ) reaction on the thermal neutron with the magnitude of cross-section usually measurable on the neutrons with more probable velocity

$$\sigma = \frac{\sigma_{mp} v_{mp}}{\bar{v}} = \frac{\sigma_{mp}}{1.128} = \sigma_{2200}/1.128. \qquad (85)$$

The energetical dependence on the cross-section for Si can be described by following approximated equation [179]

$$\begin{aligned}\sigma &= \sigma_a + \sigma_s = \sigma_a + \sigma_{fa}(1 - e^{-2w}) \\ &= \frac{0.8}{\sqrt{E_n} + 2.25(1 - e^{CE_nT})},\end{aligned} \qquad (86)$$

where E_n is the energy of neutrons (MeV), σ_a is the cross-section of absorption which is equal 160×10^{-3} b at $E_n = 25$ MeV, σ_s is the cross-section of the scattering, σ_{fa} is the cross-section of free atom, which is equal 2.25 b, e^{-2w} is the Debye-Waller factor; $C = 1.439 \times 10^{-5}$ is the normalized constant of the dependence $\sigma(E)$ to the magnitude $\sigma = 0.55$ b at $E_n = 50$ MeV at 300 K. Calculations on the ground of Equation 80 the energetical dependence on the cross-section of Si at different temperatures as well as some eksperimental data are shown in Fig. 41. As it is seen for

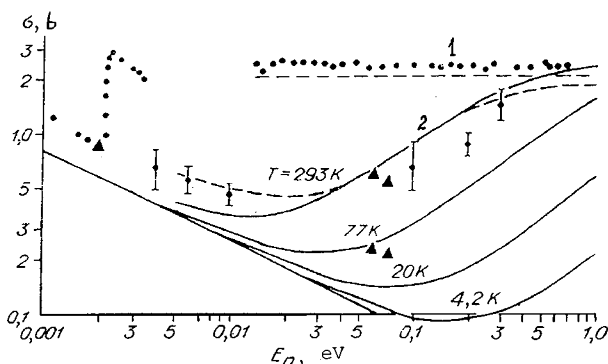


Figure 41 The dependence of the calculated on the formula the cross-section of the thermal neutrons in polycrystal (1) and monocrystal (2) of Si at different temperatures (after [151]).

polycrystals (curve 1) Si the cross-section in thermal field doesn't depend on the energy and for monocrystals (curve 2) this dependence is changed according to the law $1/v$ and depends on the temperature. The last fact indicates that the NTD of different semiconductors and one semiconductor with different impurities can irradiate at different temperature (details see [151]).

2.6. Influence of the dopants

According to common scheme of the nuclear reactions the primary products of (n, γ) reactions present the nuclear X_Z^{A+1} per one atomic unit heavier outcome ones. These can be heavier isotopes of the same or neighbor element of the periodical table. If these isotopes have in natural mixture isotopes irradiated material and appear stable ones, in such a case nuclear transformation leads to the change of the initial concentration of stable isotopes in the irradiated substance. In such a case this prices doesn't create impurities of neighbor elements of the periodical table. The possibility of the exchange of electrophysics properties with the assistance of impurities which are created by the method of nuclear doping at irradiation of neutrons proves for numerous semiconductors (see e.g., Table 2.6 in [151]). In this way there are different properties of semiconductors to connect with the nature and concentrations of the predominance impurities after respective annealing: for example Sn and Te in InSb [180–182], Ge and Se in GaAs [183–185], In in CdS [186] and phosphorus in Si [154, 181, 182, 187–190] etc.

For the estimation of the character the distribution of impurities on the thickness of the doped materials we'll outcome from fact that the attenuation of the intensity of collimated neutron flux by the layer of substance of the thickness x is confirmed by the well-know law:

$$I = I_0 \exp(-N\sigma x) = I_0 \exp(-\mu x) = I_0 \exp(-x/l_l), \quad (87)$$

where N is the amount of the atoms of the irradiated material per 1 cm^3; l_l is the middle length of the absorption [191] connected with the macroscopical absorption coefficient of the material $\mu = N\sigma$ next relation

$$l_l = 1/N\sigma = 1/\mu, \quad (88)$$

where l_l characterizes also the layer thickness of the material on which the neutron flux and respectively the impurities concentration is diminished in $e = 2.72$ times.

The mentioned above relations are correct, if to assume that the neutron cross-section of scattering is small compared to the absorption cross-section. In common case when there are both absorption and scattering of neutrons it is necessary to use the conclusions of the common theory of neutron diffusion [191]:

$$I = I_0 \exp(-x/L), \quad (89)$$

where L is the diffusion length.

For the compound semiconductors having some sorts of atoms, the attenuation of the emission beam can be considered as additive property of medium. Taking into account this fact we represent the Equation 81 in the following form

$$I = I_0 \exp\left(-\frac{\mu}{\rho} m\right). \quad (90)$$

Here ρ is the density of substance, μ/ρ is the mass coefficient of the attenuation; m is a mass of substance with the cross-section 1 cm^2 and thickness x. Then for compound substance we can write

$$\mu/\rho = \sum_i C_i (\mu/\rho)_i, \quad (91)$$

where C_i is the weight concentration of the i-element of the mixture. Equations 81–85 apply also to the attenuation of the narrow beam of γ-quantes. In this case the linear coefficient of the attenuation is determined (see also [175] and references therein) by the sum of the contributions of photo- and Compton effects, and also from creation process of electron-positron pairs, e.g.

$$\frac{1}{\mu} = N(\sigma_{ph} + \sigma_c + \sigma_p), \quad (92)$$

where σ_{ph}, σ_c and σ_p are the cross-sections of the irradiated processes of the interaction of γ-quantes with substance in estimation per one atom. The efficiency of the neutron absorption and γ-rays absorption of different semiconductor materials at one side irradiation is reflect in Table X. The magnitude of μ/ρ and respectively values l_n and l_γ are calculated using the Equation 85 and the knowledges about the cross-section of neutrons [192] and extrapolation data of mass coefficients of absorption of γ-radiation for some elements. The

TABLE X The efficiency of the attenuation of the neutrons and γ-rays by the different semiconductor materials (after [151])

Material	Density, $\rho(\frac{g}{cm^3})$	SN $N\sigma$ (cm^{-1})	SN l_n (cm)	SN L (cm)	GR $\mu/\rho(\frac{cm^2}{g})$	GR μ (cm^{-1})	GR l_γ (cm)
Si	2.42	0.008	125.0	22.2	0.024	0.058	17.2
Ge	5.46	0.25	4.0	4.7	0.39	0.213	4.7
GaAs	5.4	0.36	2.8	–	0.039	0.210	4.8
InSb	5.78	7.0	0.14	–	0.052	0.391	3.43
CdS	4.82	115.0	0.01	–	0.045	0.217	4.6

SN = Slow neutrons; GR = for γ-rays with energy $E = 20$ 30 MeV.

comparison of the data mentioned above in the Table X shows that for neutrons at increasing absorption ability from Si to CdS rather sharply decreases in the thickness of the semiconductor material layer in the boundary of which the distribution of impurities can be regarded homogenous. At that in Si the natural doping is inhomogeneous caused by the absorption and scattering of neutrons and has small and for real ingot with diameter 50 mm doesn't exceed 1% [155] and for crystal with diameter 80 mm the relation of concentrations in center ($C_{\min}$) and at the end ($C_{\max}$) of the ingot is combined

$$C_{\min}/C_{\max} \simeq 0.956 \text{ [193] (details see[151])}.$$

TABLE XI Energy available for producing displacement damage in silicon by neutron transmutation doping (after [195])

Silicon isotope	Concentration (10^{22} cm^{-3})	Cross-section (10^{-24} cm^2)	Nσ (cm^{-1})	νNσ (eV/cm)
a) Thermal Neutron Capture Recoil (Energy into displacement, $\nu = 780$ eV)				
^{28}Si $(n,\gamma)^{29}$Si	4.61	0.08	0.0037	2.88
^{29}Si $(n,\gamma)^{30}$Si	0.23	0.27	0.00062	0.49
^{30}Si $(n,\gamma)^{31}$Si	0.15	0.12	0.00018	3.51
β(1.5 MeV)$\rightarrow$ 2.61 hr $\rightarrow$ ^{31}P			Total	3.51
b) Fast Neutron Knock-on Recoil (Average recoil energy of 50 keV assumed, $\nu = 25$ keV)				
All 5 3(avg) 0.15 3.8 × 10^3				

2.7. Atomic displacement effects in NTD

Recoil atoms from γ-rays or particle emissions after thermal neutron capture, and recoil atoms from elastically and inelastically scattered fast neutrons, produce atomic displacements in solids [194]. The relative importance of the displacement damage produced by thermal and fast neutrons can be estimated by using isotope concentrations, capture or scattering cross-sections, and recoil energies. Such an estimate has been made for silicon and the results are presented in Table XI. The first three columns list the silicon isotopes, isotope concentrations, and cross-sections for thermal neutron capture [194] and for fast neutron scattering [195]. The probability of interaction (product of isotope concentration and capture or scattering cross-section) is listed in column 4. The energies available from silicon recoils for producing atomic displacements are given in column 5. For thermal neutrons, these energies were obtained from a product of the probability for thermal neutron capture and the average recoil energy [194] of 780 eV. Kirk and Greenwood [165] used 474 eV for the average silicon recoil energy, but the lower energy only emphasizes further the dominance of the fast neutrons for producing displacement damage (see also [195]). The details of the damage produced by recoiling ^{31}Si and ^{31}P atoms may, however, be important in determining the lattice location of ^{31}P introduced by NTD. Column 5 of Table XI shows that the energy available for displacements from an incident fast neutron is 10^3 times that from an incident thermal neutron. Fast neutrons will, therefore, dominate the displacement damage until thermal-to-fast ratios exceed 1000:1 (see also above). Thermal neutron capture cross-sections for germanium and gallium arsenide [195] are much larger than those for silicon (see also Table X). Consequently, displacement damage by thermal neutrons relative to fast neutrons is expected to be more important in these materials than it is in silicon. An atom recoiling in a host material creates a high-defect density (cluster of defects) along the recoil track [157]. The upper part of Fig. 42 illustrates clusters formed by recoil tracks calculated for a 50 keV silicon atom recoiling in silicon [196], and for a 10 keV germanium

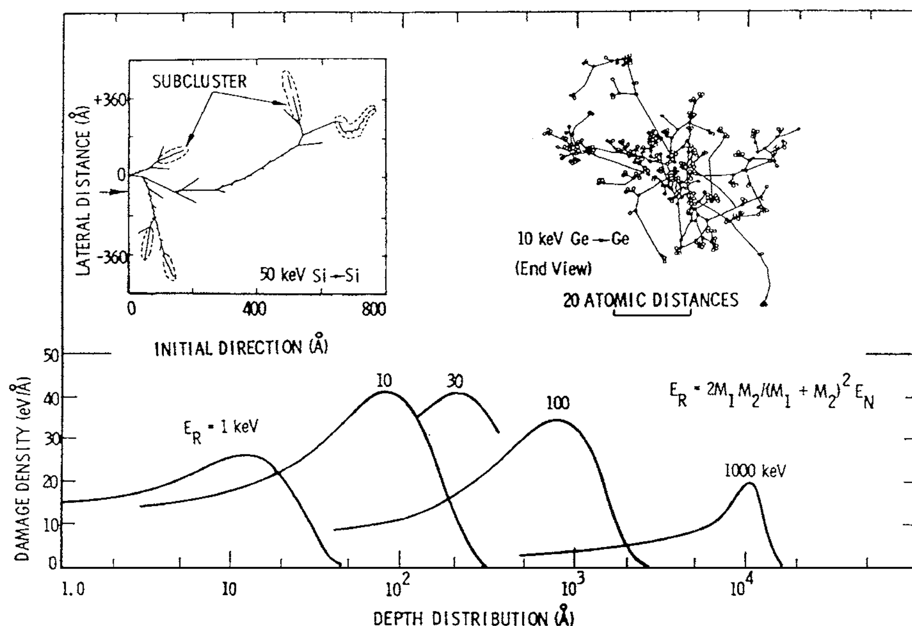


Figure 42 Shown in the upper left is a track for a 50 keV Si atom recoiling in Si according to calculation by Van Lint *et al.* [196]. Upper right is a displacement cascade for a 10 keV Ge atom recoiling in Ge according to calculations by Yoshida [197], where the open and solid circles represent interstitials and vacancies, respectively. The lower part of the figure shows the energy deposited per Å for different Si recoil energies E_r, according to the formulations of Brice [198] (after [195]).

atom recoiling in germanium [197]. The track in silicon as a side view and the authors [196] emphasize the high damage density in subclusters expected near the end of range for each recoiling silicon atom. These subcluster regions are similar to the capture-recoil damage regions produced by thermal neutron capture. Therefore, the kind of damage regions produced by thermal neutrons is included in fast neutron damage. The track in germanium is an end view where the open circles, which represent interstitials, are shown to be concentrated on the periphery of the track. The lower part of Fig. 42 shows results of calculating the depth distribution of displacement damage by averaging over a large number of tracks for silicon atoms recoiling with energy E_r in silicon [198]. The peak damage density first increases with E_r but then decreases because the damage clusters become more diffuse. Knock-on recoils in reactor-neutron irradiations have an average $E_r \approx 50$ keV. The areas under the curves give the energy spent in collision processes, and is ≈ 25 keV for $E_r = 50$ keV. The other half of E_r is spent in ionization [198]. Assuming 25 eV/displacement, there is sufficient energy to produce 1000 displacements.

One of the most graphic early examples of defect clusters produced by neutron bombardment of semiconductors was obtained by Bertolotti *et al.* [199] using etched surface replication transmission electron microscopy (TEM). Fig. 43 is a sketch taken from such results obtained on 14 MeV neutron irradiated silicon [200]. Most people would agree that the central region is probably due to the core of the displacement-damage clusters. There is less agreement on the interpretation of the outer zone. Direct TEM mesurement [201, 202] indicate a strained region around the damage core so that strain-induced differential etching may have caused the outer zone observed in papers [199, 200]. Model by Nelson [203] postulates trapping of mobile defects by

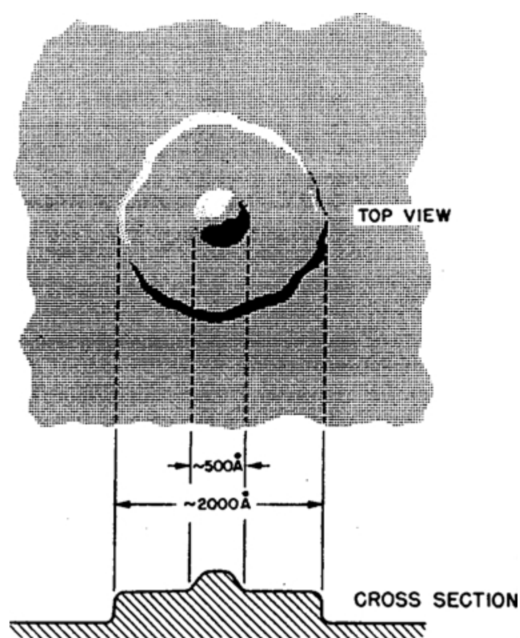


Figure 43 Sketch of a region observed on an etched surface of Si following 14 MeV neutron irradiation. Central core (500 Å) is attributed to recoil damage (after [195]).

damage clusters so that the differential etching may have caused by ab excess of trapped vacancies or interstitials. However, the interpretation that has been most extensively used to explain experimental data is due to Gossick [204] and Crawford and Cleland [205]. In this model, the outer zone represents a space charge region surrounding a p-type germanium. The damage cluster in silicon is nearly intrinsic so that a space charge region would be found in both n- and p-type silicon. The space charge model has been used to interpret minority carrier lifetime data [206], changes in carrier concentration and carrier mobility [207], the light sensitivity of neutron-produced electrical changes [207], photoconductivity [208] and EPR observations on specific defects in neutron irradiated silicon (see also [182] and references therein).

Thus calculations of energy deposition into displacement processes show that fast neutrons will dominate defect production in NTD processing of silicon unless thermal-to-fast neutron ratios exceed 1000:1. Defect clusters are produced by silicon-atom recoils from fast neutron collisions. Using an experimental value for the energy needed per unit volume to form amorphous material, it is argued in paper [195] that amorphous zone formation in silicon NTD is highly unlikely.

2.8. Experimental results

2.8.1. Ge

As was noted above, neutron transmutation (NT) is especially intriguing for semiconductors for several reasons. First, the NT process can create new elements removed by just one atomic number. Considering for the moment the elemental group-IV semiconductors Ge and Si, this means that the donors As and P will be created, respectively, following neutron capture and β-decay of isotopes of these semiconductor elements. The new elements are, of course, the prototypical donors. Neutron capture leads to NTD. Second, the number of new atoms ${}^{A+1}_{Z+1}\mathrm{N}$ created is simply (see also 2.1)

$$ {}^{A+1}_{Z}\mathrm{N} = n\sigma_{\mathrm{n}}{}^{A}_{Z}\mathrm{N}, \tag{93}$$

where n being the total neutron fluency (cm^{-2}), σ_n the cross-section for thermal neutron capture (cm^2), and ${}^{A}_{Z}\mathrm{N}$ the atom concentration of the specific isotope in the given isotope mixture (cm^{-3}) (either natural or man made). Considering that the values of σ_n lie in the $10^{-23} \div 10^{-24}$ cm^{-2} range (see above), it recognizes that very large neutron fluences are required to transmute a significant number of atoms of one element into another. Whereas this may pose problems to fulfill the medieval alchemist's dream, it is just perfectly suited for the person who wants to dope semiconductors. With the thermal neutron fluences available in modern nuclear reactor (see also 2.2) one can dope Ge up to the metal-insulator (MI) transition (2 to 3×10^{17} cm^{-3}) while Si can be doped with phosphorus to several times 10^{15} cm^{-3} [182]. As will be shown below this is due to the small atom concentration of ${}^{30}_{14}\mathrm{Si}$ and the modest value of the thermal capture cross-section. Third, there are elements which have light isotopes which

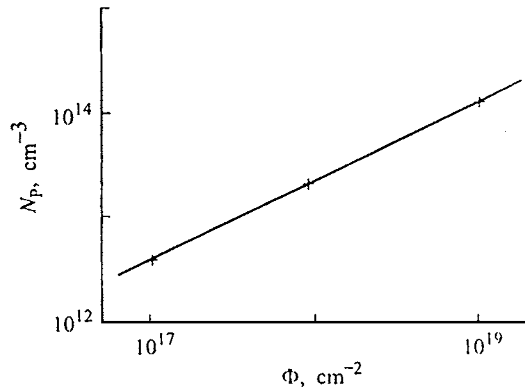


Figure 44 The dependence of the phosphorus atoms concentration on the neutron irradiation doze of Si crystals and followed annealing at 800°C during 1 h. The dependence was measured by Hall effect (after [215]).

upon neutron capture transmute to a lower Z element either by electron capture or by positron decay. In this case acceptors are created. A classical case is the transmutation of $^{70}_{32}$Ge into $^{71}_{31}$Ga.

The main advantage of the NTD method, as we know at present, is the precision doping which is connected with the linear dependence of concentration of doping impurities on the doze of neutron irradiation. Such dependence is numerous observed in the different experiments (see, e.g. [209–214]). As an example, in Fig. 44 there is shown the dependence of the concentrations doped phosphorus on the doze of irradiation the Si crystal in nuclear reactor. This dependence was measured with the help of Hall effect [198]. However, at the large doze of neutron irradiation there is observed the nonlinear dependence. On Fig. 45 is shown the results of paper [216] where was observed the deviation from linear law at the large doze of neutron irradiation of the sample of ^{74}Ge which was annealed after irradiation at $T = 460°$C during different time (see also caption of Fig. 45). More amazing effect was observed at the second irradiation of the samples of ^{74}Ge previously strong doped with As by NTD method. Instead expectable increase of the concentration free charges (electrons) n there is observed the decrease n. This decrease was direct proportional to the neutron irradiation doze of ^{74}Ge crystals. Both effects are rather details analyzed in papers [213, 216].

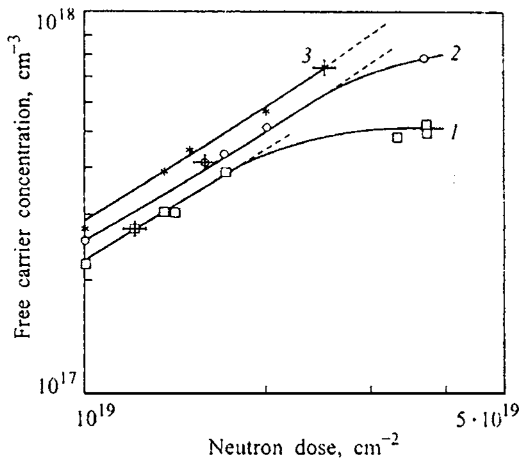


Figure 45 The dependence of the concentration of free electrons in ^{74}Ge NTD on the irradiation of the thermal neutron doze and followed annealing at 460°C during 24 (1), 50 (2) and 100 (3) hs. (after [216]).

The transmutation of the stable germanium isotopes via capture of thermal neutrons is well understood. Table XII contains all the information relevant to NTD of germanium. In paper of Haller *et al.* [209] quoted the values of the thermal neutron capture cross-section σ_n of three sources [181, 190, 217]. The information of Table XII permits the computation of the acceptor and donor concentrations for a known neutron exposure. Not only are these concentrations important but the ratio of the sum of all minority dopants (donors) and the sum of all majority dopants (acceptors) i.e., the compensation K, is crucial for the low temperature conduction. For the case of germanium, one obtains K from the following equation (see also [209])

$$K = \left(\sum \text{donors} \cdot \text{cm}^{-3}\right) \Big/ \left(\sum \text{acceptors} \cdot \text{cm}^{3}\right) = (N_{\text{As}} + N_{\text{Se}})/N_{\text{Ga}}. \quad (94)$$

The substitutional selenium impurities are double donors providing two electrons for compensation. Therefore they are counted twice in the sum of donors. Using the different values for σ_n, one finds K ranging from 0.322 to 0.405 for crystals with negligible initial donor and acceptor concentrations. It would be of great help for both the basic understanding of the hopping conduction [218] as well as for application of neutron-transmutation-doped germanium as, for example, bolometer material [209], if these cross-sections could be accurately evaluated in one or more well characterized nuclear reactors (see above).

In order to obtain the above K values and thus take full advantage of NTD, Haller *et al.* choose the purest

TABLE XII Characteristics of the transmutation process of germanium (after [209])

Isotope	Abundance (%)	NCCS (barn)	NCDR	Dopant type
$^{70}_{32}$Ge	20.5	3.4; 3.2; 3.25	$^{70}_{32}$Ge$(n, \gamma) \rightarrow$ $^{71}_{32}$Ge $\rightarrow$ $^{71}_{31}$Ga	p
$^{72}_{32}$Ge	27.4	0.98; 1.0; 1.0	$^{72}_{32}$Ge$(n, \gamma) \rightarrow$ $^{73}_{32}$Ge	
$^{73}_{32}$Ge	7.8	14.0; 14.0; 15.0	$^{73}_{32}$Ge$(n, \gamma) \rightarrow$ $^{74}_{32}$Ge	
$^{74}_{32}$Ge	36.5	0.62; 0.5; 0.52	$^{74}_{32}$Ge$(n, \gamma) \rightarrow$ $^{75}_{32}$Ge $\rightarrow$ $^{75}_{33}$As	n
$^{76}_{32}$Ge	7.8	0.36; 0.2; 0.16	$^{76}_{32}$Ge$(n, \gamma) \rightarrow$ $^{77}_{32}$Ge $\rightarrow$ $^{77}_{33}$As $\rightarrow$ $^{75}_{34}$Se	n

NCCS = Neutron capture cros-sections.
NCDR = Neutron capture and decay reactions.

available Ge crystals as a starting material. Germanium is, in this respect, ideally suited for NTD because it can be purified at present time down to concentrations of $\lesssim 10^{11}$ cm^{-3} (see e.g. [219]). Such low concentrations are negligible when compared with the dopant concentrations after NTD in the low 10^{16} cm^{-3} range. According to Haller *et al.* [209] the concentrations od electrically inactive impurities such as hydrogen, carbon, oxygen and silicon can be as high as 10^{14} cm^{-3}. Of all the isotopes of these impurities only $^{30}_{14}$Si transmutes to an electrically active impurity, phosphorus, a shallow donor. With only one silicon atom in every 4.4×10^8 germanium atoms and only 3% os all silicon atoms being $^{30}_{14}$Si which has a neutron capture cross-section much smaller than the germanium isotope cross-sections, Haller *et al.* can estimate that less than one phosphorus donor is produced for every 10^{11} gallium majority acceptors during the NTD process. These authors concluded that ultra-pure germanium crystals are virtually perfect starting material. For the NTD study, they have chosen an ultra-pure germanium single crystal which they have grown at the crystal growth facility described early (see also [209, 219] and references therein).

The measured resistivities (ρ) in paper [209] are presented in Fig. 46. The results of these measurements yield the mobility μ:

$$\begin{aligned} \rho &= (p\mu e)^{-1}, \\ R_{\mathrm{H}} &= (pe)^{-1} \\ \mu &= R_{\mathrm{H}}/\rho, \end{aligned} \tag{95}$$

where p is free hole concentration, e is charge of the electron and R_{H} is the magnitude from Hall measurement.

The mobility values are only useful down to the temperature where hopping conduction sets in. The mobility values agree well with published values for melt-doped material in the temperature range above the hopping regime. This indicates that the concentration of residual radiation damage or other free-carrier scattering centers must be very small. Fig. 46 shows the log (resistivity) versus $1000/T$ dependence for six NTD germanium samples. The number next to each curve corresponds to the acceptor (gallium) concentration in each sample. For comparison Haller *et al.* have also measured gallium-doped germanium samples which have extremely small values of K. These so-called uncompensated samples were cut from crystals which were doped in the melt and were grown in the ultra-pure germanium crystal-growing equipment, and not NTD doped. The compensating donor concentration

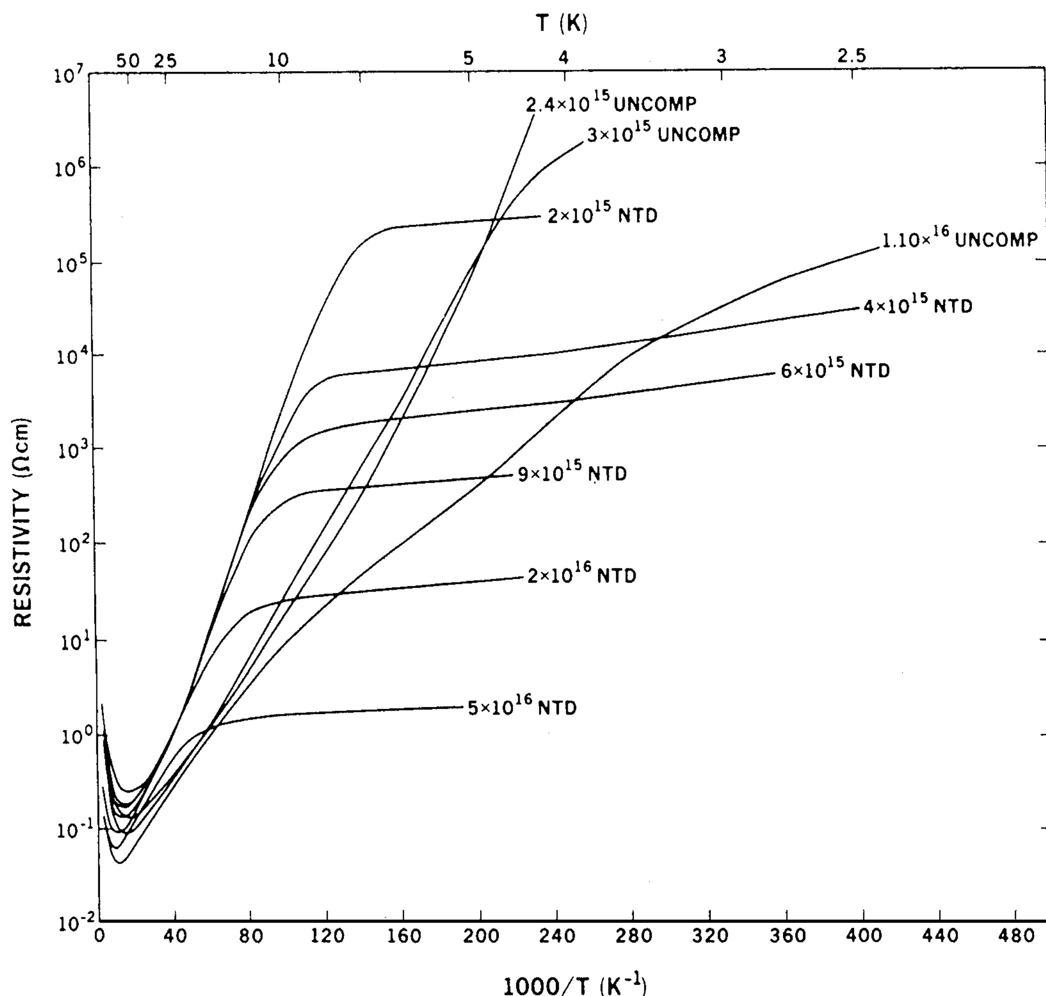


Figure 46 Resistivity as a function of $1000/T$ for NTD and uncompensated germanium samples. Each curve is labeled by the gallium concentration obtained by either NTD or melt doping (after [209]).

in these crystals is estimated to be less than 10^{11} to 10^{12} cm^{-3}. The resulting K is of the order of 10^{-4} to 10^{-5}. The resistivity-temperature dependence of these NTD samples is characterized by three regimes. At high temperatures (room temperature down to about 50 K), the resistivity decreases because the carrier mobility increases. Below about 50 K carrier freeze out begins and reduces the free hole concentration rapidly. The slope of the freeze out in highly-compensated material os proportional to the acceptor binding energy $E_A - E_V \simeq 11$ meV. At still lower temperatures, the appearance of hopping conduction causes the resistivity to increase only very slowly. All six NTD germanium samples show these three resistivity regimes very clearly. The low-compensated samples show different $\log(\rho)$ versus $1/T$ dependences. A third conduction mechanism has been proposed for such material [220]. It is based on the idea that carriers can "hop" from neutral to a neighboring neutral acceptor thereby forming a positively charged acceptor. The NTD process in high-purity germanium leads to a fixed compensation which in turn results in a certain slope of the $\log(\rho)$ versus $1/T$ dependence for a given neutron exposure.

2.8.1.1. Metal-insulator transition. In next part of this paragraph let us briefly discussed the metal-insulator transition (MIT) [210, 218, 221, 222] in transmuted Ge. In the literature there is an intensive debate whether MIT is a phase transition of first or second order and what are the experimental conditions to obtain it at finite temperatures and in real (disordered) system (see, e.g. [211, 212, 214, 223]). If the MIT is as second order phase transition a further challenge is the solution of the so called puzzle of the critical index, μ for the scaling behavior of the metallic conductivity near the MIT, i.e., just above the critical impurity concentration N_c and as small compensation, K. According to the scaling theory of the MIT for doped semiconductors [222], the conductivity at zero temperature $\sigma(0) = \sigma(T \to 0)$, when plotted as a function of impurity concentration N, is equal to zero on the insulating side of the MIT and remains finite on the metallic side, obeying a power law in the vicinity of the transition,

$$\sigma(0) \propto [(N/N_c) - 1]^{\mu}, \quad (96)$$

where N_c is the critical impurity concentration of the given system and μ is the critical conductivity exponent. The value of μ, determined experimentally, is compared with theoretical predictions. Up to present time $\mu \approx 0.5$ has been obtained with nominally uncompensated semiconductors (Si:P [224], Si:As [225, 226], Ge:As [227], Si:B [228]) while $\mu \approx 1$ has been found with compensated semiconductors (Ge:Sb [229], Si:P,B [230], Ge:Ga, As [231]) and amorphous alloys [232–234]. Exceptions are uncompensated Ge:Sb with $\mu \approx 1$ [210] and Ga_xAr_{1-x} amorphous alloys with $\mu \approx 0.5$ [235, 236]. As was shown in [211] the value $\mu \approx 0.5$ obtained with simple systems like uncompensated semiconductors turns out to be inconsistent with theoretical prediction [221, 222, 234]. In his original theory Mott considered only the electron-electron (e^-–e^-) interaction (Mott transition) and predicted a discontinuous transition of $\sigma(0)$ at N_c [237]. Although there is much evidence for the importance of e^-–e^--interactions, no experimental observation of such an abrupt transition has been reported. Anderson's ides of MIT is based solely on the disordered potential arising from randomly distributed dopants (Anderson transition) [238]. This lead to the development of the well-known "scaling theory" which predicted $\mu \approx 1$ for three dimensional systems (see also [222] and references therein). More recently, higher order calculations of the scaling theory (exclusively with disorder and no interactions) predict $\mu \approx 1.3$ [239], and more importantly, this value is shown to be independent of time reversal invariance [240] and of strength of spin-orbit interactions [241] (see also [212]). It is therefore clear that the effect of disorder alone cannot explain the experimental results of $\mu \approx 0.5$ or 1. Chayes *et al.* combined the theories of Mott and Anderson and successfully set the lowest limit $\mu > 2/3$ [242]. This result permits $\mu \approx 1$ obtained with compensated semiconductors and amorphous alloys. However, there still is no theory which convincingly explain $\mu \approx 0.5$ found for uncompensated semiconductors.

Even with today's advanced semiconductor technology, melt-doping of bulk semiconductors always leads to inhomogeneous dopant distributions due to impurity segregation and striation during crystal growth [223]. In papers [211, 223] this difficulties have overcome by applying the NTD technique to a chemically pure, isotopically enriched ^{70}Ge [211] and ^{74}Ge [223] crystals. The ^{70}Ge crystal of isotopic composition [^{70}Ge] = 96.2 at.% and [^{72}Ge] = 3.8 at.% was grown in paper [211] using the Czochralski method developed for ultra-pure Ge [219]. The as-grown crystal was free of dislocations, p-type with an electrically active net-impurity concentration less than 5×10^{11} cm^{-3}. In paper [223] was used isotopically engineered germanium which was grown from pure ^{74}Ge, enriched up to 94%, or by addition of a controlled portion of Ge with natural isotopic content to the ^{74}Ge material. In this way both, the doping as well as the compensation, are very homogeneous due to the NTD and the compensation by controlled mixtures of ^{74}Ge and ^{70}Ge which transmute to ^{75}As donors and ^{71}Ga acceptors. Four series of n-type NTD-Ge with different K were grown [223]. The values of K are proportional to the product of the isotopic abundance and the thermal neutron cross-section of all isotopes producing impurities (see above): $K = N_{Ga}/(N_{As} + N_{Se})$, whereas the impurity concentration is additionally proportional to the irradiation doze. A very small fraction of ^{72}Ge becomes ^{73}Ge which is stable, i.e., no other acceptors or donors are introduced. Use NTD since it is known [145, 243] to produce the most homogeneous, perfectly random dopant distribution down to the atomic level. Fig. 47 shows the temperature dependence of the resistivities (ρ) of 14 insulating samples in the range $N = 0.16$–$0.99N_c$ for NTD ^{70}Ge:Ga crystals. The analogous picture for NTD ^{74}Ge:Ga is shown in Fig. 48. All curves become linear only when $\ln\rho$ is plotted against $T^{-1/2}$ in good agreement with theory of variable range hopping conduction for strongly interacting

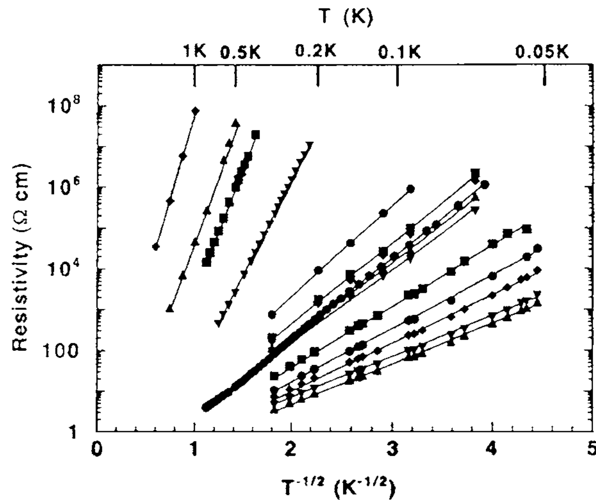


Figure 47 The logarithm of the resistivity plotted as a function of $T^{-1/2}$ for 14 insulating NTD ^{70}Ge:Ga samples. Gallium concentration from top to bottom in units 10^{16} cm^{-3} are 3.02; 8.00; 9.36; 14.50; 17.17; 17.52; 17.61; 17.68; 17.70; 17.79; 17.96; 18.05; 18.23 and 18.40 (after [211]).

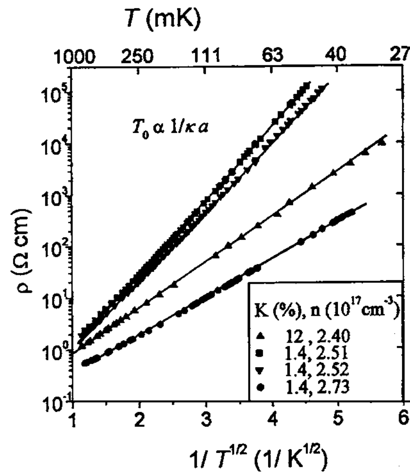


Figure 48 Typical temperature dependences of the resistivity as a function of $T^{-1/2}$ for four samples NTD ^{70}Ge:Ga crystals (after [223]).

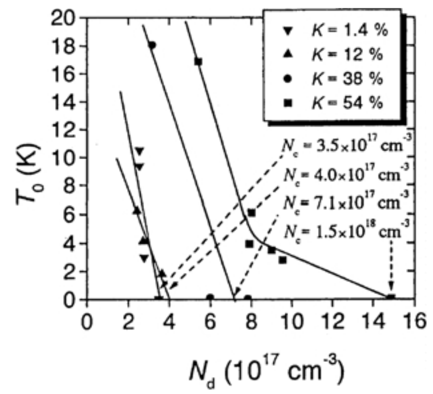


Figure 49 Determination of N_c from the extrapolation $T_0 \to 0$ in the range $T_0 > T$ (after [223]).

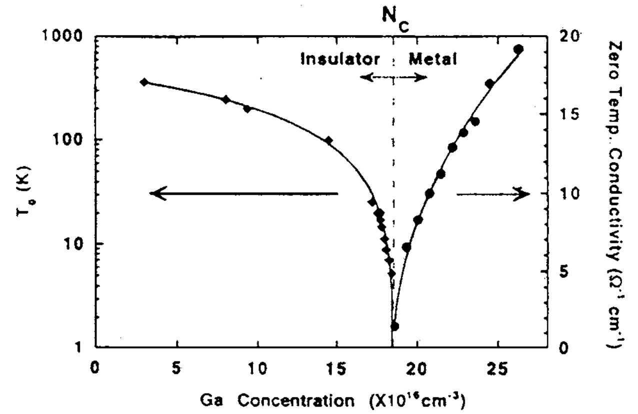


Figure 50 The left side shows (^{70}Ge) as a function of Ga concentration (◇). The solid curve is the best fit obtained with Equation 99 (with $\alpha \approx 1$). The right side shows the zero temperature conductivity σ(○) obtained from the extrapolation in Fig. 51 for the metallic samples as a function of Ga concentration (●). The solid curve is the best fit obtained with Equation 99 (after [211]).

electrons [218]:

$$\rho = \rho_0 \exp(T_0/T)^{1/2}, \tag{97}$$

where ρ_0 is a prefactor and T_0 is given by

$$T_0 \approx 2.8\mathrm{e}^2/k(N)\xi(N), \tag{98}$$

where $k(N)$ and $\xi(N)$ are the dielectric constant and localization length depending on N, respectively. Moreover, $k(N) \propto [N_c/(N_c - N)]^s$ and $\xi(N) \propto [N_c/(N_c - N)]^\zeta$ as N approaches N_c from the insulating side so that T_0 becomes [218]

$$T_0 = A[(N_c - N)N_c]^\alpha. \tag{99}$$

Here $\alpha = s + \zeta$ is to be determined experimentally [211].

Fig. 49 shows the dependence of T_0 as function of $N_d = n/(1 - K)$ for different K of ^{74}Ge:Ga [223]. As early these authors used the intersection point of these dependencies with the x-axis as a tool for the detrmination of $N_c(K)$. The left half of Fig. 50 shows the experimentally determined T_0 versus [Ga] (filled diamonds) together with the result of a three-parameter-fitting using A, N_c and α as variables in Equation 99 (solid curve) [211]. These authors deduced [Ga] for sample using following equation $[^{71}\mathrm{Ga}](\mathrm{cm}^{-3}) = 0.1155 \times n$ (cm^{-2}), since it was known the precise neutron fluency used in each irradiation. The best fit of T_0 with Equation 99 was obtained with the values $\alpha = 1.03 \pm 0.038$ and $N_c = (1.855 \pm 0.012) \times 10^{17}$ cm^{-3}. A much larger value of $\alpha \approx 2$ has been reported for Ge:As using only three samples with the highest N being far from the transition $0.56N_c$) [244]. In paper [211], it has been obtained $\alpha = 1$ with 14 homogeneously doped samples of [Ga] = 0.16–0.99N_c, all demonstrating the lh$\rho \propto T^{-1/2}$ dependence, i.e., this data set should be considered to be the first reliable determination of the hoping conductivity exponent α for a particular semiconductor system.

Fig. 51 shows the conductivity σ according to the results of paper [211] in ten metallic samples plotted against $T^{1/2}$. Extrapolation of each curve to $T = 0$ K, i.e., the determination of the zero temperature conductivity $\sigma(0)$, yields a very small error since the dependence of σ on T for all samples is very weak. The right half of Fig. 50 shows $\sigma(0)$ as a function of [Ga] (filled circles) together with a fit obtained by the scaling expression Equation 99 (solid curve). The values of the parameters determined in paper [211] from this fit are

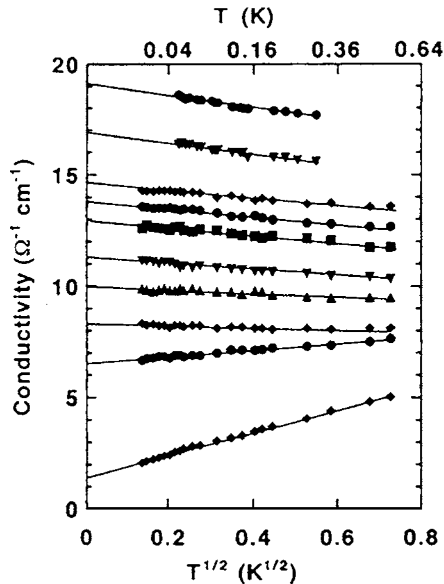
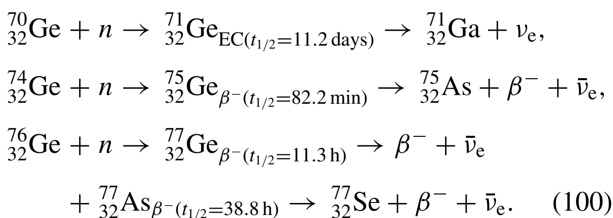


Figure 51 Conductivity plotted as a function of $T^{1/2}$ for 10 metallic NTD ^{70}Ge:Ga samples. Sold lines indicate extrapolation to $T = 0$ K. Gallium concentration from top to bottom in units of 10^{16} cm^{-3} are 18.61; 19.33; 20.04; 20.76; 21.47; 22.19; 22.90; 23.62; 24.50 and 26.25 (after [211]).

$\mu = 0.502 \pm 0.025$ and $N_c = (1.856 \pm 0.003) \times 10^{17}$ cm^{-3}. This value present $\mu \approx 0.5$ for uncompensated Ge:Ga semiconductors with high confidence, since the two values of N_c obtained from the scaling of T_0 [Equation 99] and $\sigma(0)$ [Equation 96] agree perfectly (details see also [214, 223]).

2.8.1.2. Neutral-impurities scattering. The low-temperature mobility of free carriers in semiconductors is mainly determined by ionized- and neutral-impurity scattering. The ionized-impurity scattering mechanism has been extensively studied (see e.g. [245] and references therein), and various aspects of this process are now quite well understood. Scattering by neutral impurities (see also [177]) is much less than by ionized centers, i.e., its contribution is significant only in crystals with low compensation and at very low temperatures where most of the free carriers are frozen on the impurity sites. The availability of highly enriched isotopes of Ge which can be purified to residual dopant levels $<10^{12}$ cm^{-3} has provided the first opportunity to measure neutral impurity scattering over a wide temperature range. Three Ge isotopes transmute into shallow acceptors (Ga), shallow donors (As) and double donors (Se) (see also above):



$$^{70}_{32}\mathrm{Ge} + n \rightarrow\ ^{71}_{32}\mathrm{Ge}_{\mathrm{EC}(t_{1/2}=11.2\,\mathrm{days})} \rightarrow\ ^{71}_{32}\mathrm{Ga} + \nu_e,$$

$$^{74}_{32}\mathrm{Ge} + n \rightarrow\ ^{75}_{32}\mathrm{Ge}_{\beta^-(t_{1/2}=82.2\,\mathrm{min})} \rightarrow\ ^{75}_{32}\mathrm{As} + \beta^- + \bar{\nu}_e,$$

$$^{76}_{32}\mathrm{Ge} + n \rightarrow\ ^{77}_{32}\mathrm{Ge}_{\beta^-(t_{1/2}=11.3\,\mathrm{h})} \rightarrow \beta^- + \bar{\nu}_e + \ ^{77}_{32}\mathrm{As}_{\beta^-(t_{1/2}=38.8\,\mathrm{h})} \rightarrow\ ^{77}_{32}\mathrm{Se} + \beta^- + \bar{\nu}_e. \quad (100)$$

The isotopes ^{72}Ge and ^{73}Ge are transmuted into the stable ^{73} Ge and ^{74}Ge respectively. Controlling the ratio of ^{70}Ge and ^{74} Ge in bulk Ge crystals allows fine tuning of the majority- as well as the minority carrier concentration. Currently, this is the best method to vary the free-carrier concentration independently from compensation ratio. As opposed to other doping methods, NTD yields a very homogeneous, perfectly random distribution of the dopants down to the atomic levels [246]. Thus isotopically controlled crystals offer a unique possibility to study systematically the scattering mechanism of the charge carriers in semiconductors. Extensive Hall-effect and resistivity measurements from room temperature down to 4.2 K yielded very accurate free-carrier concentrations and mobilities as a function of temperature and doping level were done in papers [105, 247, 248]. Itoh *et al.* have performed temperature-dependent Hall measurements on four different p-type and two-different n-type Ge crystals. The n-type crystals were obtained through NTD of isotopically enriched ^{74}Ge, and the p-type crystals correspondingly from NTD of isotopically enriched ^{70}Ge. The neutron cross-section for the neutron capture of the isotope for these irradiations were determined to be $\sigma_c(^{70}\mathrm{Ge}) = 2.5(5) \times 10^{-24}$ cm^2 and $\sigma_c(^{74}\mathrm{Ge}) = 0.6(1) \times 10^{-24}$ cm^2 by Itoh *et al.* [249]. To remove structural defects due to the unintentional irradiation with fast neutrons, all samples had to be thermally annealed at 650°C for 10 s in a rapid thermal annealer. Hall mobility obtained from the conductivity and free-carrier concentration data (listed in the Table XIII) are displayed in Fig. 52. A magnetic field of 3 kG was used, that is, for the temperature range of interest for the neutral impurity scattering the high-field limit $\mu B \gg 1$ is satisfied and the Hall mobility can be equated with the drift mobility.

TABLE XIII Carrier concentration of the Ge crystals used in the work of Fuchs *et al.* [105]

p-type	$N_A - N_D$	N_D	$K = N_D/N_A$
Ge:Ga #1	3.1×10^{14}	3×10^{12}	9×10^{-3}
Ge:Ga #2	7.7×10^{15}	9×10^{13}	1.2×10^{-2}
Ge:Ga #3	1.7×10^{16}	2×10^{14}	1.2×10^{-2}
Ge:Ga #4	1.0×10^{15}	1.2×10^{13}	1.2×10^{-2}
n-type	$N_D - N_A$	N_A	$K = N_A/N_D$
Ge:As #1	3.5×10^{14}	8.5×10^{12}	2.4×10^{-2}
Ge:As #2	1.2×10^{15}	1.2×10^{13}	1.0×10^{-2}

Fuchs *et al.* [105] analyzed the mobility data of Fig. 52 in terms of scattering of the carriers from phonons (μ_1), ionized impurities (μ_i) and neutral impurities (μ_n) assuming next rule

$$\frac{1}{\mu} = \frac{1}{\mu_1} + \frac{1}{\mu_i} + \frac{1}{\mu_n} \quad (101)$$

To extract the neutral impurity scattering contribution, they subtracted $1/\mu_1 + 1/\mu_i$ from the measured $1/\mu$. The relative contributions of phonon scattering ($1/\mu_1$), ionized impurity scattering ($1/\mu_i$) and the resulting neutral impurity scattering ($1/\mu - 1/\mu_1 - 1/\mu_i$) are plotted in Fig. 53 (data Ge:Ga #2). For $T > 80$ K, phonon scattering is the dominant scattering mechanism. On comparison of Figs 52 and 53, it becomes clear that the "dip" in the carrier mobility around 50 K is caused by scattering from ionized impurities, which dominate the scattering of the carriers between 20 and

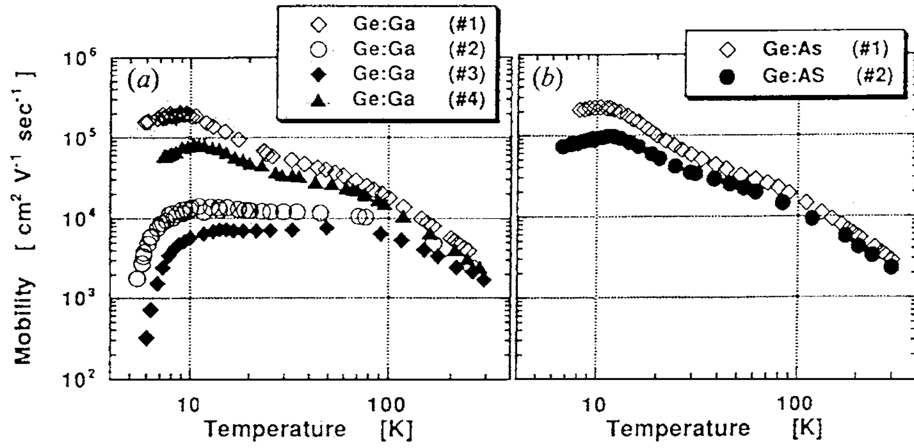


Figure 52 Temperature dependence of the carrier mobility of (a) p-type and (b) n-type NTD Ge crystals (after [105]).

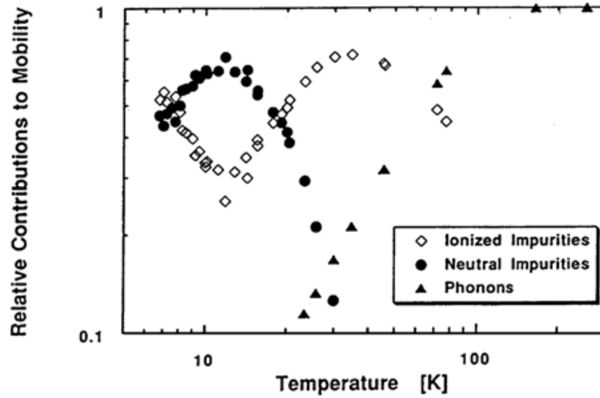


Figure 53 Temperature dependence of the relative contributions to the mobility. Note that the mobility is dominated by neutral impurity scattering below 20 K (^{70}Ge:Ga #2 crystal) (after [105]).

80 K. The flattering and saturation of the mobilities below 20 K originate from neutral impurity scattering, which can only be observed in crystals with very high crystalline quality and low compensation like isotopically enriched NTD Ge crystals used in paper [247].

The experimental data, obtained in [247] allow these authors quantitative comparison with theory. According to Erginsoy [250], the inverse relaxation time τ^{-1}, the scattering rate, for neutral-impurity scattering equals:

$$\tau^{-1} = \frac{20kN_{\mathrm{N}}\hbar^3}{m^{*2}e^2}, \tag{102}$$

where k is dielectric constant, e is the electron charge, N_{N} is the neutral-impurity concentration, and m^* is the electron effective mass. Equation 102 can be considered only as a first-order approximation because the prefactor 20 is an empirically determined constant and only the lowest s partial wave is taken into account in the phase-shift calculation (see also [251, 252]). McGill and Baron [253] have used for $\tau^{-1}_{\text{neutral}}$ the following equation:

$$\tau^{-1}_{\text{neutral}} = \frac{4\pi N_{\mathrm{N}}\hbar e}{2km^*E_{\mathrm{B}}}\sum_{\mathrm{l}=0}^{\infty}\frac{(\mathrm{l}+1)}{4w^{1/2}}[3\sin^2(\delta_{\mathrm{l}}^- - \delta_{\mathrm{l}+1}^-) + \sin^2(\delta_{\mathrm{l}}^+ - \delta_{\mathrm{l}+1}^+)], \tag{103}$$

where E_{B} is the binding energy of the scattering centres, $w \equiv E/E_{\mathrm{B}}$ and E is the incident electron energy, and δ_{l}^+ and δ_{l}^- are the lth partial shift for the singlet and triplet states respectively. Authors [253] graphically showed the accurate $\tau^{-1}_{\text{neutral}}$ as a function of w for neutral-impurity scattering in semiconductors. This result has been considered as an appropriate model for neutral-impurity scattering in semiconductors and has been discussed in detail in many standard textbooks (see, e.g. [254]).

Meyer and Bartoli reevaluated this task and obtained an analytical expression that is essentially the same as the graphical solution of authors of paper [253] but covering a wider incident-electron energy range (see also [255]):

$$\tau^{-1}_{\text{neutral}} = \frac{A(w)kN_{\mathrm{N}}\hbar^3}{m_{\mathrm{H}}^{*2}e^2}, \tag{104}$$

with

$$A(w) = \frac{35.2}{w^{1/2}}\frac{(1+e^{-5\mathrm{w}})(1+80.6W+23.7w^2)}{(1+41.3w+133w^2)} \times\left[\frac{1}{w}\ln(1+w) - \frac{(1+0.5w-1.7w^2)}{(1+w)^3}\right]. \tag{105}$$

Here m_{H}^* is the hydrogenic effective mass given by

$$m_{\mathrm{H}}^* = \frac{E_{\mathrm{B}}k^2m_0}{E_{\mathrm{H}}}. \tag{106}$$

In last equation m_0 is the electron rest mass and $E_{\mathrm{H}} =$ 13.6 eV is the binding energy of hydrogen. In total-mobility calculation Itoh *et al.* [247] employ a standard relaxation-time approximation. This approach is valid because they are limiting to low temperatures ($T <$ 25 K) where the inelastic optical-phonon deformation-potential scattering is negligible. Three scattering mechanisms are considered: neutral-impurity, ionized-impurity, and acoustic-phonon deformation-potential scattering. The neutral-impurity scattering contribution was calculated using both Equations 102 and 104 so it

can compare the models of Erginsoy and Meyer and Bartoli with experimental results of paper [247]. The concentration of neutral-impurity centres as a function of temperature $N_N(T)$ in each sample is given by next relation

$$N_N(T) = N_{MJ} - N_{MN} - n(T). \quad (107)$$

Here N_{MJ}, N_{MN} and $n(T)$ are the majority-impurity, minority-impurity, and free-carrier concentrations, respectively. For the ionized-impurity scattering, Itoh *et al.* employ the Brooks-Herring expression [255, 256]:

$$\tau_{ion}^{-1} = \frac{\pi N_I e^4 (k_B T)^{-3/2} x^{-3/2}}{(2m_{con}^*)^{1/2} k^{1/2}} \left[\ln\left[1 + \frac{4x}{a}\right] - \frac{4x/a}{1 + 4x/a} \right], \quad (108)$$

$$n(T) = \frac{2(N_{MJ} - N_{MN})}{\{[1 + (N_{MN}/gN_B)\exp(E_{MJ}/k_B T)] + \sqrt{[1 + (N_{MN} gN_B)\exp(E_{MJ}/k_B T)]^2 + (4/gN_B)(N_{MJ} - N_{MN})\exp(E_{MJ} k_B T)}\}}, \quad (114)$$

where

$$a = \frac{2\pi \hbar^2 e^2 n}{m^* k k_B^2 T^2}, \quad (109)$$

and $x = E/k_B T$ (E is the incident electron energy), m_{con}^* is the average conductivity effective mass, and N_I is the ionized-impurity concentration. The temperature-dependent N_I in each sample is given by

$$N_I(T) = n(T) + 2N_{MN}. \quad (110)$$

For the acoustic-phonon deformation-potential scattering [257]:

$$\tau_{ac}^{-1} = B_{ac}(m_{con}^* T)^{3/2} x^{1/2}, \quad (111)$$

where the constant B_{ac} has well-established values for n- and p-type Ge as shown in Table XIV. Having found τ^{-1} of all three scattering mechanisms, Itoh *et al.* calculated an average $\langle\tau\rangle$ using the Maxwell-Boltzman integration:

$$\langle\tau\rangle = \frac{4}{3\sqrt{\pi}} \int_0^\infty \frac{x^{3/2}\exp(-x)}{\tau_{ac}^{-1} + \tau_{ion}^{-1} + \tau_{neutral}^{-1}} dx. \quad (112)$$

TABLE XIV Parameters used in the total-mobility calculations (after [247])

	Ga:As (n-type)	Ge:HGa (p-type)
k	16	16
m_{con}^*	0.12 m_0	0.28 m_0
[a] B_{ac}	1.08×10^{10} g$^{3/2}$K$^{-3/2}$	9.50×10^{8} g$^{3/2}$K$^{-3/2}$
B_B(theoretical)	12.5 meB [b]	11.2 meV[c]

[a]The values of B are determined experimentally using ultrapure n- and p-type Ge of $N_{MJ} \sim N_{MN} \sim 3 \times 10^{11}$ cm^{-3}.
[b]M. Altarelli, W. Y. Hsu and R. A. Sabatini, *J.* Phys. C 10, L605 (1977).
[c]A. Baldareshi and N. O. Lipari, in Proc. 13th Inern. Conf. Phys. Semicond. (F. G. Fumi, ed, North-Holland, Amsterdam, 1976) p. 595.

Finally the total mobility μ_{tot} was given by

$$\mu_{tot} = e\langle\tau\rangle / m_{con}^*. \quad (113)$$

All parameters required for the mobility calculations are well-known in Ge (see Table XIV). The only unknown material parameters at this point are sample-dependent N_{MJ}, N_{MN} and $n(T)$ in Equations 107 and 110. All three parameters as will shown below, can be determined precisely for each sample by performing variable-temperature Hall-effect measurements. Consequently all mobility calculations are performed without any adjusable or scaling parameters. The experimental curves are fitted with the following standard semiconductor statistics [258], which describes the temperature dependence of the free-carrier concentration in semiconductors doped by shallow majority impurities N_{MJ} and compensated by minority impurities N_{MN}:

where $g = 1/2$ ($g = 4$) is the spin degeneracy for a donor (acceptor), N_B is the effective conduction-(valence-) band density of states, and E_{MJ} are the experimentally determined ionization energies: 14 and 11.07 meV for As and Ga, respectively (Table XIV).

Fig. 53 shows the relative strength of the scattering from the ionized and the neutral impurities. There is only a relatively small temperature region in which the scattering from the neutral impurities dominates. This range extends to higher temperatures as the free-carrier concentration is increased. The calculated "transition temperatures" above which the ionized impurities are the main scattering centres (see also [259]) compare very well with experimental results of Itoh *et al.* [247] (see also Fig. 54).

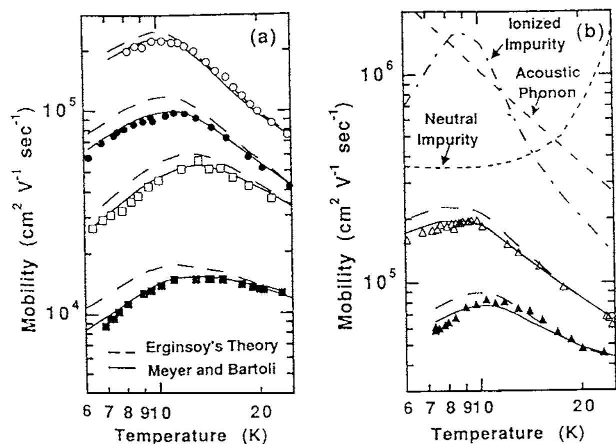


Figure 54 Data points represent experimentally measured carrier mobility in (a) four ^{74}Ge:As and (b) two ^{70}Ge:Ga samples. For a direct comparison theoretically calculated mobility using Erginsoy's model (broken line) and the model of Meyer and Bartoli (solid line) is shown for each sample. The contributions of the different scattering mechanisms to the total mobility of the ^{70}Ge:Ga—1 sample are shown in the upper half of (b) (after [247]).

We now turn the attention to the low-temperature regime where mobilities are dominated ny neutral-impurity scattering. Fig. 54 shows a direct comparison of the experimental results with theoretical total-mobility curves calculated. For each sample two theoretical total-mobility curves are calculated: one using Erginsoy's model [Equation 102] and other using Meyer and Bartoli model [Equation 104]. A strikingly good agreement was obtained between the experimental and theoretical mobilities calculated with the model of Meyer and Bartoli for all samples (see Fig. 54).

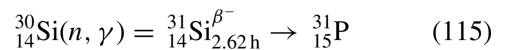
In order to demonstrate the importance of the homogeneous dopant distribution, Itoh *et al.* have performed the same study on samples cut from Ge:Ga crystals grown by the conventional Czochralski method, where Ga impurities were introduced to Ge melt during the crystal growth. These authors observed deviations of the measured mobility from the theoretical calculations, which are most likely due to inhomogeneous Ga impurity distributions in melt-doped Ge. Only the use of NTD semiconductors with randomly distributed dopants allows for an accurate test of the neutral impurity-scattering models.

2.8.2. Silicon

It is well-known that doping of silicon single crystals by incorporation of impurities from the melt during solidification in most cases leads to an inhomogeneous distribution of impurities in the solids [155, 236]. This is due to the fact that nearly all impurities in silicon have thermal equilibrium distribution coefficients much less than unity and that the solidification or crystal grows at each position of the interface is characterized by a different state of thermal inequilibrium leading to distribution coefficients that in space and time continuously change and result in a nonuniform impurity distribution [162, 260]. In actual crystal production the nonuniformity is further enhanced by lack of control of exactly constant melt volume and feed of the doping impurity. The most widely used doping elements in silicon are boron and phosphorus. Boron has a distribution coefficient between 0.9 and 1 which makes a doping uniformity of $\pm 10\%$ easily obtainable (see, e.g. [236]). The thermal equilibrium distribution coefficient for phosphorus of approximately 0.3 leads in general to the above mentioned large doping variations both on a macroscale (center to periphery) and on a microscale (striations). No other n-type doping element has a larger distribution coefficient. Because fast diffusing p-type dopants (Ga, Al) are available, because electron mobility is greater than hole mobility, and because contact alloying technology is reasonable, n-type silicon is generally used for solid state power devices [162, 261]. With avalanche breakdown voltages being determined from areas with lower resistivities, use of a conventionally doped material results in hot-spot formation prior to breakdown and too high forward voltage drop leading to excessive heat dissipation because of a safe punch through design [155, 236, 262].

Phosphorus doping by means of NTD was suggested by Lark-Horovitz [180] and Tanenbaum and Mills [154] for homogeneity purposes and has been applied for high-power thyristor manufacturing in [155, 236, 261]. Hill *et al.* [261] were demonstrated how such a homogeneous phosphorus doping may result in a "theoretical design" possibility for high-power components (see also below).

The process used for fractional transmutation of silicon into phosphorus and thereby performing n-type doping



$$^{30}_{14}\mathrm{Si}(n,\gamma) = {}^{31}_{14}\mathrm{Si}^{\beta^-}_{2.62\,\mathrm{h}} \rightarrow {}^{31}_{15}\mathrm{P} \qquad (115)$$

was first pointed out by Lark-Horovitz in 1951 [180]. Apart from special applications [263] and research, the above process was, however, not utilized to any extent until the early seventies, at which time manufacturers of high-power thyristors and rectifiers for high-voltage direct current transmission lines, in particular, initiated usage of the transmutation doping process [193, 261, 264]. The reasons for not using the neutron doping method throughout the sixties may be found in the lack of a processing technology which could benefit from a more uniform doping, insufficient availability of high resistivity starting material, and the lack of nuclear reactors with irradiation capacities in excess of that needed for testing fuel and materials for nuclear power stations.

Let us, for the following discussion, assume that completely uniform neutron doping may be accomplished. The homogeneity of the doped silicon is in this case determined by the background doping, i.e., the distribution of impurities in the starting material, where the net impurity concentration may be of either donor or acceptor type. Let us further, for simplicity, consider starting material of one conductivity type and assume complete n-type conduction after irradiation and annealing. With C_S being the net impurity concentration of the starting material and C_D the resulting donor concentration after irradiation we have, for both n- and p-type material,

$$C_\mathrm{D}^{\max} - C_\mathrm{D}^{\min} = C_\mathrm{S}^{\max} - C_\mathrm{S}^{\min}. \qquad (116)$$

In such case we may define

1. the homogeneity factors for the starting material (α_S) and for the neutron doped material (α_D), respectively

$$\alpha_\mathrm{S} = \frac{C_\mathrm{S}^{\min}}{C_\mathrm{S}^{\max}} \qquad (117)$$

and

$$\alpha_\mathrm{D} = \frac{C_\mathrm{D}^{\min}}{C_\mathrm{D}^{\max}} \qquad (118)$$

and

2. the doping factor

$$f_\mathrm{D} = \frac{C_\mathrm{D}^{\max}}{C_\mathrm{S}^{\max}}. \qquad (119)$$

TABLE XV Values for homogeneity factor α_D as function of homogeneity factor α_s of starting material and doping factor f_D as defined in text (after [193])

α_S/f_D	.1	.2	.3	.4	.5	.6	.7	.8	.9
1	.1	.2	.3	.4	.5	.6	.7	.8	.9
2	.55	.6	.65	.7	.75	.8	.85	.9	.95
5	.82	.84	.86	.88	.9	.92	.94	.96	.98
7	.87	.89	.90	.91	.93	.94	.96	.97	.99
10	.91	.92	.93	.94	.95	.96	.97	.98	.99
20	.955	.96	.965	.97	.975	.98	.985	.99	.995
50	.98	.98	.99	.99	.99	.99	.99	.996	.998
100	.991	.992	.993	.994	.995	.996	.997	.998	.999

From this is easily derived

$$1-\alpha_D=\frac{1-\alpha_S}{f_D}. \tag{120}$$

Table XV summarizes values of α_D as a function of α_S and f_D. It is seen that in order to obtain neutron-doped silicon with, for instance, a homogeneity factor greater than 0.9, it is necessary to use a doping factor of at least 7 when starting from "undoped" n-type material in which the homogeneity factor is typically not greater than 0.3 when taking the microcavitations (striations) into account. An examples of such neutron-doped silicon are shown in Figs 55 and 56. It should be noted that in terms of resistivity, which is often used for impurity characterization, a doping factor f_D means use of starting material with minimum resistivity a factor f_D or $2.8f_D$ greater than the resistivity after neutron doping for n- and p-type starting material, respectively. The difference is due to the electron mobility being 2.8 times greater than the hole mobility. In conclusion of this section it should be generally noted that in order to make neutron-doped silicon with significantly more uniform resistivity than conventionally doped material, a doping factor $f_D=5$ or more should be applied.

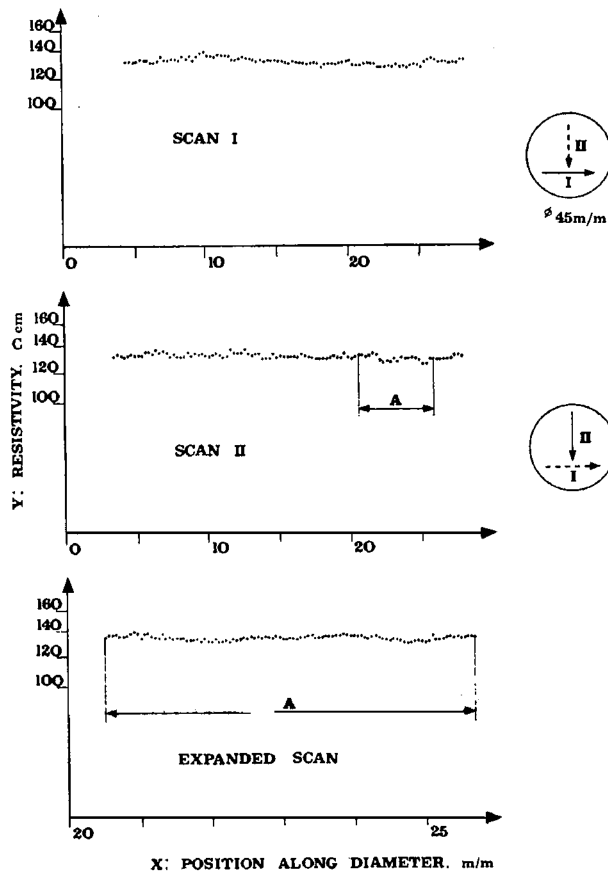


Figure 55 Spreading resistance measurements of a thermal neutron irradiation doped silicon slice. Step-length on scan 1 and 2 is 250 μm and on scan 3 step-length is 50 μm. Starting material has been selected greater than 1500Ω cm n-type (after [193]).

Following Janus and Malmros [193] let us consider further the theoretical case where a cylindrical silicon crystal is surrounded by a material with the same neutron absorption and scattering efficiency as the silicon itself (see Fig. 57). Let us furthermore assume a thermal neutron flux gradient along an x axis perpendicular to the crystal axis with the neutrons coming from an external source. In this case the neutron flux will have the form

$$\Phi=\Phi_0\cdot\exp\left(-\frac{x}{b}\right), \tag{121}$$

where b, the decay length, may be obtained from the formula

$$b=\left(3\cdot\sigma_{Si}\cdot\sigma_{Si,t}\cdot C_{Si}^2\right)^{-0.5}. \tag{122}$$

$\sigma_{Si}=0.16\cdot10^{-24}$ cm^2 is the mean of the absorption cross-sections for the 3 silicon isotopes, ^{28}Si, ^{29}Si and ^{30}Si weighted with their abundances. $\sigma_{Si,t}=2.3\cdot10^{-24}$ cm^2 is the total cross-section (absorption + scattering) and $C_{Si}=4.96\cdot10^{22}$ cm^{-3} is the total number of silicon atoms in 1 cm^3. Hence b may be calculated:

$$b_{\text{silicon}}=19\text{ cm}. \tag{123}$$

In order to improve the doping homogeneity in the cylindrical crystal this will be slowly rotated around its axis. The time average of this flux at the distance r from this axis is

$$\bar\Phi=\frac{1}{\pi}\int_0^{\pi}\Phi_0\exp\left[-\frac{r}{b}\text{cost}\right]dt = \Phi_0\left[1+\frac{1}{4}\left(\frac{r}{b}\right)^2+\cdots\cdots\right]. \tag{124}$$

The ratio between the neutron dose at the periphery and at the axis of the crystal cylinder will then be

$$\frac{\bar\Phi(a)}{\bar\Phi(0)}\simeq1+\frac{1}{4}\left(\frac{a}{b}\right)^2, \tag{125}$$

where a is the crystal radius (Fig. 57).

For intrinsic starting material the irradiation doped silicon will thus have a homogeneity factor of

$$\alpha_D\simeq1-\frac{1}{4}\left(\frac{a}{b}\right)^2\simeq0.956 \tag{126}$$

for an 80-mm-diameter crystal, i.e., the absorption limiting factor for the obtainable radial variations.

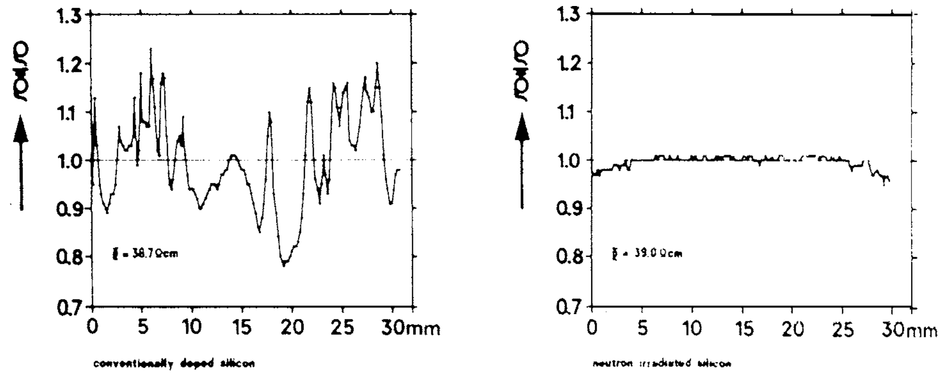


Figure 56 Typical lateral microscopic resistivity distributions in conventionally doped silicon and in silicon doped by neutron irradiation (after [236]).

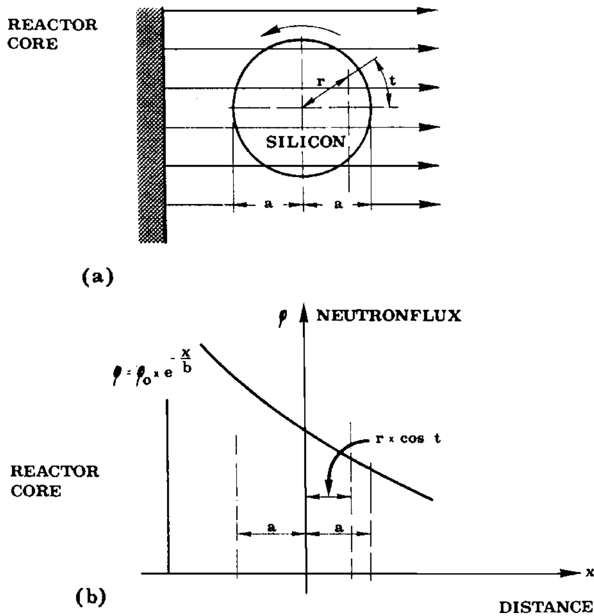


Figure 57 Irradiation configuration. (a) Top view of the facility with the cylindrical crystal situated outside the core rotating around its cylindrical axis. Arrows indicate overall direction of neutrons. The flux does not vary along the cylindrical axis. (b) The neutron flux is a function of the distance from the reactor core (after [193]).

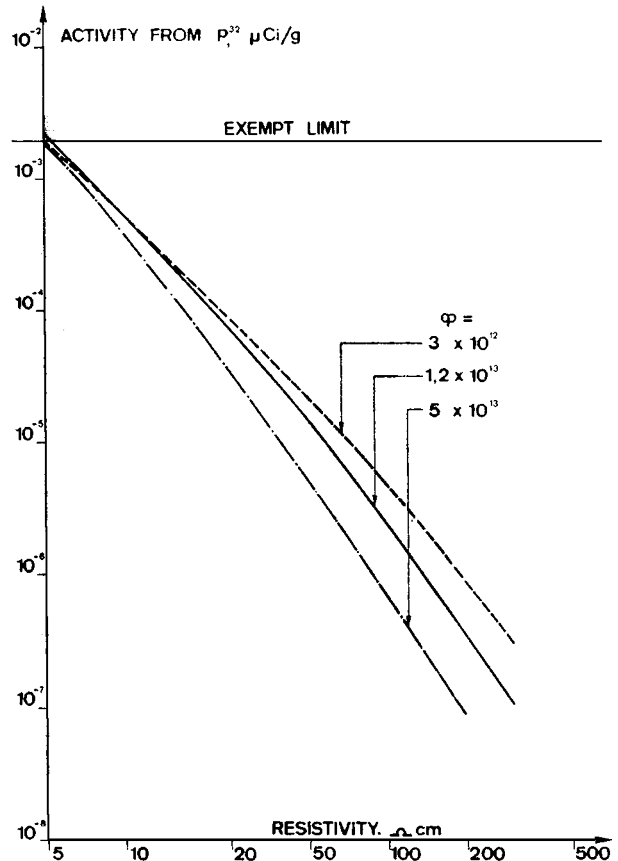
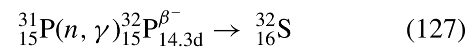


Figure 58 The radioactivity of the ^{32}P isotope in silicon after 4 days of cool down subsequent to irradiation. It may be observed that the activity as function of the resistivity obtained depends on the neutron flux used (after [193]).

In the above analysis we have neglected the effects of fast neutron moderation in the silicon. By comparison, however, of irradiations performed in reactors with fast neutron fluxes from 10^{-4} to 1 times the thermal flux and with different flux gradients, the authors [193] have observed no influence on the resistivity homogeneity due to fast neutron moderation in the silicon.

In irradiated silicon crystals for semiconductor device applications only two isotopes ^{31}Si and ^{32}P are of importance in connection with radioactivity of neutron doped material. For thermal neutron doses less than 10^{19} neutron/cm^2, no other elements have been detected emitting radiation. Futhermore, ^{31}Si, having a halflife of 2.62 h, decays to an undetectable level in 3–5 days. For this reason, it will be discussed the radioactivety only of the ^{32}P isotope. Fig. 58 pictures the ^{32}P activity as a function of final resistivity for a variety of thermal neutron flux levels typical for the nuclear test reactors in use. As was shown in [193] absolute flux determination to 1% accuracy has proven obtainable for instance by means of calorimetric boron carbide monitors.



$$^{31}_{15}\mathrm{P}(n,\gamma)^{32}_{15}\mathrm{P}\overset{\beta^-}{_{14.3\,\mathrm{d}}}\rightarrow\ ^{32}_{16}\mathrm{S} \qquad (127)$$

as a secondary one with ^{31}P concentration at each instant in time being dependent on the neutron dose received and the time allowed for the $^{31}_{14}\mathrm{Si}\overset{\beta^-}{_{2.62\,\mathrm{h}}}\rightarrow\ ^{32}_{15}\mathrm{P}$ decay.

From Fig. 58 it may be observed that neutron doping below 5 Ω · cm can be performed only when accepting cool down periods corresponding to the ^{32}P halflife of 14.3 days. The exempt limit for inactivity of 2×10^{-3} μCi/g shown on the figure is representative for most European countries, as well as being the value

recommended by the International Atomic Energy Agency (IAEA), Vienna, Austria [265]. It should be added that careful cleaning of the silicon prior to insertion in a nuclear reactor is vital to avoid radioactive surface contamination. For the safety of the personnel and the end product users, a double check upon shipping from the reactor sites and upon reception in the silicon plant, respectively is carried out to secure that only inactive material (below the exempt limit) is being further processed after the neutron doping. In general, this implies shipment from the reactor not earlier than 4 days after irradiation.

The use of NTD is of particular interest to thyristor manufacturers where n-type starting material is required for the basic p-n-p structure [259, 260]. Some advantages for high power device design and performance include:

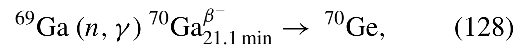
1. more precise control of avalanche breakdown voltage,
2. more uniform avalanche breakdown, i.e., greater capacity to withstand overvoltages,
3. more uniform current flow in forward direction, i.e., greater surge current capacity, and
4. narrower neutral zone and therefore narrower base and lower forward voltage drop V_f.

The summary of some points concerning the preparation of NTD silicon for special applications on an *R* and *D* scale describe in papers [260–261]. The production of large quantities of NTD silicon for power devices is described in [190]. More recently (see, e.g. [266]) the NTD technique has been also proposed for the effectual doping of P in a-Si:H films (see also [4]). The results of [266] are shown that NTD technique is an excellent method for doping of P in a-Si:H.

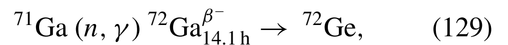
Despite intensive study over many years and considerable progress, no clear understanding has emerged of one of the fundamental issues regarding the MIT in doped semiconductors and amorphous metal-semiconductor mixtures: whether and under what circumstances the Hall coefficient diverges as the transitions is approached (see above for Ge). As is well-known in the localized regime the spatial behavior of the wave functions is usually described by an exponential decay length reflecting the spatial extent of the wave function (see e.g. [267–269]). Dai *et al.* recently are shown that the Hall coefficient of Si:P diverges at the transitions, as it does in Si:B [270] and Ge:Sb [271]. The difference in the behavior of MIT according these authors may be connected with a different degree of compensation. It is also possible that the MIT is different in a persistent photoconductor, where the disorder is particularly strong and the concentration of shallow donors is varied and controlled through illumination. (details see also [268, 269]).

2.8.3. Other compounds

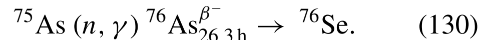
The NTD method have used with success in study of compound semiconductors: GaAs [184, 272–274] and GaP [275, 276]. NTD of GaAs is based on the following thermal neutron capture nuclear reactions (see also [184]):



$$^{69}\mathrm{Ga}\,(n,\gamma)\,^{70}\mathrm{Ga}^{\beta^-}_{21.1\,\mathrm{min}} \rightarrow\ ^{70}\mathrm{Ge}, \qquad (128)$$



$$^{71}\mathrm{Ga}\,(n,\gamma)\,^{72}\mathrm{Ga}^{\beta^-}_{14.1\,\mathrm{h}} \rightarrow\ ^{72}\mathrm{Ge}, \qquad (129)$$



$$^{75}\mathrm{As}\,(n,\gamma)\,^{76}\mathrm{As}^{\beta^-}_{26.3\,\mathrm{h}} \rightarrow\ ^{76}\mathrm{Se}. \qquad (130)$$

The relative abundances of the isotopes involved in the reactions and the cross-sections for these reactions are such that the ratio of Se and Ge concentrations produced is

$$N_{\mathrm{Se}}/N_{\mathrm{Ge}} = 1.46. \qquad (131)$$

Selenium is a typically shallow substitutional donor in GaAs with an electronic energy level a few meV from the conduction bans edge [277]. Germanium in GaAs is an amphoteric impurity which acts as a shallow donor (also a few meV from the conduction band) is situated on a Ga site and as an acceptor level at $E_V + 0.04$ eV if situated on an As site [278]. Since, if electronically active, all of the Se atoms and some portion of the Ge atoms are expected to act as donors, NTD of GaAs is expected to dope GaAs more n-type. The addition of donors moves the Fermi level (E_F) away from the valence band (E_V) to the conduction band (E_C). If a sufficiently high concentration of donors is added, E_F will move to the upper half of the bandgap and the GaAs will be converted to n-type. Analysis of Hall effect data as a function of temperature provides a means of measuring the donor content in irradiated GaAs samples. Young *et al.* were thus able to compare electrically active added donor content to the NTD-produced impurity concentrations determined from nuclear measurements. The Hall effect analysis also allows them to determine concentrations and energy levels (E) of impurities or defects in the p-type GaAs samples if the Fermi level in the material moves near E at some temperature over the range of measurements. This technique thus provides a means of identifying and measuring undercompensated acceptor content in the samples. The low temperature photoluminescence technique used in paper [184] measured donor-to-acceptor or conduction-band-acceptor luminescence. It provides an accurate determination of the position of acceptor electronic levels in the GaAs, permitting positive identification of impurities or defects with known luminescence lines. Identifications of lines due to specific impurities or defects can be made using luminescence techniques regardless of the position of the Fermi level in material. Little detailed information concerning an acceptor level can be obtained from Hall effect if that acceptor is overcompensated. However, the presence of specific acceptors can be detected by luminescence techniques even in n-type samples. On other hand, luminescence data do not provide the quantitative information obtainable from Hall effect measurements.

The results of room temperature measurement of the electrical properties of eight annealed NTD GaAs samples are summarized in Table XVI. The total NTD dose ($N_{\mathrm{Se}} + N_{\mathrm{Ge}}$), the carrier concentration and carrier

TABLE XVI Room temperature results for Hall effect samples of GaAs annealed 830°C/20 min (after [184])

Sample no	NTD dose/cm^3	n (−) or P (+) in cm^{-3}	μ, cm^2 V^{-1} s$_{-1}$
1, 2	3.8×10^{15}	$+2.3 \times 10^{16}$	360
3	8.5×10^{15}	$+2.4 \times 10^{16}$	341
4	1.7×10^{16}	$+1.9 \times 10^{16}$	337
10	2.7×10^{16}	$+8.6 \times 10^{15}$	242
12	7×10^{16}	-1.6×10^{16}	1251
20	1.5×10^{17}	-7.7×10^{16}	3960
15	2.8×10^{17}	-2.3×10^{17}	3631
16, 18	6.3×10^{17}	-4.9×10^{17}	3110

type (negative values of concentration indicate n-type), along with carrier mobility at room temperature are indicated in Table XXII. Note that following an NTD dose sufficient to produce $7 \cdot 10^{16}$ atoms/cm^3 initially present in the samples. Therefore, 7×10^{16} donors/cm^3 would indeed be expected to just overcompensate the p-type material. The results presented in Table XVI show that the p-type samples become progressively less p-type and the n-type samples progressively more n-type with increasing NTD dose. Because the donor levels in GaAs are very shallow, they remain fully ionized in the temperature range of Young *et al.* experiments, so that the measured electron concentration is practically temperature independent (see Fig. 4 in [184]). This measured n for each sample is approximately equal to total donor minus total acceptor concentration.

Fig. 59 shows the measured in [184] added electrically active donor concentration in eight NTD samples as a function of N_{Se} and of $(N_{\mathrm{Se}} + N_{\mathrm{Ge}})$ added by transmutation as determined from nuclear activity measurements. The uncertainty in determining added donor content in the p-type samples is large because of the complexity of analyzing material with multiple independent acceptor levels in closely compensated cases. The added donors can be much more accurately determined in the more highly doped n-type samples. The results shown in Fig. 59 imply that all of the selenium

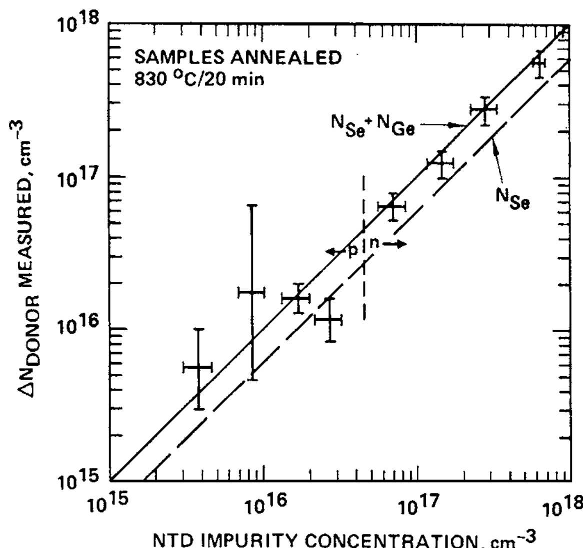


Figure 59 Measured added donors vs. NTD produced impurity content (after [184]).

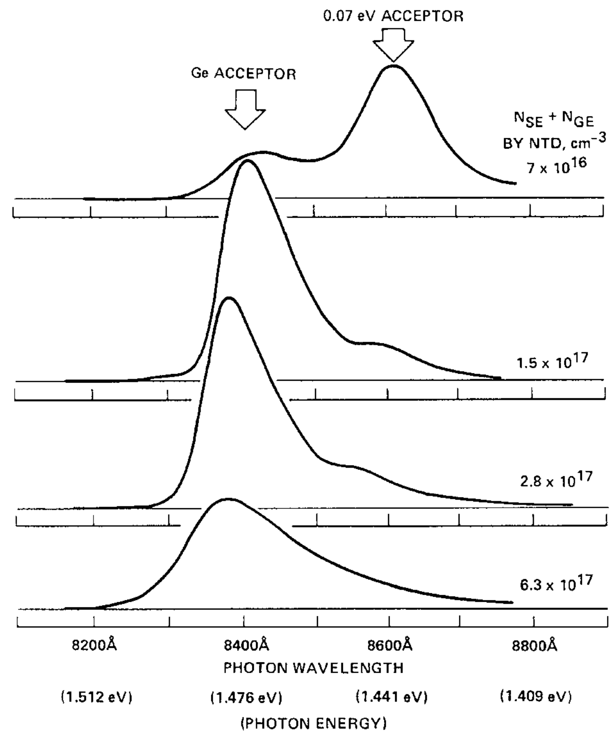


Figure 60 Relative photoluminescence spectra for four n-type NTD samples. The four spectra are not normalized with respect to each other (after [184]).

and a substantial fraction of the Ge atoms introduced by transmutation act as donors following the 830°C/20 min anneal. As will be shown below from photoluminescence measurements a fraction of Ge atoms produced by transmutation are on acceptor rather than donor sites in GaAs samples.

Fig. 60 show relative luminescence spectra for the four n-type samples respectively. The spectral positions indicated by arrows for carbon acceptor, the Ge acceptor, and 0.07 eV acceptor correspond to donor (or band) to acceptor luminescence lines. The most important conclusion to be drawn from a comparison of the spectra for the control and eight NTD samples is that Ge acceptors not present in the "starting material" control sample are introduce by the NTD process. The increase in intensity of the Ge acceptor line with increasing dose relative to both the carbon and 0.07 eV acceptor lines indicates that Ge acceptor content increases with increasing transmutation doping. Therefore, some of the Ge atoms produced by NTD in these samples are acting as acceptors rather than donors. Photoluminescence measurement studies of the control and eight annealed NTD samples at longer wavelengths indicate another new line present only in NTD samples at about 9450 Å. The intensity of this line increases with increasing NTD dose.

The characteristic lifetimes of radioactive isotopes can be used to label and identify defect levels in semiconductors which can be detected by photoluminescence [273] and Raman-scattering spectroscopy [274]. Magerle *et al.* [273] show photoluminescence spectra of GaAs doped with ^{111}In that decays to ^{111}Cd. ^{111}In is isoelectronic to Ga and hence occupies Ga lattice sites in GaAs. It decays to ^{111}Cd with a lifetime $\tau_{111_{\mathrm{In}}} = 98$ h by electron capture [279]. Since the recoil energy of the

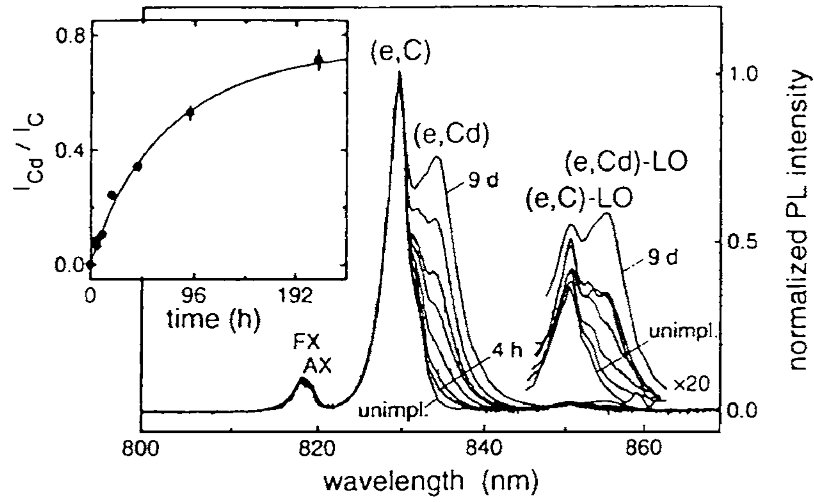


Figure 61 Photoluminescence spectra of undoped and ^{111}In doped GaAs successively taken 4 h; 7 h; 12 h; 22 h; 2 d; 4 d and 9 d after doping. All spectra are normalized to the intensity of the (e, C) peak. In the inset, the height I_{Cd}/I_C of the (e, Cd) peak in these spectra is shown as a function of time after doping with ^{111}In. The solid line is a fit to the data using Equation 139 (after [273]).

Cd nucleus due to the emission of the neutrino is much smaller than the typical displacement energy in GaAs [280], ^{111}Cd atoms on Ga sites (Cd_{Ga}) are created by the decay of ^{111}In on Ga sites ($^{111}In_{Ga}$) and act there as shallow acceptors. This chemical transmutation was monitored by photoluminescence spectroscopy. Fig. 61 shows successively taken photolumunescence spectra from the ^{111}In doped sample. A spectrum from the undoped part is also shown. The photoluminescence spectrum of the undoped part of the sample shows the features well known for undoped MBE-grown GaAs [281]. The peaks FX and AX around 819 nm are due to the recombination of free and bound excitons. The peak (e,C) at 830 nm and its LO phonon replica (e,C)-LO at 850 nm are due to recombination of electrons from the conduction band into C acceptor states. The recombination of electrons from donor states into C acceptor states appears as a small shoulder at the right-hand sides of either of these two peaks. C is a residual impurity in GaAs present in MBE-grown material with a typical concentration between 10^{14} and 10^{15} cm^{-3} [281]. Magerle *et al.* determined the height I_{Cd}/I_C of the (e, Cd) peak normalized to I_C as the function of time after doping. These was done by substracting the normalized spectrum of the undoped part from the normalized spectra of the ^{111}In doped part. The height I_{Cd}/I_C of the (e,Cd) peak remaining in these difference spectra is displayed in the insert of Fig. 61. Indicated authors fitted these data by

$$\frac{I_{Cd}}{I_C}(t) = \frac{I_{Cd}}{I_C}(t = \infty)\left(1 - e^{-\frac{t}{\tau}}\right) \qquad (132)$$

and obtained a time constant $\tau = 52(17)$ h, which is not the nuclear lifetime $\tau_{111_{In}} = 98$ h of ^{111}In. Evidently I_{Cd}/I_C is not proportional to N_{Cd}. The photoluminescence intensity I_{Cd} is proportianal to the recombination rate of excess carriers per unit area through Cd acceptors states $\Delta n_L B_{Cd} N_{Cd}$, where B_{Cd} is a recombination coefficient. The excess sheet carrier concentration in the implanted layer Δn_L can be expressed in terms of the total carrier lifetime in the implanted layer τ_L and the generation rate of excess carriers per unit area in the implanted layer $f_L G$ by using the first of the two equilibrium conditions

$$f_L G = \frac{\Delta n_L}{\tau_L} \quad \text{and} \quad f_B G = \frac{\Delta n_B}{\tau_B} \qquad (133)$$

The second one describes the balabce between the generation rate $f_B G$ and the recombination rate of excess carriers $\frac{\Delta n_B}{\tau_B}$ in the bulk. The total generation rate G is proportional to the incident photon flux and $f_L + f_B = 1$. To get an expression for τ_L, cited authors assumed two additional recombination processes in the implanted layer: the radiative recombination via Cd acceptors and nonradiative recombination due to residual implantation damage, and write the recombination rate in the small single approximation (see, e.g. [282]) as

$$\frac{\Delta n_L}{\tau_L} = \frac{\Delta n_L}{\tau_B} + \Delta n_L B_{Cd} N_{Cd} + \Delta n_L B_{nr} f_{nr} N_{Cd} \qquad (134)$$

Here $\Delta n_L B_{nr} f_{nr} N_{Cd}$ is the nonradiative recombination rate per unit area due to residual implantation damage, $f_{nr} N_{Cd}$ is the concentration of these nonradiative recombination centers, and B_{nr} is the corresponding recombination coefficient. Hence Δn_L and Δn_B can be expressed as a function of N_{Cd} and the recombination rates through all the different recombination channels and thereby the relative photoluminescence peak intensities can be deduced. I_C is proportional to the sum of the (e,C) recombination rates per unit area in the implanted layer and the bulk and within this model it can obtain

$$I_C \propto \frac{\Delta n_L + \Delta n_B}{\tau_C} = G\frac{\tau_B}{\tau_C}\left(\frac{f_L}{1 + \Phi_{Cd}/f_B b} + f_B\right). \qquad (135)$$

Here Φ_{Cd} is the dose between 10^9 and 10^{13} cm^2. Thereby $\tau_C = 1/B_C N_C$ is an effective lifetime describing the recombination probability through C acceptor states and b is a constant defined below. With help of Equations 133 and 135 it can be obtain (assuming that

the detection efficiencies of both peaks are equal) the following relation between I_{Cd}/I_C and Φ_{Cd}:

$$\frac{I_{Cd}}{I_C} = \frac{\Delta n_L B_{Cd} N_{Cd}}{(\Delta n_L + \Delta n_B)/\tau_C} = \frac{a}{1 + b/\Phi_{Cd}}, \quad (136)$$

with

$$a = \frac{f_L}{f_B} \frac{B_{Cd}}{(B_{nr} f_{nr} + B_{Cd})} \frac{\tau_C}{\tau_B} \quad \text{and}$$
$$b = \frac{d}{f_B (B_{nr} f_{nr} + B_{Cd}) \tau_B}. \quad (137)$$

This model describes quantitatively the dependence of (e, Cd) intensity of N_{Cd} and cited authors use it to describe the increase of I_{Cd}/I_C with time in the ^{111}In-doped sample. Authors [273] model the change of the carrier lifetime τ_L with time t in the ^{111}In doped sample as

$$\frac{1}{\tau_L} = \frac{1}{\tau_B} + B_{Cd} N_{In} \left(1 - e^{-\frac{t}{\tau}}\right) B_{nr} f_{nr}, \quad (138)$$

where $N_{In} = \Phi_{In}/d$ is the initial ^{111}In concentration, $\tau = \tau_{111_{In}} = 98.0$ h is the nuclear lifetime of ^{111}In, and B_{Cd}, B_{nr} and f_{nr} are the same constants as above. Thereby we assume following Magerle *et al*. that the same kinds of nonradiative recombination centers are produced by In doping as by Cd doping and that the Cd concentration are identical to the ^{111}In concentration profile. Taking into account all above saying we can write

$$\frac{I_{Cd}}{I_C} = \frac{a}{1 + b/\Phi_{In}\left(1 - e^{-\frac{t}{\tau}}\right) + c/\left(e^{-\frac{t}{\tau}} - 1\right)}, \quad (139)$$

where a and b are the same constants as above and $c = B_{nr} f_{nr}/(B_{nr} f_{nr} + B_{Cd})$. This c term accounts for the fact that the ^{111}In doped sample the concentration of nonradiative centers is not changing with Cd concentration. Magerle *et al*. fitted Equation 139 to the data shown in inset of Fig. 61, keeping $\tau = 98.0$ h, $a = 1.25$ and $b = 3.0 \times 10^{11}$ cm^{-2}, and obtained $\Phi_{In} = 4.49 \times 10^{11}$ cm^{-2} and $c = 0.5$ (2). This fit is shown as a solid line and agrees perfectly with the experimental data. In conclusion of this part it should note that this identification technique is applicable to a large variety of defect levels since for most elements suitable radioactive isotope exist (details see [279]).

Coupling between the LO phonon mode and the longitudinal plasma mode in NTD semi-insulating GaAs was studied in paper [274] using Raman-scattering spectroscopy and a Fourier-transform infrared spectrometer. Raman spectra are shown in Fig. 62 for unirradiated, as-irradiated and annealed samples. The remarkable feature is the low intensity and asymmetric linewidth of the Lo-phonon spectrum observed in annealed samples, which are annealed above 600°C. The behavior is not understood by considering the only LO phonon. We should pay attention to the electrical activation of NTD impurities, which begin to activate electrically around 600°C. In the long-wavelength limit, the valence electrons, the polar lattice vibrations, and the

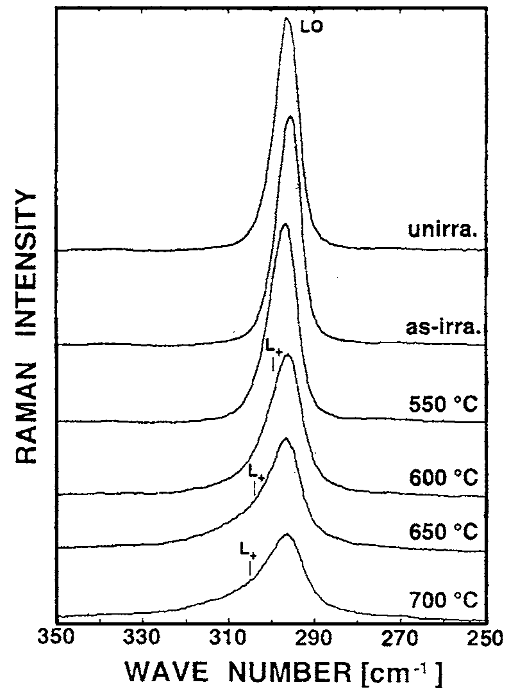


Figure 62 Raman spectra at room temperature taken for the various annealing temperatures of (100)—oriented NTD GaAs irradiated with neutron dozes (athermal neutron of 1.5×10^{18} cm^{-2} and a fast neutron of 7×10^{17} cm^{-2}). The coupling L_+ mode is observed at annealing temperatures above 600°C (see Table XVII) (after [274]).

conduction electrons make additive contributions to the total dielectric response function [283]:

$$\varepsilon_T(0, \omega) = \varepsilon_\infty + (\varepsilon_0 - \varepsilon_\infty)/\left[\left(1 - \omega^2/\omega_t^2\right) - \omega_p^2 \varepsilon_\infty/\omega^2\right]. \quad (140)$$

The high-frequency value (L_+) of the mixed LO-phonon-plasmon modes is calculated from the roots of the dielectric constant of Equation 132. The frequencies of the L_+ mod e and of the longitudinal plasma mode $\omega_p = (4\pi n e^2/\varepsilon_\infty m^*)^{1/2}$ for various annealing temperatures are listed in Table XVII. Here n is the electron concentration, m^* the effective mass in the conduction band ($= 0.07 m_0$), and ε_∞ ($= 11.3$) the optical dielectric constant. The mixed LO-phonon-plasma mode appears around 300 cm^{-1} for electron concentration of (0.8–2) $\times 10^{17}$ cm^{-3}. The phonon strength [283]

TABLE XVII Electron concentrations and the coupling modes of NTD GaAs (after [274])

Sample	EC (cm^{-3})	LO-phonon frequency (cm^{-1})	L_+ mode (cm^{-1})	PF (cm^{-1})
Unirradiated	$1 \sim 2 \times 10^7$	296.6		
As-irradiated	[a]	295.6		
500°C annealed	[a]	297.8		
600°C annealed	8.2×10^{16}	296.0	299	96.4
650°C annealed	2.2×10^{17}	296.6	304	158
700°C annealed	2.5×10^{17}	296.2	305	168

[a]Since the conduction is dominated by Mott-type hopping conduction (M. Satoh and K. Kuriyama, Phys. Rev. B40, 3473 (1989), the electron concentration can not be measured by the van der Pauw method.
EC = Electron concentration; PF = Plasma frequency.

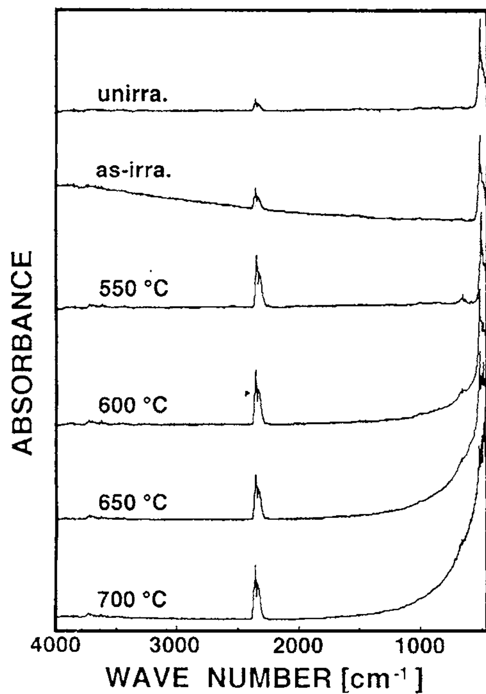


Figure 63 Infrared-absorption spectra at room temperature taken for the various annealing temperatures of the NTD GaAs used for the Raman-scattering esperiments (after [274]).

for the high-frequency mode (L_+) of the interacting plasmon-LO-phonon mode is about 0.95 for an electron concentration of 1×10^{17} cm^{-3}, while that for the low-frequency mode (L_-) is below 0.1. Therefore, the asymmetric linewidth of the Raman spectrum observed in the annealed NTD GaAs arises from both the LO-phonon and L_+ modes, but the L_- mode is not observed because of a very weak phonon strength. As a result, the LO-phonon intensity decreases with increasing coupling, and L_+ mode appears beside the LO-phonon peak.

The absorption spectra in the various annealing temperatures for NTD GaAs are shown in Fig. 63. In unirradiated samples, an absorption around 2350 cm^{-1} is assigned as the antisymmetric stretching vibration of CO_2 arising from CO_2 in an ambient atmosphere.The absorption peaks observed around 500 cm^{-1} are also assigned as a two-phonon overtone scattering [284] of transverse optical phonons (TO); these were observed at 493 cm^{-1} [2TO(X)], 508 cm^{-1} [2TO(L)], and 524 cm^{-1} [2TO(Γ)], respectively. In as-irradiated samples, a continuous absorption extending to the higher energy was observed. Although this origin cannot be attributed to interstitial anion clusters as discussed in neutron irradiated GaP [285]. In samples annealed above 600°C, the remarkable absorption was observed at wave numbers below 1450 cm^{-1}. The absorption increases with increasing annealing temperature (see Fig. 109). This behavior arises from the fre-electron absorption due to the activation of NTD impurities, which occur at annealing temperatures above 600°C. The free-electron absorption observed is consistent with a collective motion as a plasmon mode described in Raman-scattering studies.

Kuriyama *et al.* [276] were studied by a photoluminescence method the transmuted impurities Ge an S in NTD semi-insulating Gap. In NTD GaP, Ge and S impurities are transmuted from Ga and P atoms by (n, γ) reactions, respectively. Ge in GaP is an amphoteric impurity for which both the donor and acceptor states appear to be deep. The ratio between transmuted impurities Ge and S is about 16:1. Unfortunately, after the transmutation reactions, the transmuted atoms are usually not in their original positions but displaced into interstitial positions due to the recoil produced by the γ and β particles in the nuclear reactions. In addition, the defects induced by the fast neutron irradiation disturb the electrical activation of transmuted impurities. However, Frenkel type defects [275, 286] in NTD GaP were annealed out between 200 and 300°C, while P antisite (P_{Ga}) defects of $\sim 10^{18}$ cm^{-3} annihilated at annealing temperatures between 600 and 650°C. Therefore, transmuted impurities, Ge and S, would be substituted on Ga and/or P lattice sites by annealing at around 650°C.

Fig. 64 shows the photoluminescence (PL) spectra of unirradiated and NTD GaP. The PL spectrum (peak 1) of unirradiated samples shows signature of the DA pair recombination involving S donor and carbon acceptor [288]. Two (peaks 2 and 3) of the replicas occur at energies consistent with electronic transitions accompanied by zone-center optical phonons with energies 50.1 meV (LO_Γ) and 100.2 ($2LO_\Gamma$). Sulfur, silicon and carbon in GaP are the most common as the residual impurities [288]. In NTD-GaP the main transition energy was observed 1.65 eV. Since Ge in GaP is the amphoteric impurity with deep acceptor and donor levels, strong phonon co-operation will also occur. But optical transition rates will be significant only for associates. Similar situation has been proposed for Si in GaP [287], forming a nearest-neighbor Si_{Ga}-Si_P complex. Therefore, the broad emission would be expected to arise from

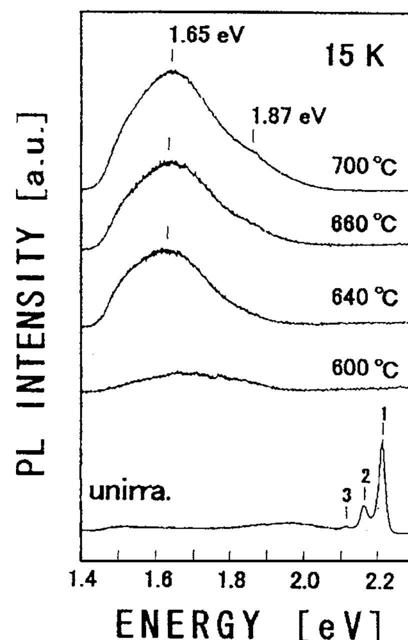


Figure 64 Photoluminescence (PL) spectra taken at 15 K for unirradiated and NTD-GaP. PL peaks 1, 2 and 3 in unirradiated GaP represent $S_P - C_P$ DA pair recombination, its LO—phonon replica, and 2LO—phonon replica, respectively. 1.65 and 1.87 eV emissions in NTD-GaP are attributed to $Ge_{Ga} - Ge_P$ complex and $S_P - Ge_P$ DA pair recombinations, respectively (after [276]).

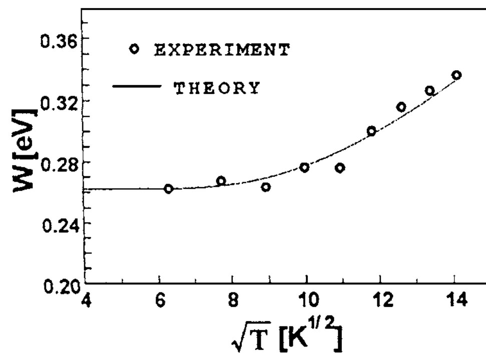


Figure 65 Variation of the half-width W with the square root of the temperature T for the 1.65 eV band in NTD-GaP. Theoretical curve is a plot of Equation 141 with $h\nu = 0.025$ eV (after [276]).

a nearest-neighbor Ge_{Ga}-Ge_P coupled strongly to the lattice. To confirm the presence of the Ge_{Ga}-Ge_P compose, the temperature dependence of the half-width, W, of the broad emission was measured. If the localized electron transitions from the excited state to the ground state of this complex center produce the characteristics luminescence, the dependence would be appear to follow the configuration-coordinate (CC) [289] model equation:

$$W = A[\coth(h\nu/2kT)]^{1/2}, \quad (141)$$

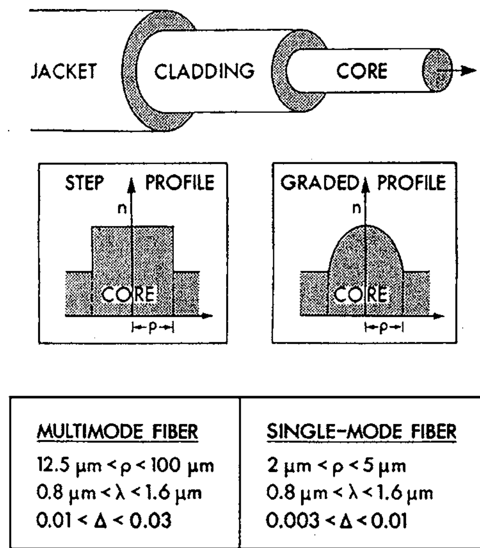
where A is a constant whose value is equal to W as the temperature approaches 0 K and $h\nu$ is the energy of the vibrational mode of the excited state. In Fig. 65, Equation 141 has been fitted to the experimental value for NTD-GaP. For the estimation of W, the spectrum of the 1.65 eV band was substracted from that of the 1.87 eV band. The value of $h\nu$ used was 0.025 eV. The good fit to this equation that was found for the Ge_{Ga}-Ge_P center in NTD-GaP shows the validity of applying the CC model. Results of paper [276] indicate that NTD method is a useful one for introducing Ge donor, resulting from a fact the Ge atoms are transmuted from Ga lattice sites in GaP. The obtained results are consistent with the presence of the Ge_{Ga}-Ge_P complex as described earlier.

Chapter 3. Optical fiber

3.1. Introduction

Optical communication using fibers is a major new technology which will profoundly impact telephone systems, computer interconnections and instrumentation (internet). Fiber links provide several major advantages over conventional electronic communications systems. These include immunity to electromagnetic interference, thinner and lighter cables, lower transmission losses (especially for very data rates) and potential kilometer-long link capabilities extending to the gigahertz region.

An optical waveguide is a dielectric structure that transports energy at wavelengths in the infrared or visible ranges [290, 291] of the electromagnetic spectrum. In practice, waveguides used for optical communications are highly flexible fibers composed of nearly transparent dielectric materials. The cross-section of these fibers is small—comparable to the thick of a human hair- and generally is divisible into three layers as shown in Fig. 66. The central region is the core, which is surrounded by the cladding, which in turn is surrounded by a protective jacket. Within the core, the refractive-index profile n can be uniform or graded, while the cladding index is typically uniform [293]. The two situation correspond to the step-index and graded-index profiles shown in the insets in Fig. 66. It is necessary that the core index be greater than the cladding index [294], at least in some region of the cross-section, if guidance is to take place. For the majority of applications, most of the light energy propagates in the core and only a small fraction travels in the cladding. The jacket is almost optically isolated from the core, so for this reason we usually ignore its effect and assume an unbounded cladding for simplicity in the analysis.



MULTIMODE FIBER	SINGLE-MODE FIBER
12.5 μm < ρ < 100 μm	2 μm < ρ < 5 μm
0.8 μm < λ < 1.6 μm	0.8 μm < λ < 1.6 μm
0.01 < Δ < 0.03	0.003 < Δ < 0.01

Figure 66 Nomenclature, profiles and ranges of dimensions for typical optical fibers, where ρ is the core radius, λ is the free-space wavelength of light and $\Delta = (1 - n_{cl}^2/n_{co}^2)/2$ (after [292]).

As usually, optical waveguides can be conveniently divided into two subclasses called multimode waveguides (with comparatively large cores) and single-mode waveguides (with comparatively small cores). The demarcation between the two is below. Multimode waveguides obey the condition (see e.g. [295]) $(2\pi\rho/\lambda)(n_{co}^2 - n_{cl}^2)^{1/2} \gg 1$, where ρ is a linear dimension in the core, e.g., the radius of the fiber core, λ is the wavelength of light in free space, n_{co} is the maximum refractive index in the core and n_{cl} is the uniform refractive index in the cladding.

As will be shown below electromagnetic propagation along optical waveguides is described exactly by Maxwell's equations. However, it is well-known that classical geometric optics provides an approximate description of light propagation in regions where the refractive index varies only slightly over a distance comparable to the wavelength of light. This is typical of multimode optical waveguides used for communication. Thus, the most direct and conceptually simple way to describe light propagation in multimode waveguides is by tracing rays along the core (see also [296, 297]). By

using classical geometric optics, we should ignored all wave effects. In multimode waveguides, wave effects are usually negligible [292], but there are exceptional situations when such effects accumulate exponentially with the distance light travels. Naturally in this cases, wave effects must be retained, since they can have a significant influence in long waveguides. In each such situation, we modify the classical geometric optics description by taking into account the local plane wave nature of light. The phenomenon of greatest practical interest in fibers used for long-distance communications is the spread of pulses as they propagate along the fiber. For idealized multimode fibers, pulse spreading is easily described by classical geometric optics. But since propagation in multimode guides is so complex that simple models and physical understanding are generally of much greater assistance than a precise, exact analysis [298].

3.2. Maxwell's equations

The spatial dependence of the electric field $\vec{E}(x, y, z)$ and the magnetic field $\vec{H}(x, y, z)$ of an optical waveguide is determined by Maxwell's equations. Further we assume an implicit time dependence $\exp(-i\omega t)$ in the field vectors, current density $\vec{J}$ and charge density σ. The dielectric constant $\varepsilon(x, y, z)$ is related to the refractive index $n(x, y, z)$ by $\varepsilon = n^2\varepsilon_0$, where ε_0 is the dielectric constant of free space. For the nonmagnetic materials which normally constitute an optical waveguide, the magnetic permeability μ is very nearly equal to the free-space value μ_0. Under these conditions, Maxwell's equations are expressible in the form (see, e.g. [299])

$$\vec{\nabla} \times \vec{E} = \frac{-\partial \vec{B}}{\partial t} = -\mu\mu_0 \frac{\partial \vec{H}}{\partial t} \qquad (142)$$

$$\vec{\nabla} \times \vec{H} = \frac{\partial \vec{D}}{\partial t} = \varepsilon\varepsilon_0 \frac{\partial \vec{E}}{\partial t} \qquad (143)$$

$$\vec{D} = \varepsilon\varepsilon_0 \vec{E} \qquad (144)$$

$$\vec{B} = \mu\mu_0 \vec{H} \qquad (145)$$

$$\vec{\nabla} \cdot \vec{D} = 0 \text{ (no free charges)} \qquad (146)$$

$$\vec{\nabla} \cdot \vec{B} = 0 \text{ (no free poles).} \qquad (147)$$

We will assume further that our guiding structure along the z direction. Thus we look for solutions to problems in which the z dependence of the field is of the form

$$\vec{E} = \vec{E}_0(x, y) \exp i(\omega t - {}^{z}k \cdot z) \qquad (148)$$

$$\vec{H} = \vec{H}_0(x, y) \exp i(\omega t - {}^{z}k \cdot z) \qquad (149)$$

In this expression we note that ω is related to ν, the frequency, by the relation $\omega = 2\pi\nu$. Furthermore, in writing the time and space variation of the field in terms of the complex exponential, it is understood, but not written, throughout this chapter that when a field is to be evaluated (e.g., for measurement) the only physical meanengful part of this complex expression is the real part. Thus, if we denote a real measurable field by ϵ and we wish to relate this to the theoretically derived field $\vec{E}$, which is given by an expression of the form of Equation 148, then

$$\epsilon = \mathrm{Re}[\vec{E}]. \qquad (150)$$

The advantage of this approach is that expressions involving the exponential of a complex quantity are more readily manipulated than the equivalent expressions involving sine and cosine. In next we will derive expressions for the field components in planar and cylindrical geometries before studying the solutions for guided waves in planar and cylindrical waveguides (see also [298]).

3.2.1. Planar geometry

The expression $\vec{\nabla} x$ in rectangular Cartesian coordinates is

$$\vec{\nabla} \times \vec{A} = \begin{vmatrix} \vec{i}\ \vec{j}\ \vec{k} \\ \frac{\partial}{\partial x}\frac{\partial}{\partial y}\frac{\partial}{\partial z} \\ {}^{x}\mathrm{A}{}^{y}\mathrm{A}{}^{z}\mathrm{A} \end{vmatrix}, \qquad (151)$$

where $\vec{i}$, $\vec{j}$, $\vec{k}$ are unit vectors in the x, y, z directions. Thus from Equation 142 we obtain for the x directed component of $\vec{\nabla} x \vec{E}$

$$\left(\frac{\partial^{z} E}{\partial y} - \frac{\partial^{y} E}{\partial z}\right) = -\mu\mu_0 \frac{\partial^{x} H_{x}}{\partial t}. \qquad (152)$$

Now we can write similar equations for $\partial^{y}H/\partial t; \partial^{z}H/\partial t; \partial^{x}E/\partial t; \partial^{y}E/\partial t$ and $\partial^{z}E/\partial t$. If then substitute the field expressions of Equations 148 and 149 into these expressions, we obtain the following results:

$$\frac{\partial^{z} E}{\partial y} + i^{z}k^{y}E = -i\mu\mu_0\omega^{x}H, \qquad (153)$$

$$i^{z}k^{x}E + \frac{\partial^{z} E}{\partial x} = i\mu\mu_0\omega^{y}H, \qquad (154)$$

$$\frac{\partial^{y} E}{\partial x} - \frac{\partial^{x} E}{\partial y} = -i\mu\mu_0\omega^{z}H, \qquad (155)$$

$$\frac{\partial^{z} H}{\partial y} + i^{z}k^{y}H = -i\varepsilon\varepsilon_0\omega^{x}E, \qquad (156)$$

$$-i^{z}k^{x}H - \frac{\partial^{z} H}{\partial x} = i\varepsilon\varepsilon_0\omega^{y}E, \qquad (157)$$

$$\frac{\partial^{y} H}{\partial x} - \frac{\partial^{x} H}{\partial y} = i\varepsilon\varepsilon_0\omega^{z}E. \qquad (158)$$

Analysis of planar structures will be restricted to infinite films that lie in the y–z plane. Thus, in addition to the assumption that the fields have the z dependence already postulated, it can further assume that the partial derivative with respect to y vanishes (hereafter $\partial/\partial y = 0$) for an infinite plane wave traveling in the z direction. With this assumption the above equations simplify and

demonstrate a fundamental relationship about the fields in such a structure. Indicated relations have forms:

$$i^{z}k^{y}E = -i\mu\mu_0\omega^{x}H \text{ (TE group)}, \quad (159)$$

$$i^{z}k^{x}E + \frac{\partial^{z}E}{\partial x} = i\mu\mu_0\omega^{y}H \text{ (TM group)}, \quad (160)$$

$$\frac{\partial^{y}E}{\partial x} = -i\mu\mu_0\omega^{z}H \text{ (TE group)}, \quad (161)$$

$$i^{z}k^{y}H = i\varepsilon\varepsilon_0\omega^{x}E \text{ (TM group)}, \quad (162)$$

$$-i^{z}k^{x}H - \frac{\partial^{z}H}{\partial x} = i\varepsilon\varepsilon_0\omega^{y}E \text{ (TE group)}, \quad (163)$$

$$\frac{\partial^{y}H}{\partial x} = i\varepsilon\varepsilon_0\omega^{z}E \text{ (TM group)}. \quad (164)$$

We can see that the fields have now split into two separate groups, namely, ${}^{y}E$, ${}^{x}H$ and ${}^{z}H$ are coupled and ${}^{y}H$, ${}^{x}E$ and ${}^{z}E$ are also coupled. The guided waves formed by the first group are described as TE modes (for transverse electric), and the latter are known as TM modes (for transverse magnetic). We can now use the above relations to derive simpler expressions for the transverse field components in terms of the ${}^{z}E$ and ${}^{z}H$ components only e.g., eliminate ${}^{y}H$ between Equations 160, 162 and 164 to obtain a relation for ${}^{x}E$. This yeilds the following

$${}^{x}E = \left(\frac{-i^{z}k}{\varepsilon\varepsilon_0\mu\mu_0\omega^2 - {}^{z}k^2}\right)\frac{\partial^{z}E}{\partial x}, \quad (165)$$

$${}^{y}E = \left(\frac{-i\omega\mu\mu_0}{\varepsilon\varepsilon_0\mu\mu_0\omega^2 - {}^{z}k^2}\right)\frac{\partial^{z}H}{\partial x}, \quad (166)$$

$${}^{x}H = \left(\frac{-i^{z}k}{\varepsilon\varepsilon_0\mu\mu_0\omega^2 - {}^{z}k^2}\right)\frac{\partial^{z}H}{\partial x}, \quad (167)$$

$${}^{y}H = \left(\frac{-i\omega\varepsilon\varepsilon_0}{\varepsilon\varepsilon_0\mu\mu_0\omega^2 - {}^{z}k^2}\right)\frac{\partial^{z}E}{\partial x}, \quad (168)$$

Substituting these expressions into Equations 159 to 164 yields two wave equations for propagation in the x direction:

$$\frac{\partial^{2z}E}{\partial x^2} - (\omega^2\varepsilon\varepsilon_0\mu\mu_0 - {}^{z}k^2)^{z}E = 0, \quad (169)$$

$$\frac{\partial^{2z}H}{\partial x^2} - (\omega^2\varepsilon\varepsilon_0\mu\mu_0 - {}^{z}k^2)^{z}H = 0. \quad (170)$$

These indicate that for the transverse dependence of the fields we should seek solutions of the form $\exp(i^{x}kx)$, where

$${}^{x}k^2 = (\omega^2\varepsilon\varepsilon_0\mu\mu_0 -^{z} k^2) = -({}^{x}\gamma)^2. \quad (171)$$

The significance of the variable ${}^{x}\gamma$ introduced here will become apparent later. Notice that Equation 171 could have been obtained much more straightforwardly by deriving the wave equation directly from Equations 142 to 147, setting $\partial/\partial y = 0$, and substituting the field Equations 148–149. However, that route would not have yielded detailed interrelationship between the vector components of the field that we will need for finding the conditions for guided waves. In conclusion, derived relationships between the vector components of the fields for a planar structure lying in the y–z plane, with a wave propagating in the z direction are summarized in Equations 159 to 164. In addition we have shown that fields in such a structure take the general form

$$\vec{E} = \vec{E}(x)\exp i(\omega t - {}^{z}kz \pm {}^{x}kx), \quad (172)$$

$$\vec{H} = \vec{H}(x)\exp i(\omega t - {}^{z}kz \pm {}^{x}kx), \quad (173)$$

$${}^{x}k^2 = \omega^2\varepsilon\varepsilon_0\mu\mu_0 - {}^{z}k^2, \quad (174)$$

$$n^2k_0^2 = \omega^2\varepsilon\varepsilon_0\mu\mu_0. \quad (175)$$

3.2.2. Cylindrical geometry

We now repeat the analysis of Section 3.2.1 but in cylindrical polar coordinates, since these are more appropriate to the analysis of optical fiber guides. The coordinates x, y and z are now replaced by r, ϕ and z. These are related to the coordinates as follows:

$$x = r\cos\phi, \quad (176)$$

$$y = r\sin\phi, \quad (177)$$

$$z = z. \quad (178)$$

Since we are still concerned with a structure that is expected to guide waves in the z direction, we should postulated fields of the form

$$\vec{E} = \vec{E}(r,\phi)\exp i(\omega t - {}^{z}kz), \quad (179)$$

$$\vec{H} = \vec{H}(r,\phi)\exp i(\omega t - {}^{z}kz). \quad (180)$$

The relation for $\vec{\nabla} \times \vec{A}$ in polar coordinates is as follows:

$$\vec{\nabla} \times \vec{A} = \begin{vmatrix} \frac{\vec{r}}{r}\phi\frac{\vec{k}}{r} \\ \frac{\partial}{\partial r}\frac{\partial}{\partial \phi}\frac{\partial}{\partial z} \\ {}^{r}\mathrm{A}r^{\phi}\mathrm{A}^{z}\mathrm{A} \end{vmatrix}, \quad (181)$$

where $\vec{k}$ is the unit vector in the z direction. We will now derive the expressions for the field components by use of Maxwell Equations 142–147. We obtain the set of relations equivalent to equations in Cartesian coordinates for the planar case:

$$\frac{1}{r}\left[\frac{\partial^{z}E}{\partial\phi} + i^{z}k(r^{\phi}E)\right] = -i\mu\mu_0\omega^{r}H, \quad (182)$$

$$i^{z}k^{r}E + \frac{\partial^{z}E}{\partial r} = i\mu\mu_0\omega^{\phi}H, \quad (183)$$

$$\frac{1}{r}\left[r\frac{\partial^{\phi}E}{\partial r} - \frac{\partial^{r}E}{\partial\phi}\right] = -i\mu\mu_0\omega^{z}H, \quad (184)$$

$$\frac{1}{r}\left[\frac{\partial^{z}H}{\partial\phi} + i^{z}k(r^{\phi}H)\right] = i\varepsilon\varepsilon_0\omega^{r}E, \quad (185)$$

$$\frac{1}{r}\left[r\frac{\partial^{\phi}H}{\partial r} - \frac{\partial^{r}H}{\partial\phi}\right] = i\varepsilon\varepsilon_0\omega^{z}E. \quad (186)$$

Solving these system of equations, we obtain expressions for the r and ϕ components only in terms of the z components:

$$ {}^{r}E = \frac{-i}{T_{k^2}}\left[{}^{z}k\frac{\partial^{z}E}{\partial r} + \mu\mu_0\omega\frac{1}{r}\frac{\partial^{z}H}{\partial\phi}\right], \quad (187) $$

$$ {}^{\phi}E = \frac{i}{T_{k^2}}\left[\frac{{}^{z}k}{r}\frac{\partial^{z}E}{\partial\phi} - \mu\mu_0\omega\frac{1}{r}\frac{\partial^{z}H}{\partial r}\right], \quad (188) $$

$$ {}^{r}H = \frac{-i}{T_{k^2}}\left[{}^{z}k\frac{\partial^{z}H}{\partial r} - \varepsilon\varepsilon_0\omega\frac{1}{r}\frac{\partial^{z}E}{\partial\phi}\right], \quad (189) $$

$$ {}^{\phi}H = \frac{-i}{T_{k^2}}\left[\frac{{}^{z}k}{r}\frac{\partial^{z}H}{\partial\phi} + \varepsilon\varepsilon_0\omega\frac{\partial^{z}E}{\partial r}\right], \quad (190) $$

$$ {}^{T}k^2 = \omega^2\varepsilon\varepsilon_0\mu\mu_0 - {}^{z}k^2 = n^2k_0^2 - {}^{z}k^2. \quad (191) $$

Here ${}^{T}k$ is the total transverse component of the $\vec{k}$ in the waveguide.

3.2.3. The electromagnetic wave equation

Below we derive the standard derivation of the wave equation. Moreover, we give the form of the Laplacian operator for rectangular Cartesian and polar coordinates. If we take the curl of the first Maxwell equation, then we obtain:

$$ \vec{\nabla} \times (\vec{\nabla} \times \vec{E}) = -\mu\mu_0\left[\vec{\nabla} \times \frac{\partial\vec{H}}{\partial t}\right]. \quad (193) $$

Differentiating Equation 143 with respect to time yields

$$ \vec{\nabla} \times \frac{\partial\vec{H}}{\partial t} = \varepsilon\varepsilon_0\frac{\partial^2\vec{E}}{\partial t^2}. \quad (194) $$

We then make use of the vector identity

$$ \vec{\nabla} \times (\vec{\nabla} \times \vec{E}) = -\vec{\nabla}(\vec{\nabla} \times \vec{E}) - \vec{\nabla}^2\vec{E} = -\vec{\nabla}^2\vec{E} \quad (195) $$

since $\vec{\nabla} \cdot \vec{E} = 0$ (see also Equation 146). Then it follows directly by substitution that

$$ \vec{\nabla}^2\vec{E} = \varepsilon\varepsilon_0\mu\mu_0\frac{\partial^2\vec{E}}{\partial t^2} \quad (196) $$

and likewise

$$ \vec{\nabla}^2\vec{H} = \varepsilon\varepsilon_0\mu\mu_0\frac{\partial^2\vec{H}}{\partial t^2}. \quad (197) $$

These equations are both of the general form

$$ \vec{\nabla}^2\vec{A} = \frac{1}{V^2}\frac{\partial^2\vec{A}}{\partial t^2}, \quad (198) $$

where V is the velocity of propagation (phase velocity) of the wave in the medium. It follows as usually that

$$ V_p = \frac{1}{\sqrt{\varepsilon\varepsilon_0\mu\mu_0}} \quad (199) $$

and that, for free space, we have the velocity of light, c, given by

$$ c = \frac{1}{\sqrt{\varepsilon_0\mu_0}}. \quad (200) $$

For planar waveguides, described by rectangular Cartesian coordinates, or circular fibers, described by cylindrical polar coordinates, the Laplacian operator has the forms

$$ \nabla^2 A = \frac{\partial^2 A}{\partial x^2} + \frac{\partial^2 A}{\partial y^2} + \frac{\partial^2 A}{\partial z^2}, \quad (201) $$

$$ \nabla^2 A = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial A}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 A}{\partial\phi^2} + \frac{\partial^2 A}{\partial z^2}. \quad (202) $$

We should stressed here that the modes considered in fiber optics are exact solution of Maxwell's equations (details see [292–294]).

3.3. Geometrical optics of fibers

As was noted above the mechanism of light propagation along fibers as small as a few wavelength in diameter can be understood almost entirely using the ray theory and well-known principles of geometrical optics. It is interesting to note that, even for smaller-diameter dielectric cylinders, which act as waveguides, the geometrical optical theory, with some modification, helps in the understanding of the complex mechanisms. We begin ray analysis of multimode optical waveguides with the planar, or slab waveguide, which is the simplest dielectric structure for illustrating the principles involved, and has application in integrated optics [290, 291]. Since we can analyze its light transmission characteristics in terms of a superposition of ray paths, it is important to fully appreciate the behavior of individual rays. We will study the trajectories of rays within planar waveguides, concentrating on those rays—the bound rays—which propagate without loss of energy on a nonabsorbing waveguide, and can, therefore propagate arbitrarily large distances. The planar, or slab, waveguide is illustrated in Fig. 67. It consists typically of a core layer of thickness 2ρ sandwiched between two layers which form the cladding. As explained in the introduction, we assume, for simplicity, that the cladding is unbounded The planes $x = \pm\rho$ are the core-cladding interfaces. Since the waveguide extends indefinitely in all directions orthogonal to the x-axis, the problem is two dimensional [300]. The z-axis is located along the

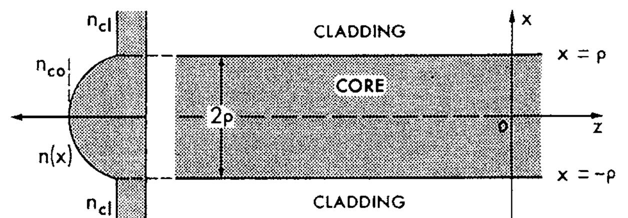


Figure 67 Nomenclature and coordinares for describing planar waveguides. A representative graded profile varies over the core and is uniform over cladding, assumed unbounded (after [292]).

axis of the waveguide midway between interfaces. The refractive-index profile $n(x)$ in Fig. 67 can be uniform or graded across the core, and assumes uniform value n_{cl} in the cladding. It is necessary that the core refractive index make some values greater than n_{cl} for the waveguide to have guidance properties. Furthermore, we assume that the profile does not vary with z, so that the waveguide is translationally invariant, or cylindrically symmetric. The parameters defined in Fig. 67 can be combined with the free-space wavelength λ of the light propagating along the waveguide to form a single dimensionless parameter V, known as the waveguide parameter, or waveguide frequency. If n_{co} is the maximum value of $n(x)$, which need not conduct with the on-axis value $n(0)$, then it will define

$$V = \frac{2\pi\rho}{\lambda}\left(n_{\mathrm{co}}^2 - n_{\mathrm{cl}}^2\right)^{1/2}. \quad (203)$$

The ray theory considered here is restricted to multimode waveguides i.e., waveguides satisfying $V \gg 1$. The step-index planar waveguide, according to Fig. 68, has the refractive-index profile defined by

$$\begin{aligned} n(x) &= n_{\mathrm{co}}, \quad -\rho < x < \rho; \\ n(x) &= n_{\mathrm{cl}}, \quad |x| > \rho, \end{aligned} \quad (204)$$

where n_{co} and n_{cl} are constants and $n_{\mathrm{co}} > n_{\mathrm{cl}}$. One of the most important problems is to determine the conditions necessary for a ray to be bound, i.e., the ray propagates along the nonabsorbing waveguide without loss of power.

Propagating within the uniform core of the step-index waveguide of Fig. 68 is along straight lines. If a ray originates at P on one interface and makes angle θ_z with the waveguide axis, it will meet the opposite interface at Q as shown in Fig. 68. The situation at Q in more details is pictured in Fig. 69 reflection in this situation is governed by Snell's law [293, 300]. While

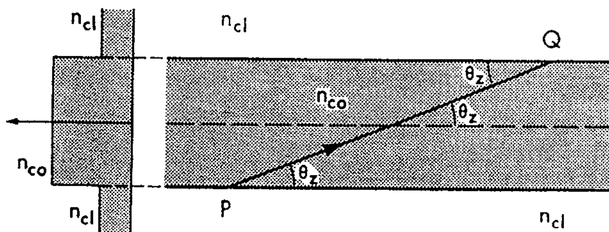


Figure 68 Propagation along a straight line between interfaces in the core of a step-profile planar waveguide (after [292]).

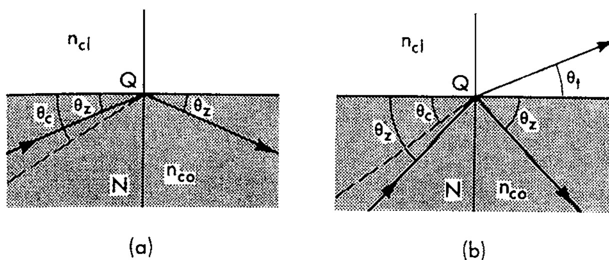


Figure 69 Reflection at a planar interface unbounded regions of refractive indices n_{co} and n_{cl}, showing (a) total internal reflection and (b) partial reflection and refraction (after [292]).

these laws are usually expressed in terms of angles relative to the normal QN, we following [292] prefer to retain the complementary angle θ_z. Thus, in terms of complementary angles, the incident ray at Q is totally internally reflected if $0 \preceq \theta_z \prec \theta_c$, and is partly refracted if $\theta_c \prec \theta_z \preceq \pi/2$, where θ_c is the complement of the critical angle, defined by [292]

$$\theta_c = \cos^{-1}\left[\frac{n_{\mathrm{cl}}}{n_{\mathrm{co}}}\right] = \sin^{-1}\left[1 - \frac{n_{\mathrm{cl}}^2}{n_{\mathrm{co}}^2}\right]^{1/2}. \quad (205)$$

In the first case, Fig. 115a shows the reflected ray leaving the interface at the same angle θ_z as the incident ray, while in the second case (see Fig. 69b) shows that the ray bifurcates, part of it being reflected at angle θ_z and part of it being transmitted into the cladding at angle θ_t to the interface, which satisfies Snell's law

$$n_{\mathrm{co}} \cos\theta_z = n_{\mathrm{cl}} \cos\theta_t. \quad (206)$$

Only total internal reflection returns all the ray power, i.e., the energy flowing along the ray, back into the core medium. A ray is reflected from the interface back into the core at angle θ_z regardless of whether partial or total reflection occur. If we repeat this procedure at successive reflections from the interfaces (see Fig. 70), in such way we construct the zig-zag paths, or trajectories. Path depicted in Fig 116a is a ray that is totally reflected at every reflection. We refer to this as a bound ray, since its path is entirely confined within the core. Path (see Fig. 70b) is for a ray that is partly reflected at each reflection. We refer to this as a refracting ray. The rays may be categorized by the value of θ_z according to [292]

$$\text{Bound rays:} \quad 0 \preceq \theta_z \prec \theta_c, \quad (207a)$$

$$\text{Refracting rays:} \quad \theta_c \preceq \theta_z \preceq \pi/2. \quad (207b)$$

Since the power of a bound ray is totally reflected back into the core at every reflection, the ray can propagate indefinitely without any loss of power. A refracting ray loses a fraction of its power at each reflection and therefore attenuates as it propagates.

As was shown in whole raw of textbooks (see, e.g. [292–298]), it is useful to introduce parameters that characterize ray propagation, as it is these parameters, rather than the spatial dependence of the ray path, that are important. As can be seen from Fig. 71 the ray trajectory is fully characterized once the angle θ_z is

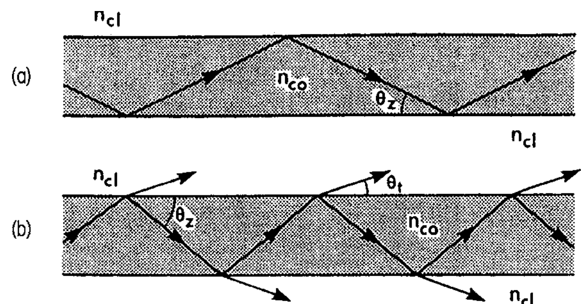


Figure 70 Zig-zag paths within the core of a step-profile planar waveguide for (a) bound rays and (b) refracting rays (after [292]).

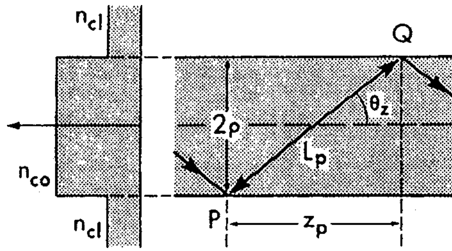


Figure 71 Path length and ray half-period z_p for a ray in the core of a step-profile planar waveguide (after [292]).

prescribed. We define the path length L_p between successive reflections to be the distance PQ. According geometry picture we have

$$L_p = \frac{2\rho}{\sin\theta_z} = \frac{2\rho n_{co}}{\left(n_{co}^2 - \bar{\beta}^2\right)^{1/2}}, \quad (208)$$

where $\bar{\beta} = n_{co}\cos\theta_z = n_{cl}\cos\theta_t$ is a ray invariant. For the ray transit time below, we require the optical path length L_0. In a homogeneous medium this is given by the product of path length and refractive index:

$$L_0 = n_{co}L_p = \frac{2\rho n_{co}}{\sin\theta_z} = \frac{2\rho n_{co}^2}{\left(n_{co}^2 - \bar{\beta}^2\right)^{1/2}}, \quad (209)$$

A quantity which will appear frequently in attenuation problems is the ray half-period z_p. This is distance between successive reflections , mesured along the waveguide axis:

$$z_p = \frac{2\rho}{\tan\theta_z} = L_p\cos\theta_z = \frac{2\rho\bar{\beta}}{\left(n_{co}^2 - \bar{\beta}^2\right)^{1/2}}. \quad (210)$$

Closely related is the number of reflection N per unit length of the waveguide, which is the reciprocal of the ray half-period. Hence

$$N = \frac{1}{z_p} = \frac{\tan\theta_z}{2\rho.}. \quad (211)$$

It is clear from these definitions that over arbitrary distance z along the waveguide, the accumulated path length, optical path length and number of reflections are given proportionally by

$$\frac{z}{z_p}L_p; \quad \frac{z}{z_p}L_0; \quad N_z = \frac{z}{z_p}, \quad (212)$$

respectively. These parameters are indicated in Table I-1 of [292]).

The most important quantity required to describe pulse spreading is the ray transit time t. This is the time a ray takes to propagate distance z along the waveguide, following the the zig-zag ray path (see also Fig. 70). The velocity of light v_g in fiber along the path is given by [300]

$$v_g = c/n_{co}, \quad (213)$$

where c is the fre-space speed of light. The transit time describes the following relation:

$$t = \frac{z}{z_p}\frac{L_p}{v_g} = \frac{zL_0}{z_p c} = \frac{zn_{co}^2}{c\bar{\beta}} = \frac{zn_{co}}{c\cos\theta_z}, \quad (214)$$

so that the greater θ_z, the longer the transit time.

We can account for material dispersion, which occurs when the refractive index varies with the wavelength of light λ, i.e., $n_{co} = n_{co}(\lambda)$. This requires more sophisticated reasoning relying on treating a ray as if it were a plane wave in local regions. Ray energy propagates at the group velocity v_g, which is given by Equation 213 in a dispersionless medium. but, allowing for material dispersion, it has the more general form [299]

$$v_g = c\left[n_{co}(\lambda) - \lambda\frac{dn_{co}(\lambda)}{d\lambda}\right]^{-1}. \quad (215)$$

It is convenient to introduce the group index n_g, described by

$$n_g = n_{co}(\lambda) - \lambda\frac{dn_{co}(\lambda)}{d\lambda}, \quad (216)$$

in which case, the transit time is expressible as

$$t = \frac{zn_g}{c\cos\theta_z} = \frac{zn_g n_{co}}{c\bar{\beta}}, \quad (217)$$

and varies with both θ_z and λ.

Early we establish the basic concepts for the ray analysis of planar waveguides. Now we extend the analysis to optical fibers, which are used for high-capacity communication over long distances. As far as ray tracing is concerned, the only difference between fibers and planar waveguides is the introduction of the third dimension. Thus, although the ray concepts are the same as early, the analysis and resulting expression are generally more complicated because of the fiber geometry [297]. Nevertheless, one of the important results of fiber optics shows that the ray transit times for step and clad power-law profile fibers of both circular and noncircular are identical to those of the corresponding planar waveguides. If this remarkable simplification is acceptable without proof, then pulse spreading in such fibers can be studied directly. An optical fiber is illustrated in Fig. 72. Unless otherwise stated, the core is assumed

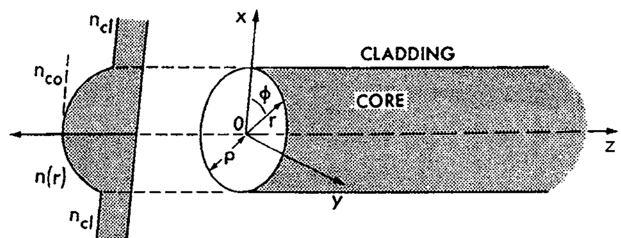


Figure 72 Nomenclature for describing circular fibers. Cartesian coordinates x, y, z and cylindrical coordinates r, ϕ, z are oriented so that the z-axis lies along the fiber axis. A representative graded profile varies over the core and is uniform over the cladding, assumed unbounded (after [292]).

to have a circularly symmetric cross-section of radius ρ, surrounded by the cladding, which, for simplicity is assumed unbounded. The core-cladding interface is the cylindrical surface $r = \rho$. Over the core, the axisymmetric refractive-index profile $n(r)$ is either uniform or graded, and it takes the uniform value n_{cl} in the cladding. The dimensionless parameter V of Equation 203 also applies to fibers, and will br referred as the fiber parameter, where ρ is the core radius. The quantity $(n_{co}^2 - n_{cl}^2)^{1/2}$ is often referred to as the numerical aperture of the fiber, while a related expression $[n_{co}^2(r) - n_{cl}^2]^{1/2}$ is sometimes called the local numerical aperture (details see [297, 298]).

3.4. Waveguide mode propagation

In a light pipe electromagnetic energy is propagated down the pipe by reflection from the walls of the structure. If the transverse dimensions are comparable to the wavelength of the light only certain field distributions (modes) will satisfy Maxwells equations and the boundary conditions. In this case the light pipe is more appropriately considered as a waveguide. Even in very large structures there are so many of them, their number increasing as the area, that in most cases a geometrical optics description is more fruitful (see above).

As is well-known, the distinction between metallic and dielectric waveguides is in the reflection mechanism responsible for confining the energy. The metallic guide does so by reflection from a good conductor at the boundary. In the dielectric waveguide, this is accomplished by total internal reflection, which is gotten by having the central dielectric made of a material of higher index of refraction than the surrounding dielectric. The two regions will henceforth be referred to as the core and cladding. In a metallic guide there are two sets of solutions, the transverse electric and transverse magnetic modes. In the dielectric guide all but the cylindrically symmetric modes TE_{0m} and TM_{0m} are hybrid, i.e., they have both electric and magnetic z components [301–303]. In general, one would expect two sets of such hybrid modes, because the boundary conditions give a characteristic equation which is quadratic in the Bessel functions describing the field in the central dielectric (details see [293, 297, 298] and references therein).

The cylindrical dielectric waveguide consists of a core of high refractive index n_{co} and radius a surrounded by a cladding of lower refractive index n_{cl}. Let the cladding material of index n_{cl} extend to infinity. We shall use both cartesian (x, y) and cylindrical polar coordinates (r, ϕ). The propagation constant β of any mode of this fiber is limited within the interval $n_{co} \geq \beta \geq n_{cl}k$, where $k = 2\pi/\lambda$ is the wavenumber in free space. If we define parameters

$$u = a\left(k^2 n_{co}^2 - \beta\right)^{1/2}, \tag{218}$$

$$w = a\left(\beta - k^2 n_{cl}^2\right)^{1/2}, \tag{219}$$

the mode field can be expressed by Bessel function $J(ur/a)$ inside the core and modified Hankel function $K(wr/a)$ outside the core [293, 302]. The quadratic summation

$$v^2 = u^2 + w^2 \tag{220}$$

leads to a third parameter

$$v = ak\left(n_{co}^2 - n_{cl}^2\right)^{1/2}, \tag{221}$$

which can be considered as a normalized frequency. By matching the fields at the core-cladding interface, we obtain characteristic functions $u(v)$ or $w(v)$ for every mode; the propagation constant and all other parameters of interest can be derived from these functions. For weak guidance, we have

$$\Delta = (n_{co} - n_{cl})/n_{cl} \ll 1. \tag{222}$$

In this case, we can construct modes whose transverse field is essentially polarized in one direction. Below we postulate transverse field components (see also [297])

$$\begin{aligned} {}^{y}E = {}^{z}H &= Z_0/n_{co} \\ &= E_l[J_l(ur/a)/J_l(u)]\cos l\phi \qquad (223a) \\ {}^{y}E = {}^{z}H &= Z_0/n_{cl} \\ &= E_l[K_l(wr/a)/K_l(w)]\cos l\phi. \qquad (223b) \end{aligned}$$

Here, as in following, the (a) holds for the core and the (b) for the cladding; Z_0 is the plane wave impedance in vacuum, and E_l the electrical field strength at the interface. Fig. 73b–e illustrate the case l = 1. Since we

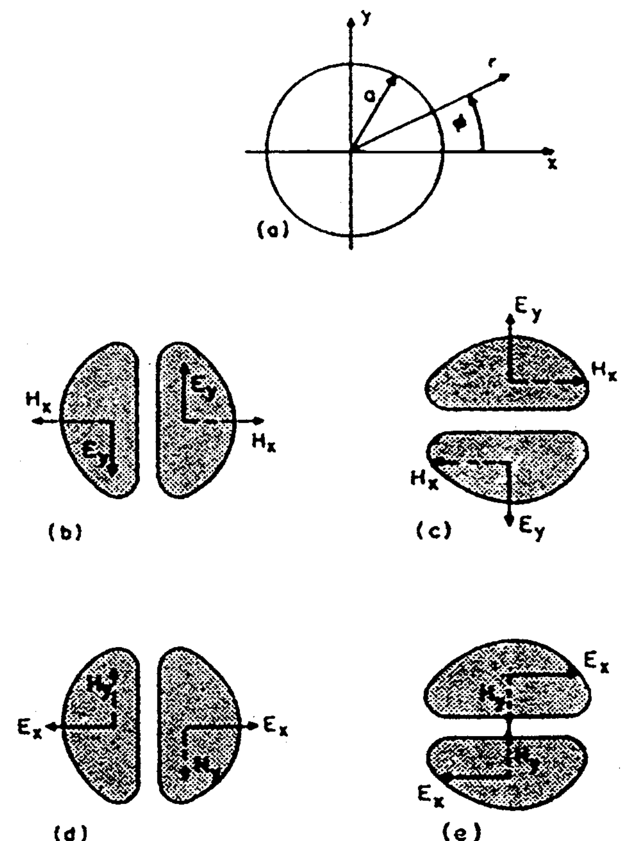


Figure 73 Sketch of the fiber cross-section and the four possible distributions of LP_{11} (after [303]).

have the freedom of choosing $\sin l\phi$ or $\cos l\phi$ in Equation 223 and two orthogonal states of polarization, we can construct a set of four modes for every l as long as $l > 0$. For $l = 0$, we have only a set of two modes polarized orthogonally with respect to each other. The longitudinal components can be obtained from the equations [304]

$$ {}^{z}E = \frac{iZ_0}{k}\left[\frac{1}{n_{co}^2}\right]\frac{\partial^z H}{\partial y}, \tag{224a} $$

$$ {}^{z}E = \frac{iZ_0}{k}\left[\frac{1}{n_{cl}^2}\right]\frac{\partial^z H}{\partial y}, \tag{224b} $$

and

$$ {}^{z}H = (i/kZ_0)(\partial^y E/\partial x). \tag{224c} $$

By introducing (446), we have

$$ {}^{z}E = \frac{-iE_l}{2ka}\left[\frac{u}{n_{co}}\frac{J_{l+1}(ur/a)}{J_l(u)}\sin(l+1)\phi + \frac{u}{n_{co}}\frac{J_{l-1}(ur/a)}{J_l(u)}\sin(l-1)\phi\right], \tag{225a} $$

$$ {}^{z}E = \frac{-iE_l}{2ka}\left[\frac{w}{n_{cl}}\frac{K_{l+1}(wr/a)}{K_l(w)}\sin(l+1)\phi + \frac{w}{n_{cl}}\frac{K_{l-1}(wr/a)}{K_l(w)}\sin(l-1)\phi\right], \tag{225b} $$

$$ {}^{z}H = \frac{-iE_l}{2kZ_0a}\left[u\frac{J_{l+1}(ur/a)}{J_l(u)}\cos(l+1)\phi - u\frac{J_{l-1}(ur/a)}{J_l(u)}\cos(l-1)\phi\right], \tag{225c} $$

$$ {}^{z}H = \frac{-iE_l}{2kZ_0a}\left[w\frac{K_{l+1}(wr/a)}{K_l(w)}\cos(l+1)\phi + w\frac{K_{l-1}(wr/a)}{K_l(w)}\cos(l-1)\phi\right]. \tag{225d} $$

For small Δ, the longitudinal components [Equations 225a–d] are small compared to the transverse components. The factors involved are u/ak and w/ak which because of Equations 218 and 219 are both of the order $\Delta^{1/2}$. Repeated differentiation of Equations 225a–d) leads to transverse components which are not identical with the postulated field [Equation 222] but small of order Δ compared to it. We shall neglect these fields in the following. It is this approximation that determines the accuracy of assumption of linearly polarized modes (see also [298]).

To match the fields at the interface let us write Equation 222 in terms of cylindrical components. We then have

$$ {}^{\phi}E = \frac{E_l}{2}[J_l(ur/a)/J_l(u)][\cos(l+1)\phi + \cos(l-1)\phi], \tag{226a} $$

$$ {}^{\phi}E = \frac{E_l}{2}[K_l(wr/a)/J_l(w)][\cos(l+1)\phi + \cos(l-1)\phi], \tag{226b} $$

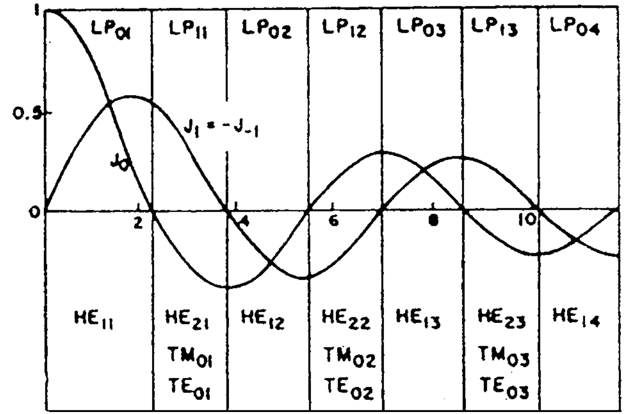


Figure 74 The regions of the parameter u for modes of order $l = 0$; 1 (after [298]).

$$ {}^{\phi}H = -\frac{E_l}{2Z_0}[n_{co}J_l(ur/a)/J_l(u)][\sin(l+1)\phi - \sin(l-1)\phi], \tag{226c} $$

$$ {}^{\phi}H = -\frac{E_l}{2Z_0}[n_{cl}K_l(wr/a)/J_l(w)][\sin(l+1)\phi - \sin(l-1)\phi], \tag{226d} $$

If we set $n_{co} = n_{cl}$ in Equations 226–227 and use the recurrence relations for J_l and K_l, we can match all tangential field components at the interface by the one equation

$$ u[J_{l-1}(u)/J_l(u)] = -w[K_{l-1}(w)/K_l(w)]. \tag{227} $$

This is the characteristic equation for the linearly polarized (LP) modes. Setting $w = 0$ yields the cutoff values $J_{l-1}(u) = 0$. For $l = 0$, this includes the roots of the Bessel function $J_{-1}(u) = -J_l(u)$, which we shall count so as to include $J_l(0) = 0$ as the first root. In such way Gloge [303] obtain the cutoff values indicated in Fig. 74 for LP_{0m} and LP_{1m}. In the limit of $w \to \infty$ we have $J_l(0) = 0$. Thus, the solution for u are between the zeros of $J_{l-1}(u)$ and $J_l(u)$. Every solution is associated with one set of modes designed LP_{lm}. For $l \geq 1$, each set composes four modes.

The accuracy of the characteristic equation can be improved if we retain n_{cl} and n_{co} as different in Equations 216 and 217. In this case, as was shown by Gloge [303], terms with $(l+1)\phi$ and $(l-1)\phi$ satisfy two different characteristic equations:

$$ (u/n_{co})[J_{l\pm1}(u)/J_l(u)] = \pm(w/n_{cl})[K_{l\pm1}(w)/K_l(w)] \tag{228} $$

By using the recurrence relations for J_l and K_l, Gloge have shown that these two equations converge into Equation 227 for $n_{co} = n_{cl}$. For $n_{cl} \neq n_{co}$, this degeneracy ceases to exist; each mode LP_{lm} breaks up into modes with terms $(l+1)\phi$, which can be identified as $HE_{l+1,m}$ or TE_m and TM_m (see also [292, 302]). A more rigorous proof of this results is given below.

As mentioned above, the problem of the dielectric cylinder with sharp index step can be solved exactly.

Using above done the nomenclature, one can write the exact characteristic equation in the form

$$(Q - D - 2\Delta\{[(l \pm 1)/\omega^2] \pm [K_l(\omega)/K_{l\pm1}(\omega)]\})$$
$$\times (Q - D) = Q^2[1 - 2\Delta(u^2/v^2)], \qquad (229)$$

where

$$Q = (l \pm 1)(v^2/u^2w^2). \qquad (230)$$
$$D = [J_l(u)/uJ_{l\pm1}(u)] \mp [K_l(w)/wK_{l\pm1}(w)], \quad (231)$$

and

$$2\Delta = \left(n_{co}^2 - n_{cl}^2\right)/n_{co}^2. \qquad (232)$$

The upper sign holds for HE_{l+1} modes and the lower sign for EH_{l-1}, TM and TE. Equation 232 agrees with Equation 222 in the case of small index differences. If Δ is set to zero in Equation 229, we find $D = 0$; and Equation 231 then becomes the simplified characteristic Equation 227. For small Δ, D is also small. Now we simplify Equation 220 to the extent that we retain terms linear in Δ or D. This result in

$$D = \Delta\{Q(u^2/v^2) - [(l \pm 1)/w^2]$$
$$\mp [K_l(w)/wK_{l\pm1}(w)]\} \qquad (231)$$

and with Equation 230

$$D = \mp\Delta[K_l(w)/wK_{l\pm1}(w)]. \qquad (232)$$

By introducing this into Equation 231 and inserting Equation 232, we find

$$(u/n_{co})[J_{l\pm1}(u)/J_l(u)]$$
$$= \pm(w/n_{cl})[K_{l\pm1}(w)/K_l(w)] \qquad (233)$$

This is exactly the characteristic Equation 228.

Evidently the guided wave traveling along the circular guide carries energy. The respective amounts are readily calculated using the Poynting vector to estimate the energy flow [298]. The Poynting vector in axial direction can be calculated from the cross product of the transverse fields given in Equation 222. Integration over the cross-section of core and cladding leads to tabulated integrals (see also [298, 301, 305]); the results are

$$P_{co} = [1 + (w^2/u^2)(1/k)](\pi a^2/2)(Z_0/n_{co})E_l^2 \quad (234)$$

and

$$P_{cl} = [(1/k) - 1](\pi a^2/2)(Z_0/n_{cl})E_l^2 \qquad (235)$$

for the power flow in core and cladding, respectively. If we ignore the small difference between n_{co} and n_{cl}, the total power in a certain mode becomes

$$P = P_{co} + P_{cl}$$
$$= (v^2/u^2)(1/k)(\pi a^2/2)(Z_0/n_{co})E_l^2. \qquad (236)$$

Practical fibers have small heat and scattering losses which cause significant attenuation over long distances. In general, these losses are attributable to certain parts of the fiber and proportional to the power propagating in this part. For considerations of this kind, it is convenient to use the power fractions

$$P_{co}/P = 1 - (u^2/v^2)(1 - k) \qquad (237)$$

and

$$P_{cl}/P = (u^2/v^2)(1 - k). \qquad (238)$$

As expected, the mode power is concentration in the core far away from cutoff. As cutoff is approached, the power of lower order modes ($l = 0, 1$) withdraws into the cladding, whereas modes with $l \geq 2$ maintain a fixed ratio of $(l-1)$ between the power in core and cladding at cutoff. The power density is related to the mode power P by

$$p(r) = k\frac{u^2}{v^2}\frac{2P}{\pi a^2}\left[J_l^2(ur/a)/J_l^2(u)\right]\cos^2 l\phi \qquad (239a)$$

$$p(r) = k\frac{u^2}{v^2}\frac{2P}{\pi a^2}\left[K_l^2(wr/a)/K_l^2(w)\right]\cos^2 l\phi. \quad (239b)$$

By averaging over ϕ at $r = a$, we obtain the mean density

$$\bar{p}(r) = k(u^2/v^2)\frac{P}{\pi a^2}\left[J_l^2(ur/a)/J_l^2(u)\right] \qquad (240a)$$

$$\bar{p}(r) = k(u^2/v^2)\frac{P}{\pi a^2}\left[K_l^2(wr/a)/K_l^2(w)\right] \quad (240b)$$

At the core-cladding interface, we have $r = a$ and

$$\bar{p}(a) = k(u^2/v^2)\frac{P}{\pi a^2}. \qquad (241)$$

The normalized density $\pi a^2\bar{p}(a)/P$ is plotted in Fig. 75. For modes of order $l = 0, 1$ this density approaches zero both at cutoff and far away from it, having a maximum in between. Modes with $l \geq 2$ have $\bar{p}(a) = [1-(1/l)]\ P/\pi a^2$ at cutoff. For $r \gg a/w$, we can replace the K functions in Equation 241 by their approximation for large argument and obtain

$$\bar{p}(r) \approx k(u^2/v^2)(P/\pi ar)\exp[-2w(r - a)/a],$$
$$\text{for } r \gg a, \quad (242)$$

as long as w is not too small. The power density decreases exponentially with the distance from the interface (see also [296]). It decreases sharply as cutoff is approached and is zero at cutoff. For sufficiently small w we may set $u = v$ and replace the K functions in Equation 241 by their approximation for small argument, obtaining

$$\bar{p}(r) \approx k_l(P/\pi a^2)(a/r)^l \qquad (243)$$

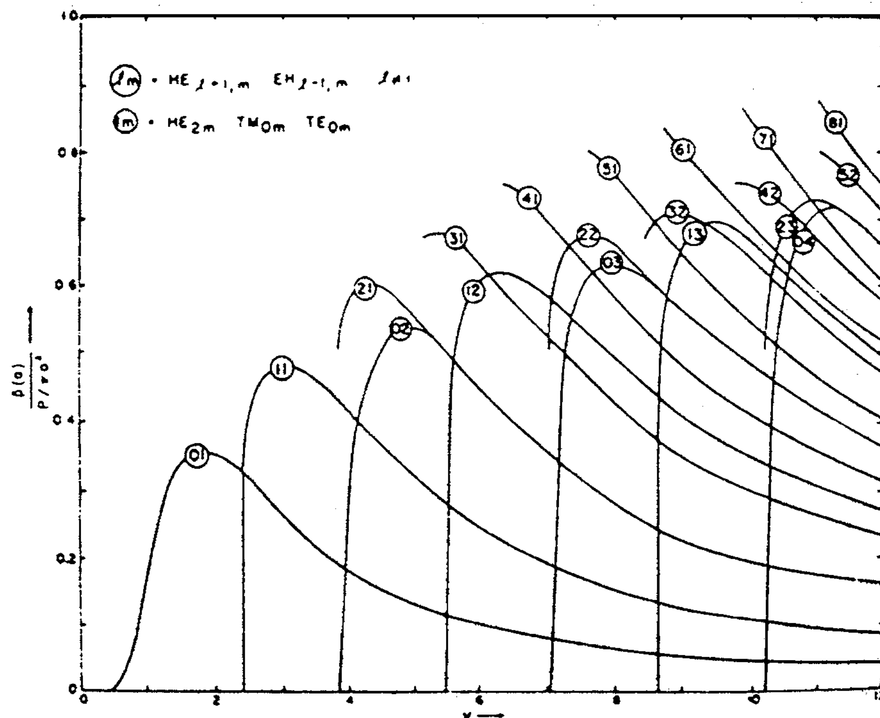


Figure 75 Normalized power density at the core-cladding interface plotted vs v (after [294]).

for $r > a$, $w = 0$. This function describes the cutoff power distribution in the cladding. It decreases with the distance from the axis for all but the lowest azimuthal order, whose cladding field is independent of the radius.

3.5. Pulse spreading

The transmission of information along optical fibers is normally achieved by sending out a sequence of pulses of light energy. However, as an individual pulse propagates, it spreads out, due to the dispersive properties of the fiber. Clearly if this spread becomes sufficiently large the pulse will overlap with adjacent pulses, leading to a decrease in information-carrying capacity because of the loss of resolution at the end of the fiber.

Below a formalism is presented for describing the propagation characteristics of graded-index, multimode fibers. The index profiles of cylindrically symmetric waveguides can be conveniently specified by the equation

$$n^2(r) = n_1^2[1 - 2\Delta f(r/a)], \tag{244}$$

where $n(r)$ is the refractive index of the waveguide as a function of distance r from the axis, and n_1 is the index along the axis. The profile function $f(r/a)$ is defined so that it is zero on axis

$$f(0) = 0 \tag{245}$$

and becomes equal to unity at the core-cladding boundary located at $r = a$,

$$f(r/a) = 1, \quad \text{for } r \geq a. \tag{246}$$

The cladding index n_2 $(= n_{\text{cl}})$ is thus defined to be

$$n_2 = n_1(1 - 2\Delta)^{1/2}. \tag{247}$$

The quantity Δ provides a useful measure of the core-cladding index difference. From Equation 247 one finds

$$\Delta = \left(n_1^2 - n_2^2\right)/2n_1^2. \tag{248}$$

Each mode of the waveguide can be specified by the pair of integers μ and ν, which, respectively specify the number of radial nodes and azimuthal nodes in the transverse electromagnetic fields of that mode. The propagation constant $\beta_{\mu\nu}$ of each mode depends explicitly on all quantities that specify the waveguide structure and on the wavelength λ of the propagating light (see also [303]),

$$\beta_{\mu\nu} = \beta_{\mu\nu}(n_1, \Delta, a, \lambda). \tag{249}$$

The propagation constants depend as usually on the wavelength explicitly and implicitly through the wavelength variation of n_1 and Δ. Although the index profile $f(r/a)$ may also vary slightly with wavelength, such effects will not considered.

For analyzing pulse transmission, one is concerned with the group delay time per unit length for the mode μ, ν. This is given

$$\tau_{\mu\nu} = \frac{\mathrm{d}\beta_{\mu\nu}}{\mathrm{d}\omega}. \tag{250}$$

If the free space propagation constant, $k = 2\pi/\lambda$, is introduced, Equation 250 can be rewritten as

$$\tau_{\mu\nu} = \frac{1}{c}\frac{\mathrm{d}\beta_{\mu\nu}}{\mathrm{d}k}. \tag{251}$$

If the fiber is excited by an impulse excitation and if there is no mode coupling, the impulse response, $P(t, z, \lambda)$, for spectral component λ at position z can be written as

$$P(t, z, \lambda) = \sum P_{\mu\nu}(\lambda, z)\delta[t - z\tau_{\mu\nu}(\lambda)], \quad (252)$$

where the summation extends over all guided modes. The distribution functions $P_{\mu\nu}(\lambda, z)$ describe the power in mode μ, ν as a function of wavelength and position. At $z = 0$ the distribution function will be determined by the spatial, angular, and spectral distribution of the source, as well as by the source-fiber coupling configuration. As the impulse propagates along the waveguide the power in each mode will change according to the attenuation occurring in that mode. Further it will be assumed that no mode coupling occurs.

In general, only the total power integrated over all source wavelengths will be detected. Hence, the quantity of practical interest is the full impulse response

$$P(t, z) = \int_0^\infty \mathrm{d}\lambda P(t, z, \lambda). \quad (253)$$

The propagation characteristics of the fiber can be described by specifying the moments $M_n(z)$ of the full impulse response. These moments are defined by

$$M_n(z) = \int_0^\infty \mathrm{d}t t^n P(t, z). \quad (254)$$

In some situations, knowledge of only the first few moments is sufficient for system design considerations. If this is the case, the required amount of pulse broadening information is reduced considerably. Equations 252–254 can be combined to yield

$$M_n(z) = z^n \int_0^\infty \mathrm{d}\lambda \sum P_{\mu\nu}(\lambda, z)\tau^n_{\mu\nu}(\lambda). \quad (255)$$

The predominant wavelength dependence of the distribution function $P_{\mu\nu}(\lambda, z)$ is determined by the spectral distribution $S(\lambda)$ of the source. Even for the relatively broad LED sources, $S(\lambda)$ is a sharply peaked function whose rms width, at most, a few percent of the mean source wavelength. One can rhus define a new distribution function $p_{\mu\nu}(\lambda, z)$ by the expression

$$P_{\mu\nu}(\lambda, z) = S(\lambda)p_{\mu\nu}(\lambda, z), \quad (256)$$

where $p_{\mu\nu}(\lambda, z)$ is a slowly varying function of λ over the range where $S(\lambda)$ is nonzero.

Proceedings with the analysis, it can be assumed that $S(\lambda)$ is normalized so that

$$\int_0^\infty \mathrm{d}\lambda S(\lambda) = 1. \quad (257)$$

Consequently, the mean source wavelength λ_0 is given by

$$\lambda_0 = \int_0^\infty \mathrm{d}\lambda\lambda S(\lambda), \quad (258)$$

and the root mean square (rms) spectral width of the source σ_s is given by

$$\sigma_s = \left[\int_0^\infty \mathrm{d}\lambda(\lambda - \lambda_0)^2 S(\lambda)\right]^{1/2}. \quad (259)$$

The influence of the source spectral distribution on the fiber's transmission properties can be studied by expanding the delay per unit length of the μ, νth mode, $\tau_{\mu\nu}(\lambda)$, in a Taylor series about λ_0. Substituting this series into Equation 255 and using Equation 256 give

$$\begin{aligned} M_n(z) = z^n \int_0^\infty \mathrm{d}\lambda S(\lambda) \sum_{\mu\nu} p(z)\{\tau^n_{\mu\nu}(\lambda_0) \\ + n(\lambda - \lambda_0)\tau^{n-1}_{\mu\nu}(\lambda_0)\tau'_{\mu\nu}(\lambda_0) \\ + n(\lambda - \lambda_0)^2/2\tau^{n-2}_{\mu\nu}(\lambda_0)\tau''_{\mu\nu}(\lambda_0) + \}. \end{aligned} \quad (260)$$

Treating $p_{\mu\nu}$ as independent of λ, Equations 255–259 can be used to integrate Equation 260 to find that

$$\begin{aligned} M_n(z) = z^n \sum p_{\mu\nu}(z)\tau^n_{\mu\nu}(\lambda_0) + \sigma_s^2/(2\lambda_0^2) \\ \times \{n\tau^{n-1}_{\mu\nu}(\lambda_0)\lambda_0^2\tau''_{\mu\nu}(\lambda_0) \\ + n(n-1)\tau^{n-2}_{\mu\nu}(\lambda_0)[\lambda_0\tau'_{\mu\nu}(\lambda_0)]^2\}) \\ + 0(\sigma_s^3/\lambda_0^3). \end{aligned} \quad (261)$$

The small size of σ_s/λ_0 allows the neglect of higher order terms. The following quantities are most useful in describing the energy distribution at z. By definition, the total power arriving at z is given by

$$M_0(z) = \sum p_{\mu\nu}(z), \quad (262)$$

the mean delay time of the pulse $\tau(z)$ is given by

$$\tau(z) = M_1(z)/M_0(z), \quad (263)$$

and the rms pulse width $\sigma(z)$ by

$$\sigma(z) = [M_2(z)/M_0(z) - \tau^2(z)]^{1/2}. \quad (264)$$

Combinations of higher moments further describe the power distribution, but the first three are the most important. To simplify the notation required in the following expressions, the symbol $\langle\ \rangle$ will be used to indicate the average value of a quantity with respect to the distribution $p_{\mu\nu}$ so, for example

$$\langle A\rangle \equiv \sum p_{\mu\nu}(z)A_{\mu\nu}/M_0. \quad (265)$$

From Equations 252–255, the full pulse delay time is found to be

$$\tau(z) = z\left[\langle\tau(\lambda_0)\rangle + \sigma_s^2/(2\lambda_0^2)\langle\lambda_0^2\tau''(\lambda_0)\rangle\right]. \quad (266)$$

For the purpose of specifying the fiber bandwidth for digital systems, Personick [306] has shown that one is

primarily concerned with the rms width $\sigma(z)$. This is given by

$$\sigma(z) = \left(\sigma^2_{\text{intermodal}} + \sigma^2_{\text{intramodal}}\right)^{1/2} + 0\left(\sigma_s^3/\lambda_0^3\right), \quad (267)$$

where the definitions

$$\begin{aligned}\sigma^2_{\text{intermodal}} &= z^2\Big\{\langle\tau^2(\lambda_0)\rangle - \langle\tau(\lambda_0)\rangle^2 \\ &\quad + \frac{\sigma_s^2}{\lambda_0^2}\big[\lambda_0^2\tau''(\lambda_0)\tau(\lambda_0) \\ &\quad - \langle\lambda_0^2\tau''(\lambda_0)\rangle\langle\tau(\lambda_0)\rangle\big]\Big\} \quad (268)\end{aligned}$$

$$\sigma^2_{\text{intramodal}} = z^2\frac{\sigma_s^2}{\lambda_0^2}\langle[\lambda_0\tau'(\lambda_0)]^2\rangle \quad (269)$$

have been introduced. We can see, that the square of rms width has been separated into an intermodal and an intramodal component. The intermodal term (262) results from delay differences among the modes and vanishes only if all delay differences vanish. This term is found to contain a dominant term and a small correction that is proportional to the square of the relative source spectral width (σ_s/λ_0). For the refractive index profiles considered below, this term is found to be negligible. The intramodal term (269) represents an average of the pulse broadening within each mode. It becomes the only term present in the dispersion of a single-mode waveguide. The intramodal dispersion arises from two distinct effects, a pure material effect that corresponds to the pulse broadening in bulk material and waveguide effect. This separation can be made by writing the modal delay time in the form

$$\tau_{\mu\nu} = N_1/c + \delta\tau_{\mu\nu}, \quad (270)$$

where N_1 is material group index,

$$N_1 = n_1\lambda \mathrm{d}n_1/\mathrm{d}\lambda, \quad (271)$$

and $\delta\tau_{\mu\nu}$ represents the correction to this introduced by the waveguide structure. The derivative $\tau'_{\mu\nu}$ can be written as

$$\tau'_{\mu\nu} = -\lambda n_1'' + \delta\tau'_{\mu\nu}. \quad (272)$$

Since the intramodal contribution to the total rms pulse width is obtained by squaring and averaging over Equation 273, one can write the intramodal contributions as

$$\begin{aligned}\sigma^2_{\text{intramodal}} &= z^2\frac{\sigma_s^2}{\lambda_0^2}\Big[\left(\lambda_0^2 n_1''\right)^2 - 2\lambda_0^2 n_1''\langle\lambda_0\delta\tau'\rangle \\ &\quad + \langle(\lambda_0\delta\tau')^2\rangle\Big]. \quad (273)\end{aligned}$$

Hence, the total intramodal contribution has a pure material component, a waveguide component, and a mixed component arising from the cross product. For graded-index waveguides the derivative of the intermodal delay differences $\delta\tau'_{\mu\nu}$ is also small, therefore the intramodal contribution is then dominated by the pure material term

$$\sigma_{\text{intramodal}} \approx \frac{\sigma_s}{\lambda_0}\left(\lambda_0^2 n_1''\right). \quad (274)$$

This quantity represents the ultimate lower limit of the pulse broadening. According to [307] the meaning of $\sigma_{\text{intermodal}}$ has next expression

$$\begin{aligned}&\sigma_{\text{intermodal}} \\ &= \frac{LN_1\Delta}{2c}\frac{\alpha}{\alpha+1}\left(\frac{\alpha+2}{3\alpha+2}\right)^{1/2} \\ &\quad \times\left[C_1^2 + \frac{4C_1C_2\Delta(\alpha+1)}{2\alpha+1} + \frac{4\Delta^2C_2^2(2\alpha+2)^2}{(5\alpha+2)(3\alpha+2)}\right]^{1/2}, \\ &\qquad (275)\end{aligned}$$

where

$$C_1 = \frac{\alpha-2-\varepsilon}{\alpha+2}, \quad (276)$$

and

$$C_2 = \frac{3\alpha-2-2\varepsilon}{2(\alpha+2)}. \quad (277)$$

As was shown in [308], the minimum of the intermodal rms width occurs for (see also [298])

$$\alpha_c = 2 + \varepsilon - \Delta\frac{(4+\varepsilon)(3+\varepsilon)}{(5+2\varepsilon)}. \quad (278)$$

Here $\varepsilon = \frac{-2n_1\lambda\Delta'}{N\Delta}$ and $\alpha = 2 + \varepsilon$. The last term on the right of Equation 278 represents a small correction to the optimal α. This correction results from a partial cancellation that occurs between the two mode dependent terms, if $\alpha - 2 - \varepsilon$ is of order Δ. The size of this correction changes as the distribution of power among the modes varies.

Fig. 76A shows the fitted values of $n_1(\lambda)$ and $n_2(\lambda)$ in the range from 0.5 μm to 1.1 μm. Figs 76B and 76C, respectively, show the first and second derivatives of the refractive indices of these two glasses. Fig. 77 shows the values of Δ and $\lambda\Delta'$ determined from the data. It can be observed in Fig. 77A that, over the range of wavelength shown, Δ decreases by about 15%. From the index data on Figs 76 and 77, $\varepsilon(\lambda)$ can be calculated, and a plot of α_c vs. λ is shown in Fig. 78. The value of α_c minimizing the intermodal rms pulse width departs significantly from the optimal profile $\alpha = 2(1 - 6\Delta/5)$, which is predicted if the effect of material dispersion is ignored. The optimal α varies quite strongly with wavelength, decreasing from about 2.5 at 500 nm to 2.2 at 1000 nm. A waveguide with a given index profile α is thus predicted to show different pulse widths according to the source wavelength used. This wavelength-dependent pulse broadening provides a tool for observing these effects and is discussed at length in [307].

According to [307] the delay time can be used to evaluate $\lambda\tau_n'$, which is required for predicting the intramodal

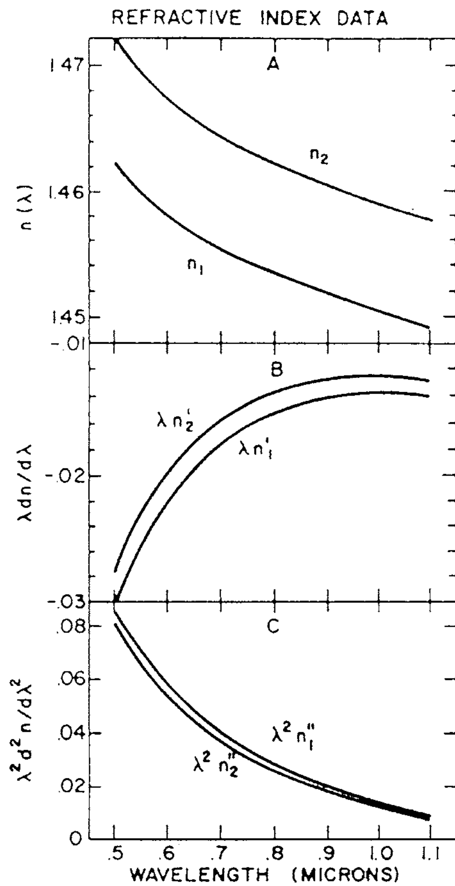


Figure 76 Refractive index data for 3.4 wt% TiO_2 doped silica (n_2) and fused silica (n_1) are shown for (A) Sellmeir fit to the refractive index, (B)—$\lambda dn/d\lambda$, (C)—$\lambda^2 d^2n/d\lambda^2$ (after [307]).

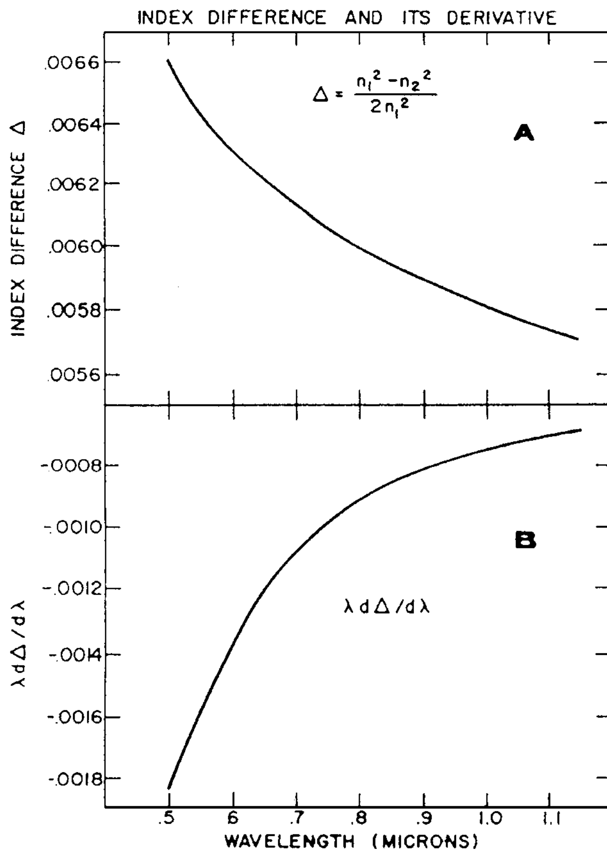


Figure 77 The index difference Δ determined from the data of Fig. 122A is shown in (A), and the derivative $\lambda d\Delta/d\lambda$ in (B) (after [307]).

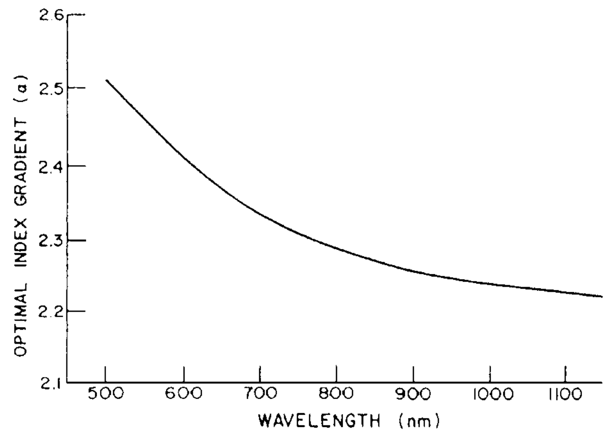


Figure 78 The α value that minimizes the pulse broadening as a function of wavelength (after [307]).

broadening has next form

$$\lambda\tau_n' = -\lambda^2 n_1'' + N_1\Delta\left(\frac{\alpha-2-\varepsilon}{\alpha+2}\right) \times\left(\frac{2\alpha}{\alpha+2}\right)\left(\frac{n}{N}\right)^{\alpha/\alpha+2}. \quad (279)$$

Since $\lambda^2 n_1''$ and Δ are the same order of magnitude, both terms contribute to $\lambda n_n'$ for large α (but with opposite signs). For α near α_c, the pure material term in Equation 273 dominates. This reflect the fact that when the intermodal delay differences are small, the derivatives of these differences are also small. Assuming equal excitation of the modes, one can perform the summation over all modes required by Equation 259 and find that

$$\sigma_{\text{intramodal}} = \frac{\sigma_\lambda}{\lambda}\left[(-\lambda^2 n_1'')^2 - 2\lambda^2 n_1''(N_1\Delta) \times\left(\frac{\alpha-2-\varepsilon}{\alpha+2}\right)\left(\frac{2\alpha}{2\alpha+2}\right) + (N_1\Delta)^2\left(\frac{\alpha-2-\varepsilon}{\alpha+2}\right)^2\left(\frac{2\alpha}{3\alpha+2}\right)\right]^{1/2}. \quad (280)$$

The total rms pulse widths for α profiles can be predicted from Equations 275 and 280. In Fig. 79, this rms pulse width is shown as a function of α for three types of GaAs sources, all operating at $\lambda = 0.9\ \mu$m, but having differential spectral bandwidths. The three curves correspond to an LED, an injection laser, and a distributed-feedback laser having typical rms spectral widths of 150 Å, 10 Å and 2 Å, respectively. The waveguide is assumed to have $n_1(\lambda)$ and $n_2(\lambda)$ of Fig. 76 and to propagate equal power in all modes. For comparison, a dashed curve represents the rms width predicted if material dispersion and intramodal broadening are ignored. We shoul add, that for the LED source, pulse broadening of less than 1.5 ns/km can be achieved if α is within 25% of the optimal value. For the injection laser, an α within 5% of the optimal will give pulse widths less than 0.2 ns/km, and for the distributed-feedback laser, widths of 0.05 ns/km are predicted if 1% control on α can be achieved.

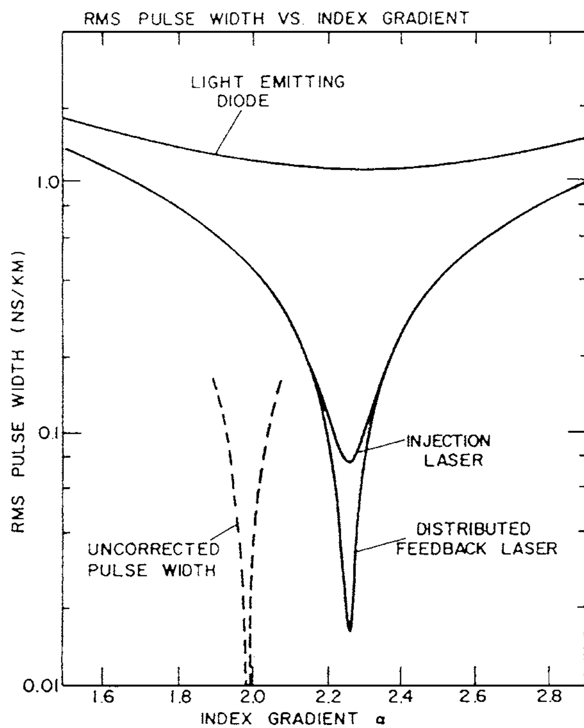


Figure 79 Assuming equal power in all modes, the rms pulse width is shown as a function of α for three different sources, all operating at 0.9 μm. The sources are taken to be an LED, a gallium arsenide injection laser, and a distributed feedback laser having rms spectral widths of 150 Å, 10 Å and 2 Å, respectively. The dashed curve shows the pulse width that would be predicted if al material dispersion effects were neglected (after [298]).

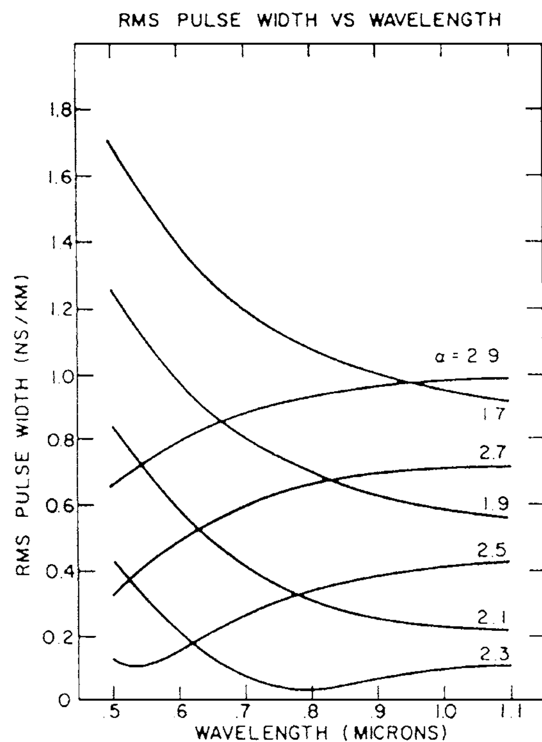


Figure 80 For a source with 2 Å spectral width, the rms pulse width is shown as a function of wavelength for several different values of α (after [307]).

In the preceding sections it was found that the α value that minimizes the intermodal pulse dispersion departs significantly from the parabolic profile and that the magnitude of this departure depends on the wavelength of the source. In Fig. 79, one sees that the calculated rms pulse width depends very strongly on the difference between the α of the waveguide and the optimal α. In Fig. 80, the calculated rms width is shown as a function of source wavelength for waveguides with index profiles in the range $1.5 \leq \alpha \leq 2.9$. For the purpose of illustration, it was assumed [307] in these calculation that the rms spectral width of the source is 2 Å. As is known, for this range of α values both the intermodal and intramodal contributions must be considered. The wavelength dependence of the intermodal term depends on the specific α value. For $\alpha \prec 2.2$, the intermodal pulse width decreases as λ increases, because for these α's the difference $|\alpha - \alpha_c|$ decreases with λ. The opposite occurs for $\alpha \succ 2.5$, and the intermodal pulse width increases with λ. For waveguides with $2.2 \prec \alpha \prec 2.5$, the minimal intermodal broadening occurs at one wavelength in the range 500 nm $\prec \lambda \prec$ 1100 nm. As λ is varied through this range, the intermodal broadening decreases until the optimal wavelength is reached, and thereafter it increases. The intramodal contribution to the rms width is dominated by material dispersion. It is largest at the shorter wavelength and decreases quite rapidly with wavelength. For a source spectral width of 2 Å, the rms width decreases from 0.1 ns/km to 0.006 ns/km between 500 nm and 1100 nm. The pulse width behavior shown in Fig. 80 reflects the combined effect of intermodal and intramodal pulse broadening.

3.6. Materials for optical fibers

The choice of materials to be used in the fabrication of fibers is influenced by the need to satisfy simultaneously many requirements. Obviously the material must be formable into a fine filament, transparent and available with two different refractive indices for core and cladding, respectively. These requirements alone more or less limit the field to plastics or glasses, although a liquid has been used to form the core of a fiber drawn from a hollow tube of glass. Many plastics are excluded from further consideration because the presence of hydrogen in their structures give rise to very high losses and because their molecular size leads to large scattering losses. And within the infinite number of possible glasses, most are rule out by other consideration. To appreciate this situation more fully, we need to examine in more detail the physical mechanisms involved, particularly those controlling optical loss since most optical communications systems require fibers of exceptionally low attenuation at the optical carrier frequency. Usually less than 20 dB/km is sought [309, 310].

Glasses are formed from fused mixtures of metal oxides, sulfides or selenides. Because they are fused mixtures rather than fixed compounds with crystal structures, their compositions are infinitely variable within certain regions of their respective phase diagrams, and large numbers of different glasses are manufactured by industry. Most of these fall into the category of oxide glasses, since these are the optically transparent ones, the sulfide and selenide glasses being used in the infrared region approximately 0.6 μm to 14 μm or more.

Of the oxide glasses, by far the most common are silica (SiO_2); sodium calcium silicates, frequently used for plate and window glass; sodium borosilicates, often

used for oven ware and chemical apparatus; and lead silicates, which are the crystal glasses having relatively high refractive indices and thus appearing "shiny". Typical starting materials for these glasses are sodium and calcium carbonates, boric oxide (B_2O_3) or boric acid, silica (sand), and lead oxide. For the typical fiber materials, new conditions arise, namely, the need for very low loss, which means very high chemical purity, so that the materials sources used are usually different from those serving large scale industry. The structure of glasses is noticeably different from that of the solids usually encountered in the electronic industry, namely, crystals. In the latter the individual atoms are well defined in space according to very precise and repeated patterns, lying in exact three-dimensional lattices. Glass, on other hand, consists of a loosely connected network formed by groups, which can be added to or modified by other components. For example, the addition of sodium (a network modifier) tends to break up the SiO_2 network, as shown in Fig. 81, in a sodium silicate glass. The result is less strongly bound than pure silica, and the melting temperature is thereby lowered. Since B_2O_3 is also a network former like SiO_2, a series of glasses such as sodium borates exist, paralleling sodium silicates but having very much lower melting temperatures since the B—O bond is much weaker than the Si—O bond.

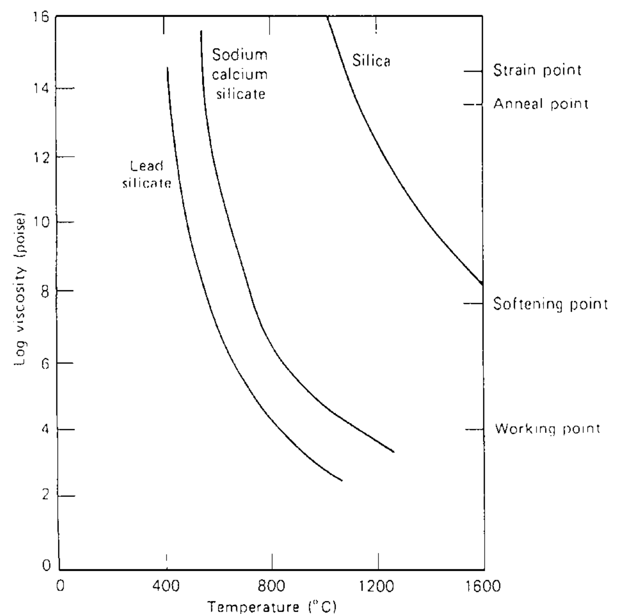


Figure 82 Viscosity versus temperature for a number of glasses (after [298]).

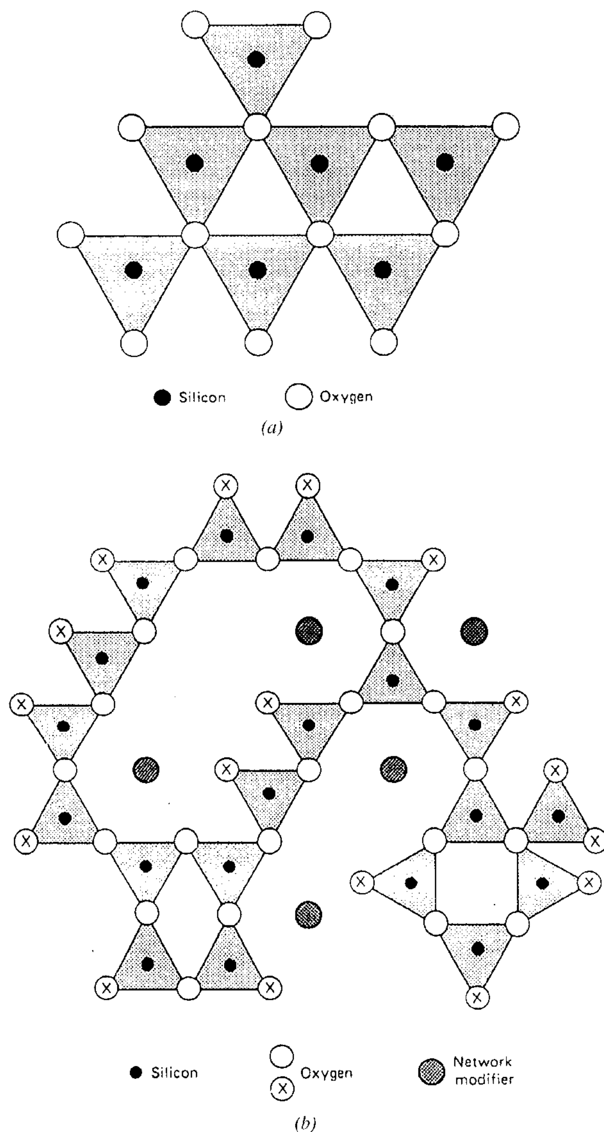


Figure 81 Networks involving SiO_2 groups, shown schematically in two dimensions. (a) A regular SiO_2 lattice (two-dimensional quartz crystal). (b) The effect on the lattice of the addition of a network modifier (after [298]).

Fig. 82 shows viscosity-temperature curves for some commonly manufactured glasses. Particularly noteworth is the fact that viscosity of a pure silica glass is much higher than that of multicomponent glasses at the same temperature. Just as the viscosity of a glass is a function of its composition, so too are the refractive index and the thermal expansion coefficient. Since there is an infinity of glass compositions, full data are not available for all glasses; in fact, comprehensive data are restricted to a few glass groups that have been extensively studied because of large scale commercial applications. The sodium calcium silicate (NCS) group is one such. A second glass system of great current interest is the sodium borosilicate (NBS) group of glasses. Another group of glasses that may be of interest in fiber manufacture is lead silicates. These are of potential interest since they allow large refractive index differences to be obtained between core and cladding glasses. For example, a pair with indices of 1.5 and 1.65 showing a 10% index difference can probably be obtained by using the highly polarizable lead ions to dope the higher index material. Fibers with such large index difference will be characterized by a higher level of Rayleigh scattering from the lead (of the order 10 dB/km, compared to 1 dB/km for silica at 900 nm) and have not yet been produced with very low loss (i.e., under 15 or 20 dB/km). But for many short link systems applications these disadvantages are more than offset by the large acceptance angle for such a high index difference fiber. All these three groups of glasses are of interest because they can be made by nearly conventional glass melting techniques in large quantities at low cost. They are formed by mixing the appropriate powders and fusing them in a crucible to

form a glass. But there is another group of materials, as already noted, all based upon pure silica with small additions of one or more other oxides.

3.6.1. Absorptive losses in glasses

Optical fiber choose to operate in one of two wavelength bands, the GaAs device region from 800 to 900 nm or the Nd laser region at 1.06 μm, with remote possibility of an extension to perhaps 1.2 or 1.3 μm using other semiconductor sources (see also [309, 310]) This choice is dictated by good physical reasons. Wavelengths shorter than about 600 to 650 nm are ruled out because Rayleigh scattering in the glass becomes severe and impurity absorption spilling over the band edge absorption is a problem. Beyond about 1.3 μm the first overtone of the OH stretching vibration appears at 1.4 μm, giving heavy absorption. Some spillover from multiphonon bands from the glass constituents probably adds to this, and detectors become far less efficient because the quantum energy $h\nu$ rapidly approaches kT under conditions at room temperature operation.

Five absorption mechanisms are of concern in a glass that is considered for use in optical fibers. The first two are associated with the basic glass constituents themselves, typically a combination of oxides of silicon, sodium. boron, calcium, germanium and so on. The glass has a band edge absorption somewhere in the ultraviolet (UV) region of the spectrum. Such absorption is extremely intense; and although the wavelength of interest for operation of a system is a considerable distance away, there has been serious debate as to whether the tail of the band edge could provide a significant loss mechanism (see also [311–316]).

Fig. 83 shows the losses of some fibers with very low water contents made with different constituents in the core by the chemical vapor deposition technique [317], and in Fig. 84 the infrared spectra of some of these same materials, showing the relative size and position of the infrared absorptions, are presented. The work suggests that GeO_2 doping for the core is the most favorable because of the longer wavelength at which the GeO_2 stretching vibration occurs, and this leads to an estimate [318] for the lowest loss for this type of fiber, at about 1.5 microns, of about 0.3 dB/km, as shown in Fig. 85. Evidently, to obtain figures even remotely approaching these, very low water contents must be achieved. The remaining absorption mechanisms are all related to impurities or defects in the glass [314] and are therefore not intrinsic. Besides that, a problem with most optical fibers is to reduce the hydroxyl ion (OH) content to a sufficiently low level, usually a few parts per million (details see [298]).

3.6.2. Rayleigh scattering

The phenomena of Rayleigh scattering is well-known to all of us as the mechanism responsible for blue sky. It is scattering of light from microirregularities in the dielectric medium through which the electromagnetic wave is propagating. The physical scale of the irregularities is of the order of one-tenth wavelength or less, so that each irregularity acts as a point source for scattered radiation. The resulting relation pattern from the induced dipole is doughnut shaped, being uniform in the plane perpendicular to the dipole and varying as $\sin\phi$ in the plane containing the dipole, where ϕ is the angle between the observation direction and the dipole axis. In the sky, Rayleigh scattering arises from the minute density fluctuation in the atmospheric gas caused by the constant thermal fluctuation of the medium. while the $\sin\phi$ component causes it to be highly polarized at

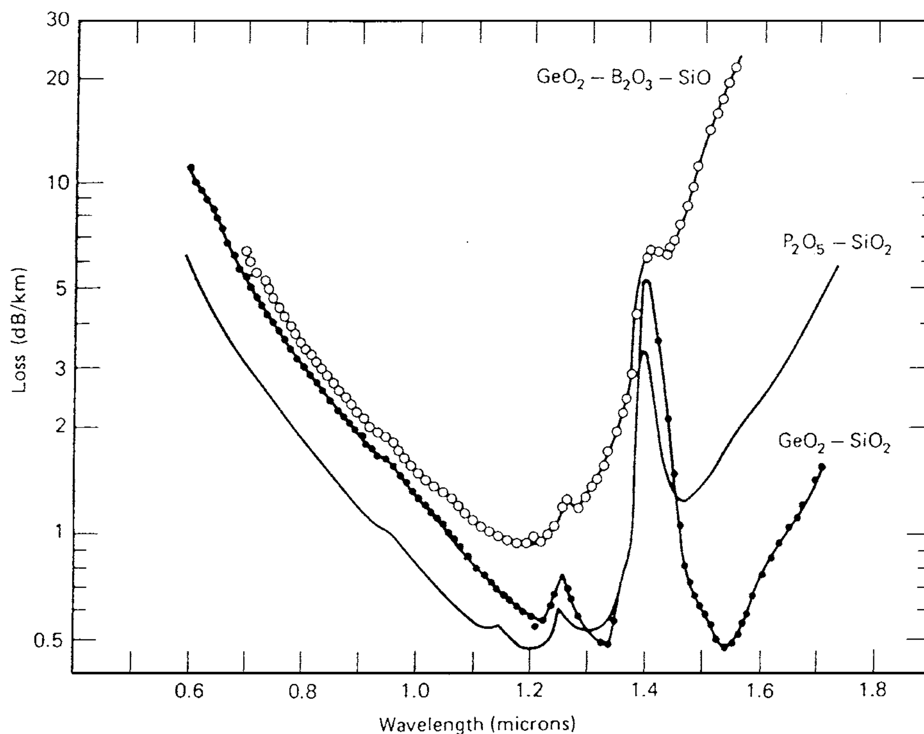


Figure 83 Losses versus wavelength of some low water content fibers showing the matrix influence on the infrared transmission (after [318]).

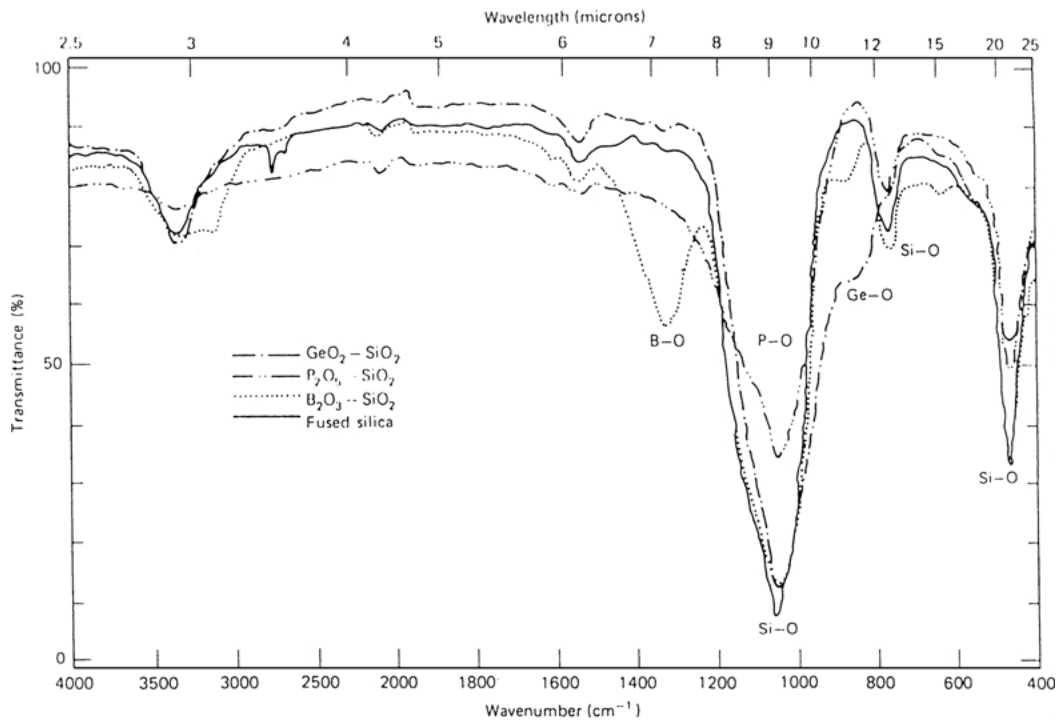


Figure 84 Infrared transmission spectra of the materials associated with the fibers of Fig. 83, showing the various absorption bands involved (after [318]).

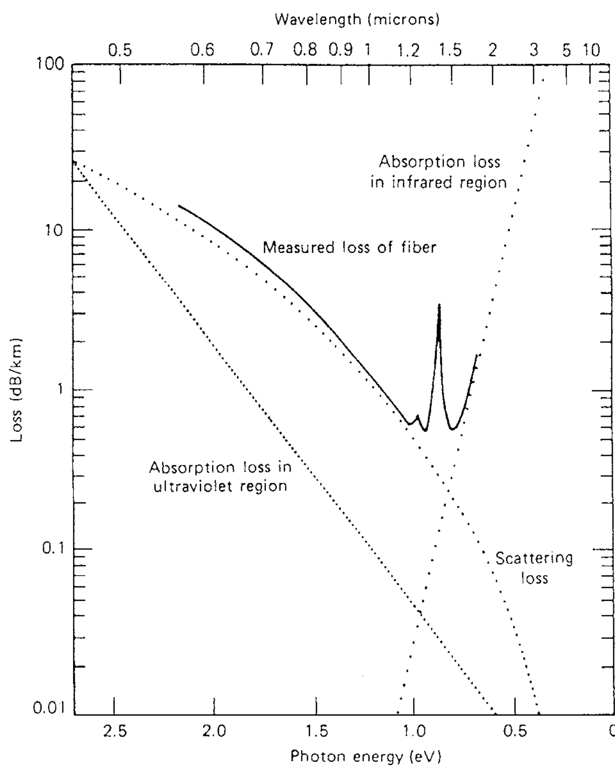


Figure 85 Predicted losses arising both the ultraviolet edge absorption and the infrared absorption for $GeO_2 - SiO_2$ fiber (after [317]).

ground level. In glass, Rayleigh scattering can arise from two separate effects, density and composition fluctuations.

We have already seen that glasses are disordered structures, loosely connected in a largely random sequence. Evidently, in such a structure there are local regions in which the average density is higher than in other regions. The dense and less dense regions can be traced back to thermally driven fluctuations in density arising from the Brownian movement of the liquid glass before it froze. The magnitude of the fluctuation is thus expected to be related to the freezing temperature—the higher the temperature, the greater the density fluctuation. One might therefore expect that a high melting material such as silica would show a higher Rayleigh scattering loss than a lead glass of lower melting temperature. But this is not so for another reason.

The magnitude of the density fluctuation scattering is given by the following expression [319]:

$$\alpha_{\text{scat.}\rho} = \frac{8\pi^3 n^3}{3\lambda^4}(n^8 p^2)(kT_{\text{f}})\beta_{\text{T}}. \quad (280a)$$

Here p is the photoelastic coefficient for the glass; T_{f} is the fictive temperature, defined as the temperature at which it becomes possible for the glass to reach a state of thermal equilibrium and closely related to the anneal temperature; k is Boltzmann's constant; and β_{T} is the isothermal compressibility. Certainly, the energy kT_{f} is the driving energy for the fluctuation giving rise to the loss mechanism described by $\alpha_{\text{scat.}\rho}$.

The expression for the scattering due to composition is more complex, taking the form [319]:

$$\alpha_{\text{scat,c}} = \frac{32\pi^3 n^2}{3\lambda^4 \rho N_{\text{A}}} \sum_{\text{j}=1}^{\text{m}} \left[\left(\frac{\partial n}{\partial x_{\text{j}}} \right)_{T_{\text{f}}, x_{\text{i}} \neq x_{\text{j}}} + \left(\frac{\partial n}{\partial \rho} \right)_{T_{\text{f}}, x_{\text{i}}} \left(\frac{\partial \rho}{\partial x_{\text{j}}} \right)_{P, T_{\text{f}}, x_{\text{i}} \neq x_{\text{j}}} \right]^2 M_{\text{j}} x_{\text{j}}. \quad (281)$$

In this formula, M_{j} and x_{j} are, respectively, the molecular weight and the weight fraction of the jth modifier, and N_{A} is Avagadro's number. The partial derivatives

of the refractive index and the density, ρ, may be determined from experimental data so that the summation can be performed. Schroeder [320] has derived an expression for the mean composition fluctuation in the glass, which takes the form:

$$\langle \Delta C^2 \rangle_{T_f} = \frac{\rho_0}{C_0}\left(\frac{kT_f}{V}\right)\left(\frac{\partial \mu}{\partial C}\right)^{-1}_{C_0, T_f}, \qquad (282)$$

where C is the mole fraction of one of the constituents in the composition C_0, V is the molar volume, and μ is the chemical potential difference between the major and minor constituents of th (binary)glass. Evidently, the composition fluctuation is minimized if $\partial\mu/\partial C$ is large and the fictive temperature is small.

Since the refractive index of a glass is a sensitive function of composition, it is frequently the case that the composition fluctuation a swamps the density fluctuation component. Thus a low melting glass may well have a much higher scatter loss than silica, and in general this is bound to be the case. However, some particular glass compositions have been found to lead to exceptionally low total Rayleigh scattering losses. The potassium silicates [320] are one such group which show a minimum scattering loss at one composition, as shown in Fig. 86. In Table XVIII it can list the Rayleigh scattering at 900 nm wavelength for a number of glasses that are of interest for optical fiber systems.

3.7. Fiber preparation

In this part we briefly consider the methods by which the different materials can be formed into fibers with the desired properties. Several techniques have been described for melting glasses that are specifically intended

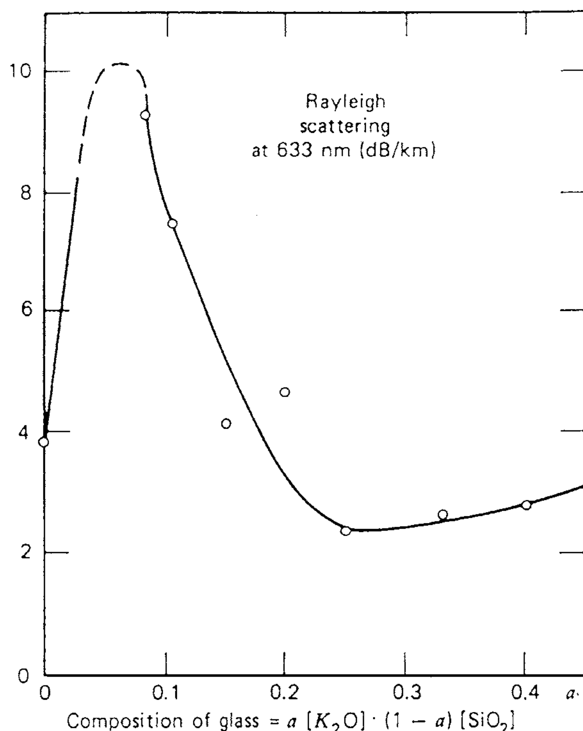


Figure 86 Rayleigh scattering loss at 633 nm versus composition for a sodium silicate glass (after [320]).

TABLE XVIII Rayleigh scattering losses for a number of fiber glasses (quoted for 900 nm) (after [298])

Glass	Loss (dB/km)
Silica	~1.2
Sodium borosilicate[a]	~1.5
Phosphosilicate[b]	~1.6
Germania silicate[c]	~1.2
Selfoc[d]	~2.0

[a]Ref. [298].
[b]D. B. Keck, R. D. Maurer and P. C. Schultz, Appl. Phys. Lett. 22, 307 (1973) (0.7% index difference from silica).
[c]M. Kawachi, A. Kawara and T. Miyashita, Electron. Lett. 13, 442 (1977) (0.18% index difference from silica).
[d]G. W. Morey, The Properties of Glass, 2nd ed. (Reinhold, New York, 1954).

for use in the preparation of optical fibers. Each is based upon the use of powders which are premixed, heated in a crucible until they fuse, and then agitated to produce a homogeneous mix. The heating may be applied to the crucible through black-body radiative coupling from the walls of an electrically heated furnace; it may be generated by the coupling of radio-frequency (RF) radiation to the crucible when the crucible is made of a conducting material such as platinum; or, alternatively, RF energy may be coupled directly to the melt glass, provided that is preheated to a temperature at which it begins to conduct. In the last case, the crucible remains relatively cold, thus helping to reduce problems of crucible contamination. In all cases it is necessary to use for the melting an enclosure which excludes contamination from the laboratory environment. Typically, silica liners are used to provide such isolation.

In the electric furnace, energy is transferred from the furnace wall to the crucible by black-body radiation. A great deal of early melting work was done using platinum crucibles contained within silica enclosures for clean-lines, but with the heat applied through RF energy coupled from an encircling coil, as shown in Fig. 87. This technique has the attraction that there is no refractory insulating anywhere near the melt [322]. However, precise temperature control is generally more difficult, the thermal time constant of the apparatus being much shorter, and the apparatus is far more expensive. It also requires a conducting crucible, ruling out the use of silica.

The production of fiber using a double-crucible apparatus is conceptually extremely simple [323, 324]. The basic apparatus is illustrated in Fig. 88. Two concentric crucibles are held with their axes vertical. Each crucible has in its base a central circular nozzle, the inner being carefully aligned to be concentric with the outer and perhaps 1 cm above it. The inner crucible is filled with core composition glass, and the outer with cladding composition glass. With the glasses molten, the core glass flows through the inner nozzle into the cladding stream and is subsequently carried out of the cladding outer nozzle surrounded by cladding glass, so that a composite glass flow is produced. The molten glass exudes into the space outside the outer crucible, and a filament is pulled from the exudant to form a cored glass fiber. In practice, the production of low loss fiber by this process requires a great deal of careful attention

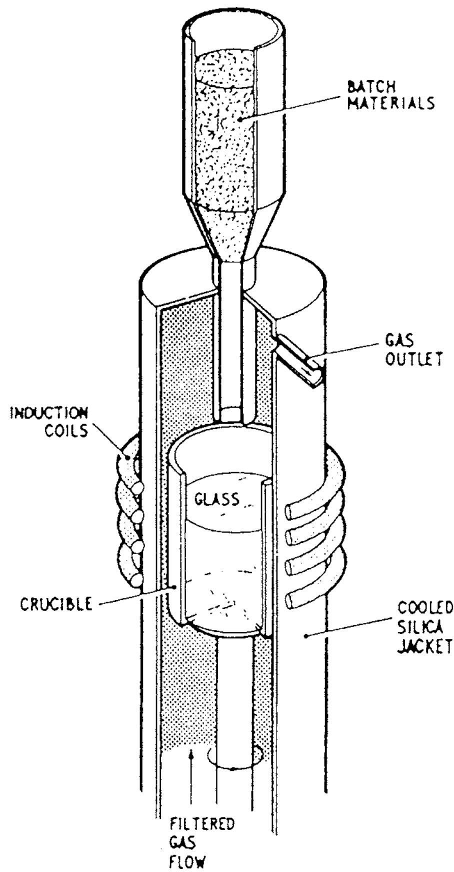


Figure 87 Radio frequency heating of a platinum crucible for a glass melting in a clean enclosure (after [321]).

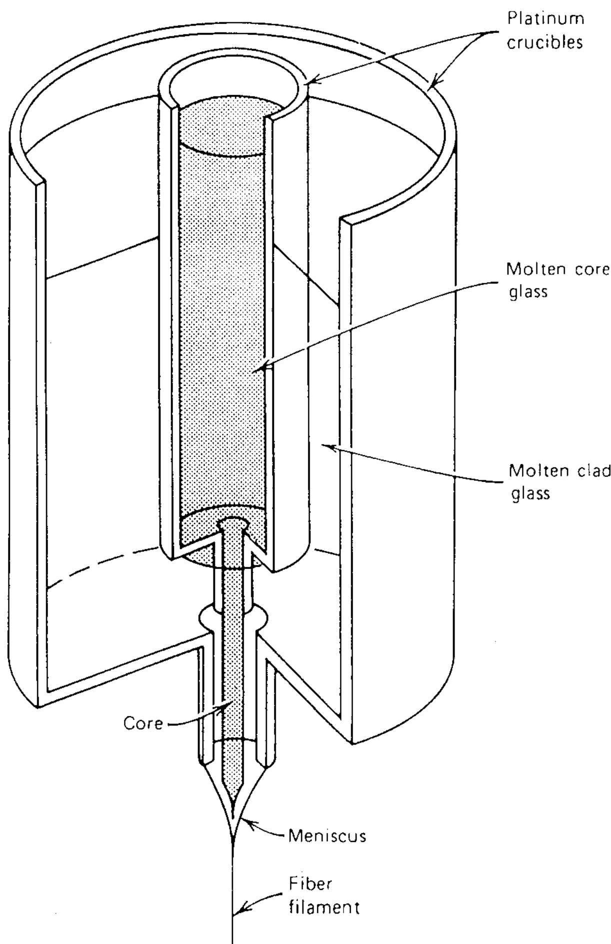


Figure 88 Apparatus for double crucible fiber pulling (after [321]).

to detail at every stage of the process; the glass preparation, cleaning of crucibles, assembly of apparatus, loading of the glass, atmosphere control, stabilization of temperature and so on.

We should note that increasing interest centers upon the deposition of more than one dopant for a given fiber and also upon depositing multiple layers, the concentration in each layer being carefully monitored and varied in a predetermined manner. In the former category a typical deposition might be of B_2O_3-SiO_2 to form a cladding and GeO_2-SiO_2 to form the core. Such a pair could be chosen to enable the viscosities of the two layers to be more equally matched, thus facilitating the collapse stage. Also, it is found that the presence of some of the dopants allows greatly increased gas flow and reaction rates to be used, a factor of considerable commercial importance in implementing a process.

The presence of layers arising from the deposition process is sometimes clearly evident in the preforms and fibers. This appears to be particularly true of CVD fiber and less so of outside deposition soor fiber. Interferometric analysis of the index profiles of preforms shows the presence of an "oscillating" index profile following a mean curve of the form desired, as illustrated schematically in Fig. 89. To conclusion of this part we pay attention to new class of fibers—the holey fibers, which have the other technology (details see [325] and references therein).

3.8. Isotopes in fibers

As a first example of possible major applications of isotopic engineering it will be considered isotopic fiber-optics and isotopic optoelectronics at large (see also [1, 100, 177]). It is known that for typical solids the lattice constant variations of isotopically different samples are usually within the limits

$$\frac{\Delta d}{d} \sim 10^{-3} \div 10^{-4}. \qquad (283)$$

Let us define an isotopic fiber as a structure in which core and cladding have the same chemical content but different isotopic composition. The boundary between different isotopic regions form an isotopic interface. The difference in the refractive index on both sides of the isotopic interface could lead to the possibility of total internal reflection of light and, consequently, could

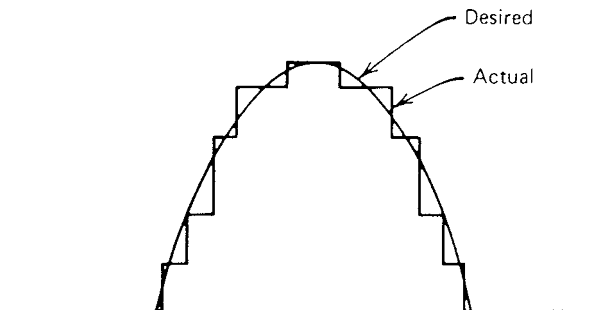


Figure 89 A schematic picture of the index distribution of a graded cor fiber produced by layer deposition (after [298]).

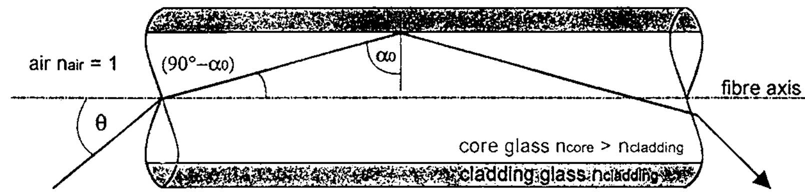


Figure 90 Light guidance in an optical waveguide by total internal reflection.

provide an alternative route to the confinement of light. For a quantitative estimate let us consider a boundary between SiO_2 (the main component of silica) where body sides are identical chemically and structurally but have a different isotopic content—e.g., $^{28}Si^{16}O_2$ and $^{30}Si^{18}O_2$ respectively (Fig. 90). In the first approximation the refractive index n is proportional to the number of light scatterers in the unit volume. From the Clausius-Mosotti relation for the refractive index one can deduce the following proportion (at $\Delta n \ll n$)

$$\frac{\Delta n}{n} \simeq 3c\frac{\Delta d}{d}, \qquad (284)$$

where c is a dimensionless adjustment factor of the order of unity. Substituting Equation 283 to Equation 284 we can obtain

$$\frac{\Delta n}{n} \sim 3 \times 10^{-3} \div 10^{-4}. \qquad (285)$$

Using the Snell law of light refraction we obtain the following expression for the ray bending angle Φ when the light travels through the refractive boundary

$$\theta \simeq \text{arc } \sin\left(\frac{n_1}{n_2}\sin\alpha_0\right), \qquad (286)$$

where α is the angle between the falling ray and the direction normal to the interface. For a sliding ray ($\alpha \simeq 90^\circ$), which is the control case for light confinement in fibers, the combining of Equations 285 and 286 leads to an estimate

$$\theta \sim 1.5 \div 4.5^\circ.$$

Thus, the isotopic fibers in which core and cladding are made of different isotopes the half-angle of the acceptance-cone could be up to several degrees. The resulting lattice mismatch and strains at the isotopic boundaries are correspondingly one part per few thousand [1] and, therefore, could be tolerated. Further advancement of this "isotopic option" could open the way for an essentially monolitic optical chip with built-in isotopic channels inside the fully integrated and chemically uniform structure.

Besides that we should pay attention to the fact that composition (different isotopes) fluctuation are subject to the restoring force of the total free energy of the glass system which will also seek to minimize itself. Usin g isotope pure materials for core and cladding we should receive significant less Rayleigh scattering (see e.g. [4]).

Chapter 4. Laser materials

4.1. Some general remarks

As is well-known, the word laser is an acronym for "light amplification by the stimulated emission of radiation", a phrase which covers most, though not all, of the key physical processes inside a laser. Unfortunately, that concise definition may not be very enlightening to the nonspecialist who wants to use a laser but has less concern about the internal physics than the external characteristics. A general knowledge of laser physics is as helpful to the laser user as a general understanding of semiconductor physics is to the circuit designer. From a practical standpoint, a laser can be considered as a source of a narrow beam of monochromatic, coherent light in the visible, infrared or UV parts of spectrum. The power in a continuous beam can range from a fraction of a milliwatt to around 20 kilowatts (kW) in commercial lasers, and up to more than a megawatt in special military lasers. Pulsed lasers can deliver much higher peak powers during a pulse, although the average power levels (including intervals while the laser is off and on) are comparable to those of continuous lasers.

The range of laser devices is broad. The laser medium, or material emitting the laser beam, can be a gas, liquid, crystalline solid, or semiconductor crystal and can range in size from a grain of salt to filling the inside of a moderate-sized building. Not every laser produces a narrow beam of monochromatic, coherent light. A typical laser beam has a divergence angle of around a milliradian, meaning that it will spread to one meter in diameter after traveling a kilometer. This figure can vary widely depending on the type of laser and the optics used with it, but in any case it serves to concentrate the output power onto a small area. Semiconductor diode lasers, for example, produce beams that spread out over an angle of 20 to 40° hardly a pencil-thin beam. Liquid dye lasers emit at a range of wavelengths broad or narrow depending on the optics used with them. Other types emit at a number of spectral lines, producing light is neither truly monochromatic nor coherent. Practically speaking, lasers contain three key elements. One is the laser medium itself, which generates the laser light. A second is the power supply, which delivers energy to the laser medium in the form needed to excite it to emit light. The third is the optical cavity or resonator, which concentrates the light to stimulate the emission of laser radiation. All three elements can take various forms, and although they are not always immediately evident in all types of lasers, their functions are essential. Like most other light sources, lasers are inefficient in converting input energy into light; efficiencies can range from under 0.01 to around 20% [326–329].

4.2. Absorption and induced emission

The idea of stimulated emission of radiation, as well-known, originated with Albert Einstein [330]. Until that time, physicists had believed that a photon could interact with an atom in only two ways: it could be absorbed and raise the atom to a higher energy level or be emitted as the atom dropped to a lower energy level. Einstein proposed a third possibility-that a photon with energy corresponding to that of an energy—level transition could stimulate an atom in the upper level to drop to the lower level, in the process stimulating the emission of another photon with the same energy as first. In the normal world, stimulated emission is unlikely because at thermodynamic equilibrium more atoms are in lower energy levels than in higher ones. Thus a photon is much more likely to encounter an atom in a lower level and be absorbed than to encounter one in a higher level and stimulate emission.

Below we consider a cavity whose walls are at temperature T, containing radiation and an ensemble of atoms and let each atom be represented by a two-level quantum mechanical system with an energy level separation of $\hbar\omega_0$. In thermal equilibrium the energy density per unit angular frequency range at ω_0 is given by [331]

$$\rho_{\omega_0} = \frac{\hbar\omega_0^3}{\pi^2 c^3}\frac{n^3}{e^{\hbar\omega_0/\mathrm{kT}} - 1} = \frac{\omega_0^2}{\pi^2 c^3}\frac{\hbar\omega_0 n^3}{e^{\hbar\omega_0/\mathrm{kT}} - 1}. \quad (287)$$

Then

$$\rho_{\omega_0}(e^{\hbar\omega_0/\mathrm{kT}} - 1) = \frac{\hbar\omega_0^3 n^3}{\pi^2 c^3}. \quad (288)$$

In addition, because of detailed balance

$$AN_2^e + B_{21}\rho_{\omega_0}N_2^e = B_{12}\rho_{\omega_0}N_1^e, \quad (289)$$

where N_1^e and N_2^e are the equilibrium populations of atoms in the lower and upper levels, respectively, and $B_{21}\rho_{\omega_0}$ and $B_{12}\rho_{\omega_0}$ are the probabilities per unit time for induced downward and upward transitions, respectively and A is Einstein coefficient.

Now we can write

$$A\frac{N_2^e}{N_1^e} + B_{21}\rho_{\omega_0}\frac{N_2^e}{N_1^e} = B_{12}\rho_{\omega_0} \quad (290)$$

and

$$Ae^{-\frac{\hbar\omega_0}{\mathrm{kT}}} + B_{21}\rho_{\omega_0}e^{-\frac{\hbar\omega_0}{\mathrm{kT}}} = B_{12}\rho_{\omega_0} \quad (291)$$

or

$$\rho_{\omega_0}\big(B_{12}e^{-\frac{\hbar\omega_0}{\mathrm{kT}}} - B_{21}\big) = A. \quad (292)$$

Set

$$B_{12} = B_{21} = B. \quad (293)$$

$$\rho_{\omega_0}\big(e^{-\frac{\hbar\omega_0}{\mathrm{kT}}} - 1\big)B = A. \quad (294)$$

$$\frac{A}{B} = \rho_{\omega_0}\big(e^{-\frac{\hbar\omega_0}{\mathrm{kT}}} - 1\big) = \frac{\hbar\omega_0^3 n^3}{\pi^2 c^3} \quad (295)$$

$$B = \frac{\pi^2 c^3}{\hbar\omega_0^3 n^3}A = \frac{\pi^2 c^3}{n^3\hbar\omega_0^3}\frac{4n\omega_0^3|\mu|^2}{3\hbar c^3} = \frac{4\pi^2}{3\hbar^2 n^2}|\mu|^2 \quad (296)$$

or

$$A = \frac{4n\omega_0^3}{3\hbar c^3}|\mu|^2\,(\mathrm{s}^{-1});$$

$$B = \frac{4\pi^2}{3\hbar^2 n^2}|\mu|^2\,(\mathrm{cm}^3\mathrm{erg}^{-1}\ \mathrm{s}^{-2}). \quad (297)$$

$$\frac{A}{B} = \frac{\hbar\omega_0^3 n^3}{\pi^2 c^3}\,(\mathrm{erg}\cdot\mathrm{s}\cdot\mathrm{cm}^{-3}) \quad (298)$$

Let us call $w(\omega)\mathrm{d}\omega$ the probability per unit time that an atom undergoes an induced transition by absorbing or emitting a photon with angular frequency in $(\omega, \omega + \mathrm{d}\omega)$. It is

$$w(\omega)\mathrm{d}\omega = \mathrm{B}g(\omega)\rho_\omega\mathrm{d}\omega = \frac{4\pi^2}{3\hbar^2 n^2}|\mu|^2\rho_\omega g(\omega)\mathrm{d}\omega = \frac{4\pi^2}{3nc\hbar^2}|\mu|^2 I(\omega)g(\omega)\mathrm{d}\omega, \quad (299)$$

where

$$I(\omega)\mathrm{d}\omega = \rho_\omega\frac{c}{n}\mathrm{d}\omega = \text{intensity of radiation with angular frequency in } (\omega, \omega + \mathrm{d}\omega). \quad (300)$$

Now we can write

$$\int w(\omega)\mathrm{d}\omega = \frac{4\pi^2}{3\hbar^2 n^2}|\mu|^2\rho_\omega = \frac{4\pi^2}{3\hbar^2 nc}|\mu|^2 I(\omega_0) \quad (301)$$

Further let us assume that a plane wave goes through a certain medium in the x-direction. Let the medium consist of atoms which have two possible energy levels and let $N_1(N_2)$ be the concentration of atoms in the lower (higher) energy level (see also [327])

$$\mathrm{d}I(\omega) = -w(\omega)(N_1 - N_2)\hbar\omega\mathrm{d}x\left(\frac{\mathrm{erg}}{\mathrm{cm}^2}\right) \quad (302)$$

But from (301) we have

$$w(\omega) = \frac{4\pi^2}{3nc\hbar^2}|\mu|^2 I(\omega)g(\omega). \quad (303)$$

Then

$$\mathrm{d}I(\omega) = -\left[\frac{4\pi^2}{3nc\hbar^2}|\mu|^2 I(\omega)g(\omega)\right](N_1 - N_2)\hbar\omega\mathrm{d}x = -\left[\frac{4\pi^2(N_1 - N_2)}{3nc\hbar^2}|\mu|^2 I(\omega)g(\omega)\right]I(\omega)\mathrm{d}x = -\alpha(\omega)I(\omega)\mathrm{d}x, \quad (304)$$

where

$$\alpha(\omega) = \frac{4\pi^2(N_1 - N_2)}{3nc\hbar^2}|\mu|^2\omega g(\omega)$$
$$= \frac{n\hbar\omega}{c}B(\omega)(N_1 - N_2).$$

(absorption coefficient.) (305)

The solution of (304) is,

$$I(\omega, x) = I(\omega, x = 0)e^{-\alpha x}. \quad (306)$$

We define the sbsorption cross-section of the radiative transition as follows (details see also [328])

$$\sigma(\omega) = \frac{\alpha(\omega)}{N_1 - N_2} = \frac{4\pi^2}{3nc\hbar^2}|\mu|^2\omega g(\omega)\ (\mathrm{cm}^2) \quad (307)$$

Note the following

$$\int \alpha(\omega)\mathrm{d}\omega = \frac{4\pi^2(N_1 - N_2)}{3nc\hbar^2}|\mu|^2\omega_0$$
$$= \frac{n\hbar\omega_0}{c}B(N_1 - N_2)(\mathrm{cm}^{-1}\,\mathrm{s}^{-1}). \quad (308)$$

4.3. Semiconductor lasers

4.3.1. Heterojunction lasers

Semiconductor lasers, like other lasers, have population inversions which lead to stimulated emission of photons. Semiconductor laser is different from another lasers primarily because the energy levels in semiconductors must be treated as continuous distributions of levels rather than as discrete levels [332]. We shall assume that the densities of states in the conduction and valence bands of the semiconductor are known functions of energy, and that the occupations of these levels are characterized by quasi-Fermi levels [257]. Then the probability that the state of energy E in the conduction band is occupied by an electron is

$$f_c(E) = 1/\{1 + \exp[(E - F_c)/kT]\}, \quad (309)$$

where F_c is the quasi-Fermi level for the conduction band, k is Boltzmann's constant, and T is absolute temperature. A corresponding expression applies in the valence band, with quasi-Fermi level F_v. For a system in thermal equilibrium, the quasi-Fermi levels are equal to each and become the Fermi level E_F. In an excited system we have $F_c \succ F_v$, and we can use the separation of the quasi-Fermi levels as a measure of the excitation. The use of quasi-Fermi levels greatly simplifies the treatment of systems with many energy levels, or with continuous distributions of levels, because one quantity represents the occupation probability of many levels. The concept of quasi-Fermi level in an excited system is valid provided the scattering rate of carriers within a band is rapid compared to the recombination rate between bands, i.e., if the carriers within the conduction band and within the valence band rapidly establish a quasi-equilibrium among themselves although the conduction band and the valence band are not in equilibrium with each other For semiconductors with substantial numbers of free carriers, carrier-carrier scattering will lead to the establishment of quasi-equilibrium (see e.g. [333, 334]).

The original semiconductor lasers were p-n junctions prepared by diffusion of acceptor impurities into n-type GaAs, and this is still one of the most common structures. Semiconductors with $\vec{k}$—conserving transitions at the energy gap are strongly favored for lasing [335], but some impurity levels can lead to stimulated emission in indirect-gap semiconductors, e.g. [336]. All the p-n junction lasers are excited by passing current through the p-n junction, and the excitation rate is characterized by the current density. When a forward current flows, electrons are injected into the p-type material and holes are injected into the n-type material, te latter to a much smaller extent partly because of the lower hole mobility. In heterojunctions, potential barriers play a major role in the injection of carriers [337]. The excess of electron and hole concentrations over their equilibrium values creates a local population inversion and leads to stimulated emission of photons at sufficiently high excitation levels. The layer near the p-n junction where this occurs is called the active region or active layer of the device. Fig. 91 shows that the effective thickness of the active layer in graded junction lasers increases as the current density increases. This leads to smaller quasi-Fermi level separations and to less efficient use of the excitation for lasing. Heterostructure lasers (see also [337]) contain built-in potential barriers for the electrons which tend to confine them to regions of fixed width. Thus the excitation can be used most effectively.

A second class of excitation methods involves excitation with photons [339] or with an electron beam [338]. For optical excitation, the active layer thickness will be of the order of $1/\alpha$, where α is the absorption coefficient of the incident photons. For electron beam excitation, the active layer thickness will be of the order of the penetration depth of the electrons, which is a function of their energy (see, e.g. [340]). In both cases, diffusion of carriers will add a distance of the order of the diffusion length to the thickness given. If sufficiently thin samples are used, the excitation state may be relatively uniform through the sample, provided surface recombination is unimportant. The excitation rate for the externally excited structure can be converted to an equivalent current density. For photons, the absorbed photon flux is multiplied by the electronic charge provided that each absorbed photon give rise to an electron-hole pair. It should be added that values of about $3E_g$ are necessary to create one *e-h* pair [339]. This means that the main part of the incident pump energy is transferred into heat. This is one of the disadvantages of *e*-beam pumping.

The value of the gain coefficient g_{th} at the lasing threshold is given by a very simple calculation (see also [338]). If the laser cavity is bounded by parallel surfaces with reflectivities R_1 and R_2, separated by a distance L, then the amplification of the photon intensity on a

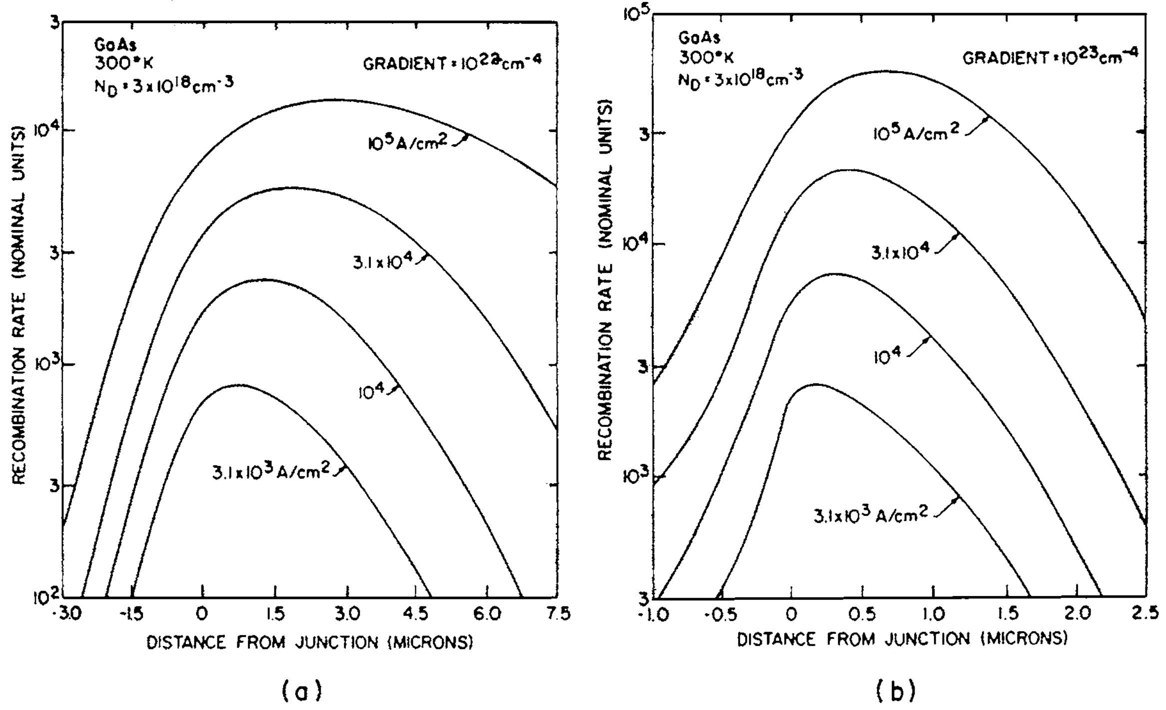


Figure 91 Spatial distribution of the recombination rate for graded GaAs p-n junction lasers with the indicated forward current densities at 300 K. The donor concentration is 3×10^{18} cm^{-3}, and the acceptor gradients at the junction are (a) 10^{22} cm^{-4} and (b) 10^{23} cm^{-4}. The ordinate is the nominal current density defined in Equation 317, in A/cm^2, equivalent to 6.2×10^{22} recombination/cm^{3}s. Note the increase in effective active layer thickness as the current density increases, particularly in (a). (after [338]).

round trip through the device is $R_1R_2\exp[2(g - \alpha)L]$, where α is an effective absorption coefficient, e.g., for free-carrier absorption. The lasing condition is that this amplification factor be unity. Thus

$$g_{th} = \alpha + \left(\frac{1}{L}\right)\log\left(\frac{1}{R}\right), \qquad (310)$$

where $R = (R_1R_2)^{1/2}$. The current density for which lasing occurs, i.e., the lowest current density for which g reaches g_{th} at any photon energy, is the threshold current density J_{th}, and is one of the principal characteristics of a semiconductor laser.

The rate at which photons are emitted per unit volume can be written as (see, e.g., [341])

$$r(E)\mathrm{d}E = [r_{spon}(E) + Nr_{stim}(E)]\mathrm{d}E, \qquad (311)$$

where N is the average number of photons per mode, given for thermal equilibrium by

$$N(E) = \frac{1}{\exp(E/kT) - 1}. \qquad (312)$$

The term r_{spon} (311) gives the rate of spontaneous downward transitions of the electronic system, and r_{stim} is the difference between the stimulated rates of downward and upward transitions.

The spontaneous and stimulated functions can be written

$$r_{spon}(E) = a\int \rho_c(E_u)\rho_v(E_l)|M|^2 f_u(1 - f_l)\mathrm{d}E_u, \qquad (313a)$$

$$r_{stim}(E) = a\int \rho_c(E_u)\rho_v(E_l)|M|^2(f_u - f_l)\mathrm{d}E_u, \qquad (313b)$$

$$a = 4Ne^2E/m^2\hbar^2c^3, \qquad (313c)$$

where f_u and f_l are the probabilities that the upper and lower states involved in the transition are occupied, N is the index of refraction, and $E_l = E - E_g - E_u$ with the convention for measuring energies from nominal band edges separated by the nominal energy gap E_g. The squared matrix element $|M|^2$ is averaged over all polarizations of the radiation, and is a function of both E_u and E_l. Equation 313 ignores Coulomb interaction between electron and hole (see below), which should be relatively weak when large carrier concentrations are present. Lasing in materials for which exciton effects are important has been treated by Haug [342].

When the populations are characterized by quasi-Fermi levels as in Equation 309, the relation between r_{stim} and r_{spon} is

$$r_{stim}(E) = r_{spon}\{1 - \exp[(E - \Delta F)/kT], \qquad (314)$$

where $\Delta F = F_c - F_v$ is the difference of the quasi-Fermi levels for the conduction band and the valence band. The relation between the local gain coefficient g_{local} and r_{stim} is

$$g_{local}(E) = -\alpha(E) = (\pi c\hbar^{3/2}/NE)^2 r_{stim}(E), \qquad (315)$$

where the minus sign arises because we have defined r_{stim} and g_{local} to be positive when radiation is emitted, while the absorption coefficient α is positive when

radiation is absorbed. Combine (314) and (315) we obtain

$$r_{\text{spon}}(E) = (NE/\pi c\hbar^{3/2})^2 \times \alpha(E/\{\exp[(E-\Delta F)]/kT) - 1\}. \quad (316)$$

in thermal equilibrium, when $\Delta F = 0$, this reduces to the well-known result of Van Roosbroeck and Shokley [343]. For semiconductor lasers the term in ΔF is extremely important. Equation 314 shows that there can be gain only for photon energies $E \prec \Delta F$.

The local gain coefficient is a function of photon energy, but for most laser cases we are interested primarily in its peak value $g_{\max}$. In Fig. 92 it plots $g_{\max}$ vs. excitation rate, which is characterized for this purpose by the nominal current density

$$J_{\text{nom}} = 10^{-4} e \int r_{\text{spon}}(E)\mathrm{d}E, \quad (317)$$

the current density required to maintain the actual excitation rate in a homogeneous layer 1 μm thick. The

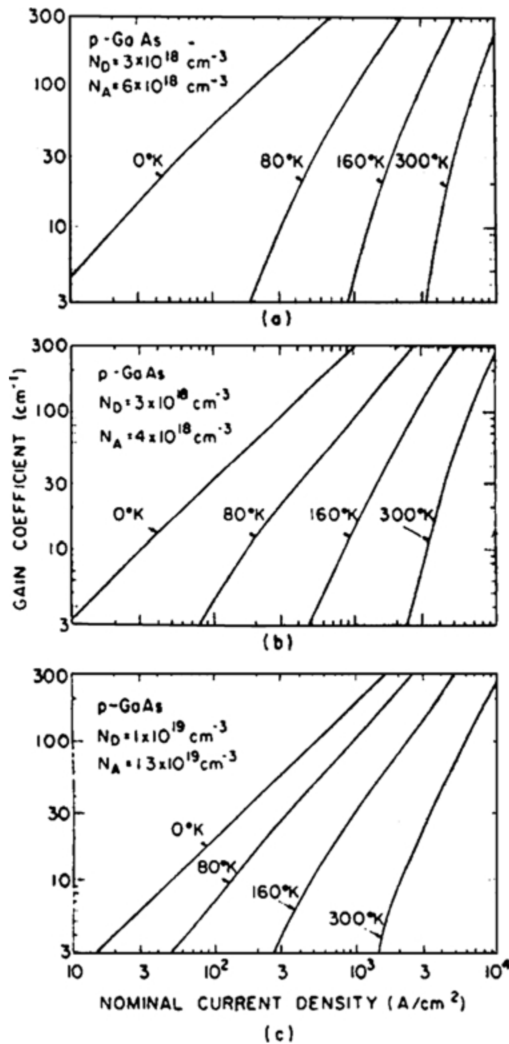


Figure 92 Maximum local gain coefficient (the value of $g_{\text{local}}(E)$ at the photon energy with maximum gain) vs nominal current density, defined in Equation 317, for various temperature and for (a) $N_D = 3 \times 10^{18}$, $N_A = 6 \times 10^{18}$ cm^{-3}; (b) $N_D = 3 \times 10^{18}$, $N_A = 4 \times 10^{18}$ cm^{-3}; (c) $N_D = 10^{19}$, $N_A = 1.3 \times 10^{19}$ cm^{-3}. The ion screening temperature is 30,000 K (after [338]).

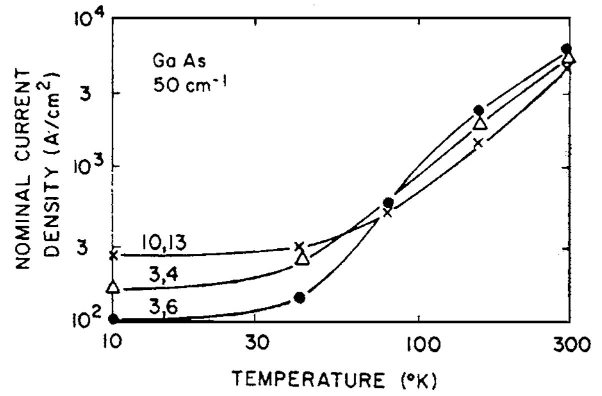


Figure 93 Temperature dependence of the nominal current density required to reach a gain coefficient $g_{\max} = 50$ cm^{-1}, taken from results of Fig. 92. The curves are labeled by the values of donor and acceptor concentration, respectively, in units of 10^{18} cm^{-3}. The temperature dependence of the threshold current density is also a function of the temperature dependence of the losses and of the effective active layer thickness (after [338]).

curves of Fig. 92 are given for several sets of values of donor and acceptor concentrations, and for several temperatures from absolute zero to 300 K. Note that increasing the compensation, i.e. increasing $N_D + N_A$ for a given $|N_D + N_A|$, raises the nominal current density required to reach a specified gain at low temperature, but lowers it at room temperature (cf. Figs 92a–c). Fig. 93 shows the nominal current density required to reach a gain of 50 cm^{-1} for each case in Fig. 92. It is an indication of the temperature dependence of the threshold current density for typical semiconductor junction lasers but makes no allowance for the effect of the structure or for losses such as free carrier absorption. A similar analysis has also been made by Hwang [344], whose curves of gain versus current density are similar to those given in Fig. 93. Both calculations agree well wth experiment. The photon energy at which lasing occurs is found by Hwang [344] to be in good agreement with experimental results (details see [338, 339, 342]).

4.3.2. Study of excitons lasing

With increasing excitation intensity, frequently laser action is observed in the excitonic luminescence. However, the direct recombination of an exciton can never give rise to laser action, because the coupled exciton-photon system corresponds in the resonant approximation to two linearly coupled harmonic oscillators. The equations of motion of this system do not contain the nonlinearity which is necessary to describe laser action. The participation of a third field is required in order to introduce the possibility of laser action [342], i.e., the laser action in exciton systems has to be a parametric process in which a pump field, a signal field and an idler field participate.

Below we describe a scheme for lasing action involving excitons in a pure crystal. In Refs. [345, 346] a general theory of various spontaneous and stimulated exciton recombination processes has been developed. In our analysis we assume a two-band model for the crystal. Atoms at lattice sites interact with each other and an electromagnetic field. The analysis is addressed

to tight-binding excitons in the first step. According to [346] the weak-binding case of Wannier-Mott excitons follows a similar description but is more involved. In next section we follow very close the results of Liu and Liboff [346].

The Hamiltonian of the entire system of crystal and electromagnetic field is [346]

$$H = H_{\text{xtal}} + H_{\text{em}} + H_{\text{int}} \tag{318}$$

The Hamiltonian of the crystal can be written in the Wannier representation

$$H_{\text{xtal}} = \sum_{\rho} E_{\text{e}} c_{\rho}^{\dagger} c_{\rho} + \sum_{\rho} E_{\text{h}} d_{\rho}^{\dagger} d_{\rho} + \hbar \sum_{\rho_1,\rho_2} W(\rho_1 - \rho_2) c_{\rho_1}^{\dagger} c_{\rho_1} d_{\rho_2}^{\dagger} d_{\rho_2}. \tag{319}$$

In this expression c_ρ and d_ρ operators refer, respectively to electrons and holes at site ρ. Their respective energies are E_{e} and E_{h}. Indicated operators obey the anticommutation relations:

$$\begin{aligned} \{c_\rho, c_{\rho'}^{\dagger}\} &= \{d_\rho, d_{\rho'}^{\dagger}\} = \delta_{\rho,\rho'}, \\ \{c_\rho, d_{\rho'}\} &= \{c_\rho, d_{\rho'}^{\dagger}\} = \{c_{\rho'}^{\dagger}, d_{\rho'}\} \\ &= \{c_{\rho}^{\dagger}, d_{\rho'}^{\dagger}\} = 0. \end{aligned} \tag{320}$$

The function in Equation 319 represents the interaction between electrons and holes and has the integral representation

$$\hbar W(\rho_1 - \rho_2) = \iint \mathrm{d}\vec{r}\,\mathrm{d}\vec{r}'\, w_{\text{c}}^{*}(\vec{r} - \vec{\rho}_2) w_{\text{v}}^{*}(\vec{r}' - \vec{\rho}_2) \times \frac{e^2}{|\vec{r} - \vec{r}'|} w_{\text{c}}(\vec{r}' - \vec{\rho}_2) w_{\text{v}}(\vec{r} - \vec{\rho}_1), \tag{321}$$

where $w_{\text{c}}(\vec{r} - \vec{\rho})$ and $w_{\text{v}}(\vec{r} - \vec{\rho})$ are Wannier functions relevant to the conduction and valence bands, respectively.

The Hamiltonian H_{xtal} may be diagonalized as follows (see also [347]). For this purpose we define the operators

$$B_{\rho}^{\dagger} = c_{\rho}^{\dagger} d_{\rho}^{\dagger}, \tag{322a}$$

$$B_{\rho} = d_{\rho} c_{\rho}. \tag{322b}$$

The operator $B_{\rho}^{\dagger}$ creates an electron-hole pair whereas B_{ρ} annihilates the pair. The operator

$$\mu_\rho = \frac{1}{2}(c_{\rho}^{\dagger} c_\rho + d_{\rho}^{\dagger} d_\rho) \tag{323}$$

indicates if the atom at the site ρ is excited or not. Its eigenvalues are (1, 0). Combining the latter three equations gives

$$[B_{\rho'}, B_{\rho}^{\dagger}] = \delta_{\rho,\rho'}(1 - 2\mu_\rho). \tag{324}$$

We further define the operator

$$b_{\text{q}}^{\dagger} = \frac{1}{\sqrt{N}} \sum_{\rho} e^{\text{i}\text{q}\rho} B_{\rho}^{\dagger}, \tag{325}$$

which represents the creation of an exciton in the propagating state $\vec{q}$. The total number of atoms in the crystal is N. The commutation relation for exciton operators the appears as [342]

$$[b_{\text{q}}, b_{\text{q}'}^{\dagger}] = \delta_{q,q'} - \frac{2}{N} \sum_{\rho} e^{\text{i}(\text{q}-\text{q}')} \mu_\rho. \tag{326}$$

In this manner wereach the important result that exciton operators are boson operators only when the number of excited atoms in the crystal is small compared to the total number of atoms N. In this even, the second term of Equation 326 can be neglected, giving $[b_{\text{q}}, b_{\text{q}'}^{\dagger}] = \delta_{\text{q},q'}$. The transformation Equation 325 then diagonalizes H_{xtal} as

$$H_{\text{xtal}} = \sum_{\text{q}} E_{\text{q}} b_{\text{q}}^{\dagger} b_{\text{q}}, \tag{327a}$$

where

$$E_{\text{q}} = E_{\text{e}} + E_{\text{h}} + \epsilon_{\text{q}}, \tag{327b}$$

and

$$\epsilon_{\text{q}} = \hbar \sum_{\rho} W(\rho) e^{-\text{i}\text{q}\rho}. \tag{327c}$$

The diagonalization Equation 327a demonstrate that excitons are collective excitations of the entire crystal.

In the limit of intense lasing, cited authors [346] replaced the exciton and photon amplitudes with their classical counterparts (see also [348–350]:

$$b_{\text{q}}^{\dagger} = \beta_{\text{q}} \exp\left[i\left(\omega_{\text{q}}^{0} + \Omega_1 + i\eta\right)t\right], \tag{328}$$

$$a_{\text{j}}^{\dagger} = a_{\text{j}}\left[i\left(\omega_{\text{j}}^{0} + \Omega_2 + i\eta\right)t\right], \tag{329}$$

where $\hbar\omega_{\text{q}}^{0} = E_q$. The operator $a_{\text{j}}^{\dagger}$ creates a photon of frequency ω_{j} and γ is the decay rate of this mode. The coefficient η is introduced to describe the decay of excitons. The frequency Ω_1 and Ω_2 remain arbitrary. Equations for $a_{\text{j}}^{\dagger}$ and β_{q} have following expressions:

$$(i\Omega_1 - \eta)\beta_{\text{q}} = -\frac{i}{\sqrt{N}} \sum_{\rho} G_\rho \sigma_\rho a_{\text{j}}^{\dagger} e^{\text{i}\text{q}\rho}, \tag{330}$$

$$(i\Omega_2 - \gamma)a_j^{\dagger} = \frac{i}{\sqrt{N}} \sum_{\rho} \sum_{\text{q}} G_{\rho}^{*} \beta_{\text{q}} e^{-\text{i}\text{q}\rho}. \tag{331}$$

Substituting Equation 330 into 331, and take into account of the resonance condition, we have

$$\omega_{\text{q}}^{0} + \Omega_1 = \omega_{\text{j}}^{0} + \Omega_2, \tag{332}$$

and equating the real and imaginary parts of the resulting equation gives

$$\gamma = \frac{\eta}{\eta^2 + \Omega_1^2} \sum_\rho |G_\rho|^2 \sigma_\rho, \tag{333}$$

$$\Omega_2 = -\frac{\Omega_1}{\eta^2 + \Omega_1^2} \sum_\rho |G_\rho|^2 \sigma_\rho. \tag{334}$$

It follows that

$$\frac{\Omega_1}{\Omega_2} = -\frac{\eta}{\gamma}. \tag{335}$$

While the difference frequency

$$\Delta \equiv \Omega_1 - \Omega_2, \tag{336}$$

Equation 560 yields

$$\Omega_1 = \frac{\Delta}{1 + \frac{\gamma}{\eta}}, \quad \Omega_2 = -\frac{\Delta}{1 + \frac{\gamma}{\eta}}. \tag{337}$$

Substituting the exponential forms (322) and (329) into next equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\sigma_\rho = \frac{1}{\tau}(s_0 - \sigma_\rho) + \frac{2i}{\sqrt{N}}\left[\sum_{\mathrm{q}} G_\rho b_{\mathrm{q}} a_{\mathrm{j}}^{\dagger} e^{\mathrm{iq}\rho} - \sum_{\mathrm{q}} G_\rho^* b_{\mathrm{q}}^{\dagger} a_{\mathrm{j}} e^{-\mathrm{iq}\rho}\right] \tag{338}$$

in steady state we find

$$0 = \frac{1}{\tau}(s_0 - \sigma_\rho) + \frac{2i}{\sqrt{N}}\left[\sum_{\mathrm{q}} G_\rho b_{\mathrm{q}}^* a_{\mathrm{j}}^* e^{\mathrm{iq}\rho} - \sum_{\mathrm{q}} G_\rho^* b_{\mathrm{q}} a_{\mathrm{j}} e^{-\mathrm{iq}\rho}\right]. \tag{339}$$

Inserting Equation 330 into the latter equation gives

$$\sigma_\rho = s_0 - \frac{4\tau}{\eta[1 + \Delta^2/(\eta+\gamma)^2]}|G_\rho|^2\langle n\rangle\sigma_\rho, \tag{340}$$

where s_0 is an asymptotic inversion value. Here, we have written

$$n = \langle a_{\mathrm{j}}^* a_{\mathrm{j}}\rangle \tag{341}$$

for the average number of photons in the jth mode.

Solving Equation 340 for the local inversion, we obtain

$$\sigma_\rho = s_0\left[1 + \frac{8\pi\tau\langle n\rangle e^2 d_{\mathrm{f}}^2}{\hbar\omega_{\mathrm{j}}^0 m^2 V\eta\{1 + \Delta^2/(\eta+\gamma)^2}\sin^2\vec{k}_{\mathrm{j}}\vec{\rho}\right]^{-1}. \tag{342}$$

For the average inversion

$$\langle\sigma\rangle = \frac{1}{N}\sum_\rho \sigma_\rho, \tag{343}$$

we find

$$\langle\sigma\rangle = \frac{s_0}{N} \times \sum_\rho\left[1 + \frac{8\pi\tau\langle n\rangle e^2 d_{\mathrm{f}}^2}{\hbar\omega_{\mathrm{j}}^0 m^2 V\eta\{1 + \Delta^2/(\eta+\gamma)^2}\sin^2\vec{k}_{\mathrm{j}}\vec{\rho}\right]^{-1}. \tag{344}$$

In the continuum limit Equation 344 becomes

$$\langle\sigma\rangle = \frac{s_0}{V}\int \mathrm{d}r\left[\frac{1}{1 + (4\tau n\gamma/\rho S)\sin^2\vec{k}\vec{r}}\right], \tag{345}$$

where $N = \rho V$ and $\langle n\rangle = nV$. Thus, ρ is atomic density and n is photon density. Furthermore,

$$S = \frac{\hbar\omega_{\mathrm{j}}^0 m^2\gamma\eta}{2\pi e^2\rho\delta_{\mathrm{f}}^2}\left[1 + \frac{\Delta^2}{(\eta+\gamma)^2}\right]. \tag{346}$$

Substituting the form for γ given by Equation 343 into the last equation, together with the continuum form of next equation

$$G_\rho = -\frac{e}{m}\sqrt{\frac{2\pi}{\hbar\omega_{\mathrm{j}} V}}\sin(\vec{k}_{\mathrm{j}}\vec{\rho})e_{\mathrm{j}}\delta_{\mathrm{j}} \tag{347}$$

gives the following relation

$$S = \frac{s_0}{V}\int \mathrm{d}r\frac{\sin^2\vec{k}\vec{r}}{1 + (4\tau n\gamma/\rho S)\sin^2\vec{k}\vec{r}}, \tag{348}$$

with $\int \mathrm{d}rV$, Equations 345 and 347 may be manipulated to yield

$$n = \frac{\rho s_0}{4\pi\gamma}\left(1 - \frac{\langle\sigma\rangle}{s_0}\right). \tag{349}$$

Finally, integrating Equation 345 we obtain

$$\frac{\langle\sigma\rangle}{s_0} = \frac{1}{\sqrt{1 + 4\tau n\gamma/\rho S}}. \tag{350}$$

With Equation 574 this expression becomes

$$\langle\sigma\rangle = \frac{S}{2}\left(1 + \sqrt{1 + \frac{4s_0}{S}}\right). \tag{351}$$

For laser action to be realized, one must have the photon density (at frequency ω_{j}), $n > 0$. From Equation 349 this condition is satisfied providing $s_0 > \langle\sigma\rangle$, or from Equation 287, $s_0 > S$, which with Equation 349 gives

$$s_0 > \frac{\hbar\omega_{\mathrm{j}}^0 m^2\gamma\eta}{\pi e^2\rho\delta_{\mathrm{f}}^2}\left[1 + \frac{\Delta^2}{(\eta+\gamma)^2}\right]. \tag{352}$$

At resonance, $\Delta = 0$ and the right-hand side of the preceding expression is minimized. In this limit Equation 348 reduces to the same form given by the

Schawlow-Townes criterion [351] for randomly distributed atoms. However, in Equation 348, the matrix element $\delta_f \equiv \langle v\rho|p|c\rho\rangle$ is appropriate to excitons whereas the matrix element in the Schawlow-Townes expression is relevant to a single atom.

For Wannier-Mott excitons the criterion Equation 348 maintains with δ_f replaced by

$$\delta_{W-M} \equiv \sum_{\rho'} F(\rho') \int \mathrm{d}r w_v^*(r-\rho) p w_c(r-\rho-\rho'). \tag{353}$$

In this expression $F(\rho)$ is the wavefunction of a Wannier-Mott exciton [352]. In preceding investigation Liu and Liboff have considered rigid lattices. If lattice vibrations are present, exciton-phonon interactions will cause exciton diffusion [352] which may diminish coherence for lasing. However, lasing will still occur [343] providing the relaxation time for exciton diffusion is longer than the exciton relaxation time, τ in Equation 328. This will be the case for sufficiently weak excito-phonon interaction (details see below).

As was shown above optical transitions in pure III–V compounds which can be used for laser action are band-band transitions. In II–VI compounds (as well as LiH [1] and etc.), the recombination process of electrons and holes via exciton states is more favorable than the band-band transition [352–354]. During last four decades laser action has been obtained in II–VI compounds by electron beam bombardment [355–357], by optical excitation [353, 358–360]. The laser transitions involve the A_1 - nLO phonons, where $n = 1$, 2. Gain measurements [357, 359] and simultaneous measurements of the emission intensities of the A_1 line (direct A - exciton recombination [353]) and the A_1 - LO line also confirm the statement that in CdS the A_1 - LO (A_1 - 2LO) line starts to lase for sufficiently high pump rates (see also Fig. 94) (details see [356]). In cited papers Haug [342] calculated the temperature

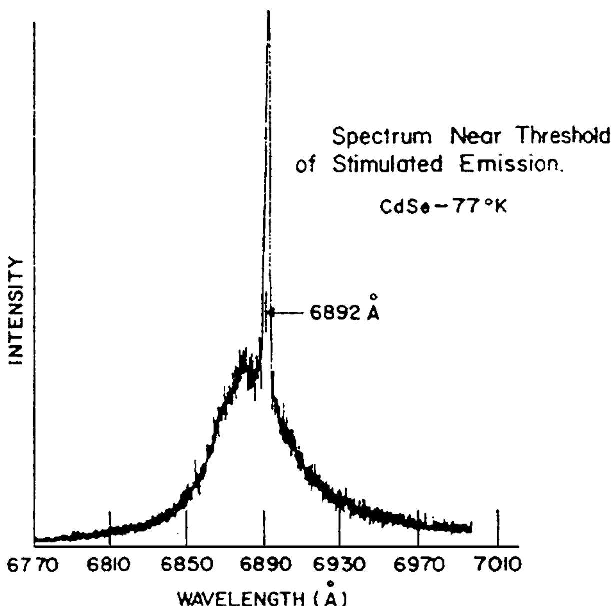


Figure 94 The onset of stimulated emission in CdSe at 77 K (after [356]).

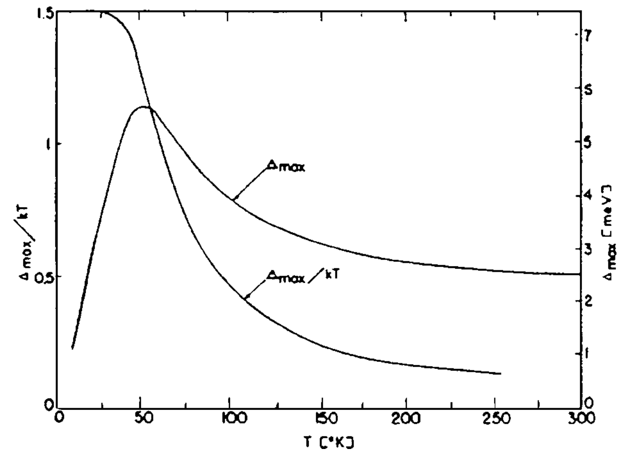


Figure 95 Maximum gain frequency Δ_{max} vs temperature (after [342]).

dependence of the maximum gain frequency at threshold (see Fig. 95). The result is simple in the low-temperature limit

$$\Delta_{max} \to \frac{3}{2}kT \tag{354a}$$

and also in the high-temperature limit

$$\Delta_{max} \to (3\kappa kT/B)^{2/5} \exp(-2\hbar\nu/5kT, \tag{354b}$$

where $2\kappa = 1.25 \times 10^{12}\ \mathrm{s}^{-1}$, corresponding to losses of $100\ \mathrm{cm}^{-1}$, $B = 1.55 \times 10^{35}\ \mathrm{erg}^{-3/2}\,\mathrm{s}^{-1}$ for CdS crystals. These limiting results have also been given by Mashkevich *et al.* [360]. The typical experimental gain I_{thr} (T) dependence, obtained in paper [359] is presented in Fig. 96. There are shown some spectra of stimulated emission at different temperatures. Authors of [359] indicated some contradiction of their experimental results with theoretical description.

4.4. Nonlinear properties of excitons in isotope-mixed crystals

Another application of isotope pure and isotope mixed crystals that will be discussed here is related to the possibility of using an isotopically mixed medium (e.g., $\mathrm{LiH}_x\mathrm{D}_{1-x}$ or $^{12}\mathrm{C}_x\ ^{13}\mathrm{C}_{1-x}$) as an oscillator of coherent radiation in the ultraviolet spectral range [361, 362]. To achieve this, the use of indirect electron transitions involving, say, LO phonons was planned [342, 363]. As was shown above using indirect electron transitions involving phonons to degenerate coherent radiation in semiconductors was originally proposed by Basov *et al.* (see [355] and reference therein). Kulevsky and Prokhorov [358] were the first to observe stimulated radiation using emission lines of LO phonon repetitions in CdS crystals on two photon excitation (see also [364]). The detection of LO phonon replicas of free-exciton luminescence in wide-gap insulators attracted considerable attention to these crystals (see e.g. Plekhanov [5], Plekhanov [6]). At the same time it ia allowed one to pose a question about the possibility of obtaining stimulated emission in UV (VUV) region (4–6 eV) of the spectrum, where no solid state sources for coherent radiation exist yet. In the first place this

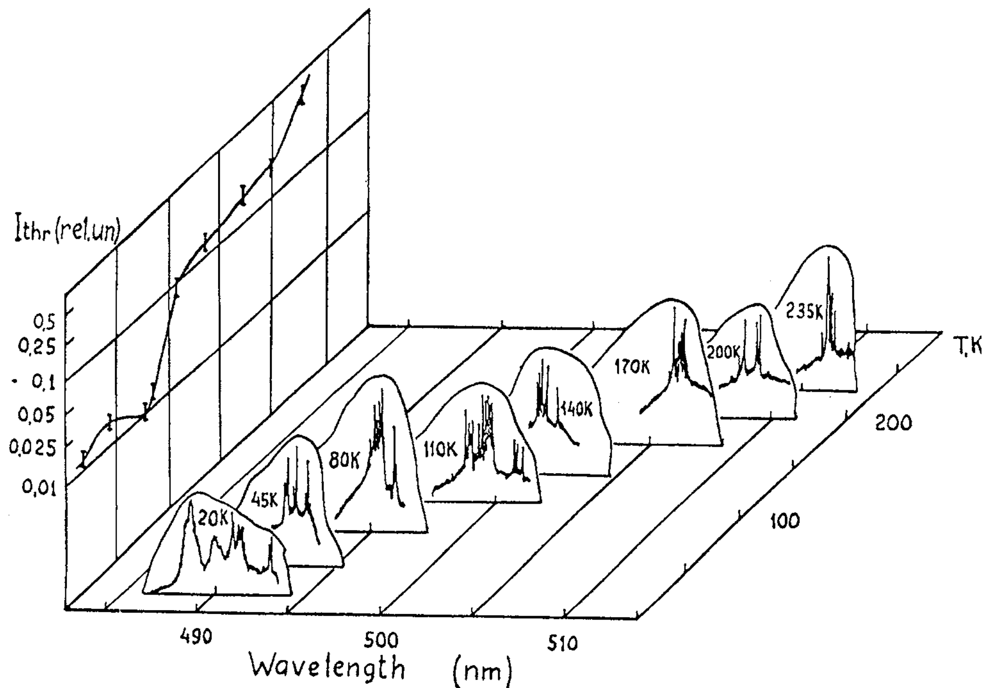


Figure 96 The dependence I_{thr}(T) and some examples of (above-threshold) lasing spectra (in the range of A_1–1LO; A_1–2LO phonons) at different temperature (after [359]).

related to the emitters working on the transitions of the intrinsic electronic excitation (exciton). The last one provides the high energetical yield of the coherent emission per unit volume of the substance. The results obtained on solidified xenon (Basov *et al.* [365]) and argon (Schwenter *et al.* [366]) under electron beam excitation with following excimer molecules emission form an exception.

In this part we will discuss the investigation results of the influence of the excitation light density on the resonant secondary emission spectra of the free-exciton in the wide-gap insulator LiH_xD_{1-x} ($LiH_{1-x}F_x$) crystals. The cubic LiH crystals are typical wide-gap ionic insulator with $E_g = 4.992$ eV [1] with relatively weak exciton-phonon interaction however: $E_B/\hbar\omega_{LO} = 0.29$ where E_B and $\hbar\omega_{LO}$ are exciton binding energy and longitudinal optical phonon's energy, respectively. Besides it might be pointed out that the analogous relation for CdS, diamond and NaI is 0.73, 0.45 and 12.7, respectively (Plekhanov [367]). Fig. 97 depicts, as an example, the exciton luminescence spectrum of pure (LiH and LiD) and mixed (LiH_xD_{1-x}) crystals at a low temperature. Analogous results for $^{12}C_x$ $^{13}C_{1-x}$ mixed diamond crystals are shown in Fig. 98. A common feature of all three spectra depicted in Fig. 97 is a phononless emission line of free excitons and its 1LO and 2LO phonon replicas. An increase in the density of the exciting light causes a burst of the radiation energy in the long-wave wing of the emission of the 1LO and 2LO repetitions (see Fig. 99) at a rate is higher for the 1LO repetion line [361]. A detailed dependence of the luminescence intensity and the shape of the 2LO phonon replica line are presented in Fig. 100 and Fig. 101, respectively. The further investigations have shown [369]

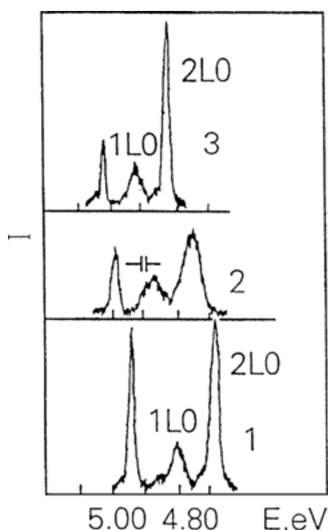


Figure 97 Photoluminescence spectra of free excitons at 4.2 K: 1—LiH; 2—LiH_x D_{1-x} and 3—LiD crystals (after [5]).

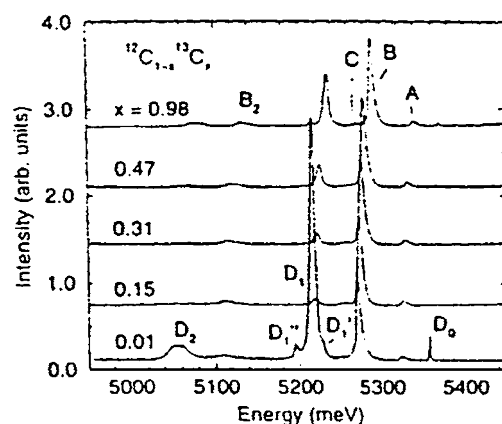


Figure 98 Cathode—luminescence spectra of isotopically modified diamond at 36 K. Intrinsic photo-assisted recombination peaks are labelled in the top spectrum, those from boron-bound excitons in that at the bottom (after [368]).

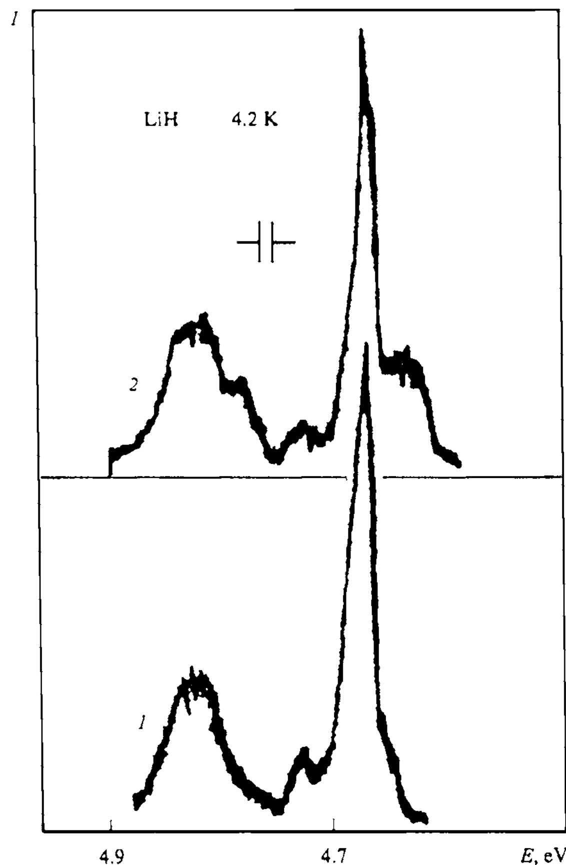


Figure 99 Luminescence spectra of free excitons in LiH crystals in the region of the emission lines of 1LO and 2LO phonon repetitions at 4.2 K for low (1) and high (2) density of excitations of 4.99 eV photons. The scales of the different curves are different (after [369]).

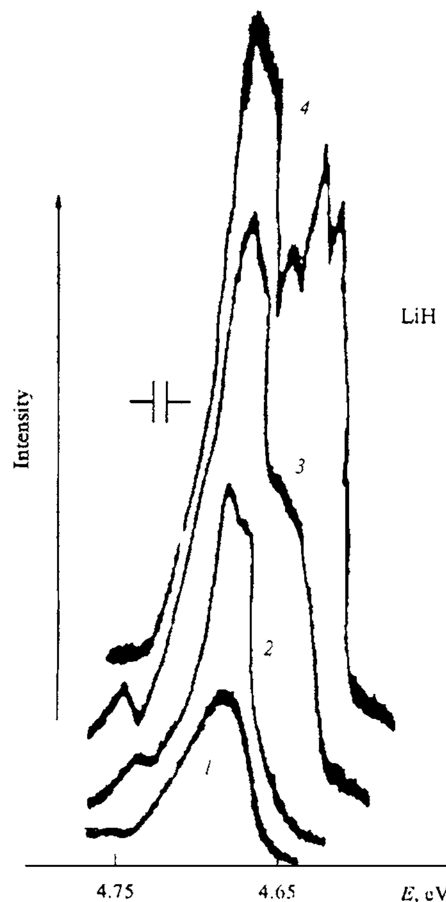


Figure 101 The dependence of the shape of 2LO replica line on the excitation intensity (I_0) light: 1—0.05I_0; 2—0.09I_0; 3—0.35I_0; 4—I_0 (after [369]).

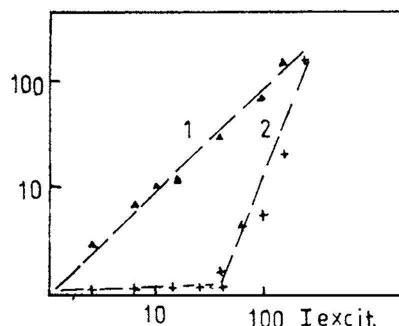


Figure 100 The dependence of the intensity in the maximum (1) and on the long-wavelength side (2) of 2LO replica emission line of LiH crystals on the excitation light intensity (after [369]).

that with the increase of the excitation light intensity at the beginning a certain narrowing can be observed, followed by widening of the line of 2LO phonon replica with a simultaneous appearance of a characteristics, probably mode structure. From Fig. 100 it can be seen that the coupling between longwavelength luminescence intensity and excitation light intensity is not only linear, but, in fact, of a threshold character as in case of other crystals [332, 364]. A proximity of the exciton parameters of LiH and CdS (ZnO) crystals allowed to carry out the interpretation of the density effects in LiH on the analogy with these semiconducting compounds. Coming from this in the paper [369] it was shown that for the observed experimental picture on LiH crystals to suppose the exciton-phonon mechanism of light generation [342] is enough the excitons density about 10^{15} cm^{-3}. This is reasonable value, if the high quality of the resonator mirrow—the crystal cleavage "in situ" and relatively large exciton radius (r = 40 Å [5] is taken into account. To this light mechanism generation must be also promoting a large value of the LO phonon energy ($\hbar\omega_{LO}$ = 140 meV) . Owing to this the radiative transition is being realized in the spectral region with a small value of the absorption coefficient, and thus with a small losses in resonator (details see also [367]).

In the present part we briefly analyse the shift of free exciton luminescence on the crystal lattice dependence in the first step it will be considered F doped of LiH crystals. The reflectance spectra of the investigated crystals with clean surface (cleaved in LHeT) had a distinctly expressed excitonic structure. Typical reflection spectra of $LiH_{1-x}F_x$ with mirror surface is depicted in Fig. 102 [370]. The crystal cleavage is carried in superfluid helium of helium cryostat bath. In the Fig. 102 for the comparison is shown the reflection spectrum of pure crystal of LiH (curve 1). All spectra possess the identical longwavelength exciton structure: we can see 1 s and 2 s exciton states. It's clearly seen, that the F addition in LiH crystals leads, as it naturally expected, to shortwavelength shift of the spectrum as the whole. With the growth x is increased the energetic interval between 1s and 2s exciton states and at the same time their shortwavelength shift is different and has saturated character

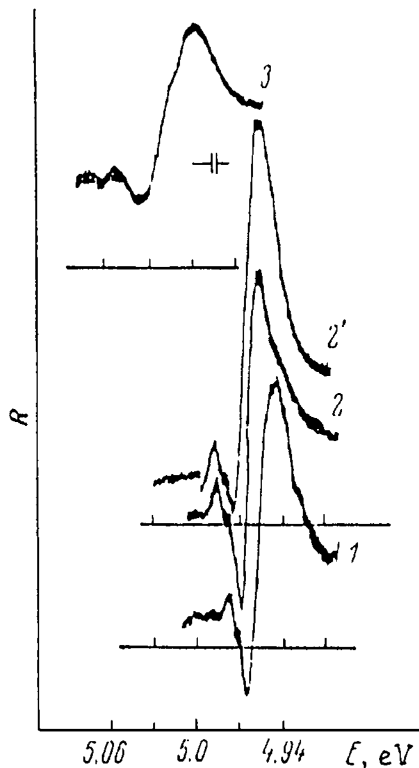


Figure 102 Reflection spectra of LiH (1); $LiH_{1-x}F_x$ ($x = 0.06\%$) (2); and ($x = 1.6\%$) (3) at 78 and 4.2 K (2′) (after [370]).

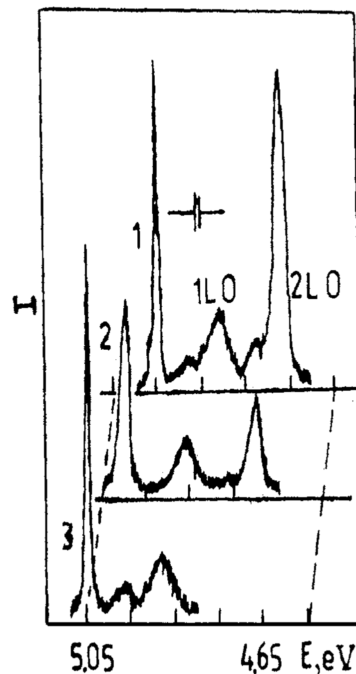


Figure 103 Free excitons luminescence spectra of LiH (1); $LiH_{0.984}F_{0.016}$ and $LiD_{0.992}F_{0.008}$ crystals cleaved in liquid helium (after [373]).

[363]. At the maximum value of $x \leq 2.5\%$ the exciton Rydberg (E_B), obtained on the hydrogen-like formulae is equal 75 ± 3 meV ($E_B = 40$ meV for LiH). For $LiD_{1-x}F_x$ crystals at $x \leq 1.6\%$ exciton binding energy is equal $E_B = 57 \pm 2$ meV. Supposing the linear dependence E_B on the x concentration in $LiH_{1-x}F_x$ crystals, we obtain $E_B = 62 \pm 25$ meV for $x = 1.6\%$ and the experimental meaning of this parameter is equal 67 ± 3 meV. Such a large error at the theoretical extrapolation is connected with the large errors when determining E_B for LiF crystals [371, 372]. However, despite of the identical structure of all free-excitons luminescence spectra, it is necessary to note a rather big variation of the luminescence intensity of the crystals from the different batches observed in the experiment. Therefore the crystals possessing the maximum value of the free exciton luminescence quantum yield were chosen for measurements of the density effects.

The luminescence spectra of virgin and mixed crystals are very likely and consist of narrow zero phonon line and its more wider LO replicas [6, 369]. As well as in the reflection spectra (see above) the dopant of LiH crystals with deuterium or fluorine is drawn to shortwavelength shift of the luminescence spectrum as a whole (Figs 97 and 103). The increasing of the deuterium concentrations leads to the widening of the luminescence line (see also [5]). The increasing of the fluorine concentration is causing (Figs 103, 104), except the spectrum shift, the sharp ignition of zero phonon line intensity in comparison with the lines of Lo-replicas intensity (see also [370]). Except indicated effects, the fluorine activation of the LiH (LiD) crystals is shifting the temperature quenching of the free excitons luminescence in the more high region. As on

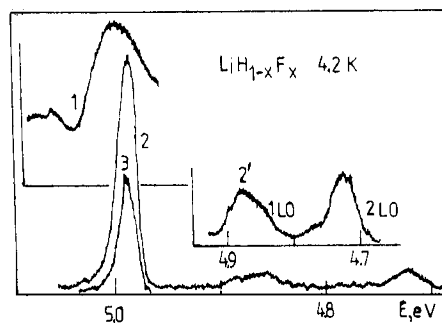


Figure 104 Reflection (1) and luminescence (2, 2′) spectra at 4.2 K and 260 K (3) of $LiH_{1-x}F_x$ mixed crystals cleaved in liquid helium (after [373]).

example on the Fig. 104 is depicted the reflection and luminescence spectra of $LiH_{1-x}F_x$ crystals in the wide temperature range [5]. It can be seen that free excitons zero-phonon emission line in these crystals is reliably registrated practically up to the room temperature. Taking into account the mixed crystals lattice potential relief it could (compare Fig. 105) not be excluded absolutely the possibility to obtain the laser generation on the zero-phonon line emission that was already obtained in the paper [374].

In conclusion of this section we should underlined that if the observable mode structure is really caused by the laser generation it may be smoothly tuned in the region of energies 4.5 ± 5.1 eV owing to smooth transition of the line emission energy in the LiH_xD_{1-x} (LiH_xF_{1-x}; LiD_xF_{1-x}) mixed crystals [5, 6] as well as in the range 5.35–5.10 eV in $^{12}C_x\,^{13}C_{1-x}$ mixed crystals (see also Fig. 20 in [1]).

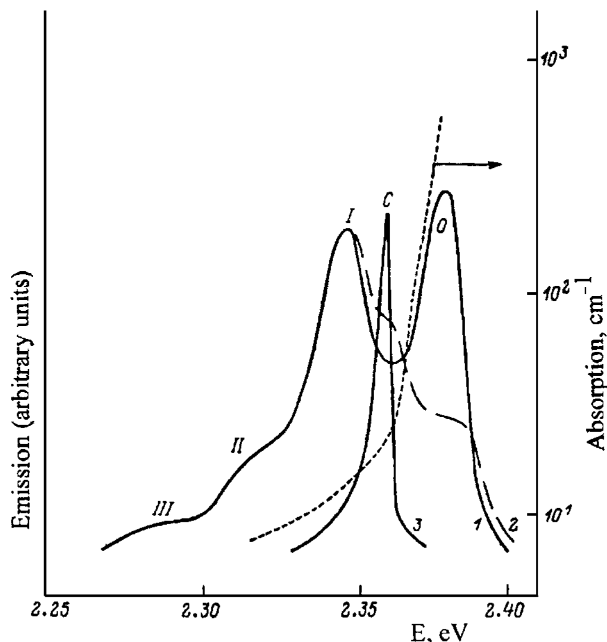


Figure 105 Emission spectra of $CdS_{0.9}Se_{0.1}$ at different levels of excitation: 10^2(1); 10^4(2) and 10^6 (3) Wt/cm^2. 0—zero-phonon line; I—III—LO—phonon replicas and dashed line is the absorption spectra, $T = 2$ K (after [374]).

Chapter 5. Other unexplored applications of isotopic engineering

5.1. Isotopic information storage

The current rapid progress in the technology of high-density optical storage makes the mere announcing of any other thinkable alternatives a rather unthankful task. An obvious query 'who needs it and what for?' has, nevertheless, served very little purpose in the past and should not be used to veto the discussion of non-orthodox technological possibilities. One such possibility, namely the technology of isotopic information storage (IIS) is discussed in this paragraph.

Isotopic information storage may consist in assigning the information 'zero' or 'one' to mono-isotopic microislands (or even to a single atoms) within a bulk crystalline (or thin film) structure. This technique could lead to a very high density of ROM-type (read-only memory or permanent storage) information storage, probably up to 10^{20} bits per cm^3. The details are discussed by Berezin *et al.* [100, 375, 376]: here it notes only that the use of tri-isotopic systems (e.g., ^{28}Si; ^{29}Si; ^{30}Si) rather than di-isotopic (e.g., ^{12}C; ^{13}C) could naturally lead to direct three dimensional color imaging without the need for complicated redigitizing (it is known that any visible color can be simulated by a properly weighted combination of three prime colors, but not of two).

Indeed, let us assume that the characteristic size of one information-bearing isotopic unit (several atoms) is 100 Å. Then 1 cm^3 of crystalline structure, e.g. diamond, is able to store roughly $(10^8)^3/100 = 10^{22}$ bits of information [376]. This capacity greatly exceeds that need to store the information content of all literature ever published ($\cong 10^{17}$ bits), including all newspapers.

The main potential advantage of isotope-mixed crystals lies in the fact that the information is incorporated in the chemically homogeneous matrix. There are no chemically different impurities (like in optical storage with color centres) or grain boundaries between islands of different magnetization (like in common magnetic storage). The information in isotope-mixed crystals exists as a part of the regular crystals lattice. Therefore, the stored information in isotope-mixed crystals is protected by the rigidity of the crystal itself. There are no "weak points" in the structure (impurities, domain wells, lattice strain etc) which can lead to the information loss due to bond strains, enhanced diffusion, remagnetization, etc. Differences in the bond lengths between different isotopes (e.g., ^{28}Si–^{29}Si or ^{29}Si–^{30}Si; H–D and so on) are due to the anharmonicity of zero-point vibrations (see, e.g., [28]). This is not enough for the development of any noticeable lattice strains although these differences are sufficiently large to be distinguishably detected in IIS-reading).

The mechanism potentially available in IIS for the writing-in of the information may be divided into two general categories. The first category refers to all techniques which are able to direct externally a particular atom to a specified position on the surface of the growing crystal structure. Any beam technique with ability of focusing on 1 Å scale could, in principle, appear feasible for such purpose. The second category relates to all 'internally operated' possibilities, i.e., delivering of the required isotope as a part of the molecule and depositing it in a particular position through some chemical process (e.g., exchange reaction, chemisorption, etc). This group of possibilities is, in fact, similar to the DNA-RNA mechanism in actual biological information transfer in living systems. Some chemically very simple crystals have, nevertheless, a very complex lattice structure. One known example is elementary boron [377, 378], which can crystallize in a beta-rhombohedral structure with 105 atoms in a unit cell with 15 crystallographically nonequivalent positions. Moreover, various atoms have 3 different coordination numbers: 91 atoms have the coordination number 6; 12 atoms, 8 and 2 "special" atoms have 9 nearest neighbors [378]. This peculiarity of the crystalline boron is rather surprising in view of the fact that it is an elementary (monoatomic) crystal. Similar possibilities exist for lattices with 2 elements, e.g., the silicides of manganese are known to form very complex structures [379]. It is therefore, legitimate to consider the ability of such structures with complex crystal lattice to provide the basis for compact information storage within the frameworks of the model of an alternative (nonorganic) genetic code. In principle, isotopic combinations could provide the basis for the storage information even in "simple" crystals (e.g., in carbon or silicon-based structures) and not only just in crystals with complex unit cells. It is possible even to raise the question what (if any) effects could be connected with isotopic permutations in "regular" biology (e.g., ^{12}C- and ^{13}C combinations in various fragments of DNA) Nevertheless, in crystals with large and sophisticatedly constructed unit cells the already "preexisting" significant level of structural complexity makes them, apparently, more preferable candidates for the evolution game of isotopic information-bearing arrays (details see [378]).

The possible key to a 3D-access could, in principle, be provided by any method which is able to probe the nuclear mass and/or magnetic moment of a single atom at a particular lattice site below surface. Without detailed elaborations the following possibilities present themselves:

1. Spectroscopy of localized crystal vibrations (see also, [1] which generally contains the information on the vibrational frequencies of an individual atom (due to square-root-of mass) dependence of the vibrational frequencies this could be a spectroscopically pronounced effect.
2. Recoil phenomena (e.g., Rutherford backscattering).
3. Nuclear magnetic resonance (NMR).
4. Spin-sensitive neutron scattering.

Of course, there has to be a great deal of perfectioning refinement to these or other techniques before they can actually be applied for their use in isotope-mixed crystals.

5.2. Isotopic structuring for fundamental studies

Isotopic substitution has made it possible to produce the objects of research that earlier were simply inaccesible (with the exception of the LiH-LiD system). The use of such objects allows the investigation of not only the isotope effects in lattice dynamics (elastic, thermal, and vibrational properties (see reviews [2–6]) but also the influence of such effects on the electronic states (the renormalization of the band-to-band transition energy E_g, the exciton binding energy E_B, and the size of the longitudinal-transverse splitting Δ_{LT} [5, 6]).

Furthermore, it is widely known that the melting and boiling points of ordinary water and heavy water (D_2O) differ by a few degrees centigrade. For elements heavier than hydrogen the isotopic differences in melting points (ΔT) of elemental and complex solids are generally smaller but also detectable. It is quite surprising, however, that there are almost no reports of direct measurement of these differences in the literature.

Another noticeable fact is that sometimes the isotope effect shows a drastic "self-amplification," e.g., isotopic replacements of Ba and Ti in $BaTiO_3$ (both are heavy elements) can shift the phase transitions temperatures by as much as 200 K [380]. The reason(s) for such selective anomalies are not yet clearly established. This part are widely considered early in the book [1].

5.3. Other possibilities

Here we shall briefly list a few additional possibilities of isotopic structuring (see also [2–4]).

1. Very perspective direction of isotope engineering could be based on exploiting the differences in thermal conductivity (see above) between isotopically pure and isotopically mixed solids for purposes such as phonon focusing, precise thermometry based on isotopically-gradiented structures, etc.
2. The use of isotopically structured Ni-films for neutron interference filters has been reported by Antonov *et al.* [381].
3. Isotopically structured light devices. This could slightly shift the spectral characteristics and lead to some changes in the kinetics of energy transfer, modify the lifetimes, recombination rate, etc.
4. Since the speed of sound is proportional to $\sqrt{M}$, variations in isotopically structured acousto-electronic devices (transducers, surface acoustic wave devices, etc.) could be significant, especially in achieving phase differences over the relatively short isotopically distinguished paths (see also [177]).
5. The possibility to get at rather low pressure the transition of metal-insulator with metallic conductivity on the zone genetic related with hydrogen in LiH crystals (see also [177] and references therein).
6. The use of the isotope boundary for Mössbauer filtration of synchrotron radiation, since this makes it possible to get rid of the background noise caused by the interaction between synchrotron radiation and the electrons in matter [382].
7. Isotope-based quantum computers (see e.g. [177, 383, 384]).

Above we have outlined several, mostly untested possibilities arising from exploiting differences in various stable isotopes and purposeful isotopic structuring. The above examples of the potential capabilities of isotopic engineering by no means an exhaustive list.

6. Conclusion

In this review, we have presented briefly the results of experimental and theoretical studies of the objects of research that earlier were simply in accessible (naturally with exception of LiH_xD_{1-x} crystals). The use of such objects allows the investigation of not only the isotope effects in lattice dynamics (elastic, thermal and vibrational properties) but also the influence of such effects on the electronic states via electron-phonon coupling (the renormalization of the band-to-band transition energy E_g, the exciton binding energy E_B and the size of the longitudinal-transverse splitting Δ_{LT}).

Substituting a light isotope with a heavy one increases the interband transition energy E_g (excluding Cu-salts) and the binding energy of the Wannier-Mott exciton E_B as well as the magnitude of the longitudinal-transverse splitting Δ_{LT}. The nonlinear variation of these quantites with the isotope concentration is due to the isotopic disordering of the crystal lattice and is consistent with the concentration dependence of line halfwidth in exciton reflection and luminescence spectra. A comparative study of the temperature and isotopic shift of the edge of fundamental absorption for a large number of different semiconducting and insulating crystals indicates that the main (but not the only) contribution to this shift comes from zero oscillations whose magnitude may be quite considerable and comparable with the energy of LO phonons. The theoretical description of the experimentally observed dependence of the binding energy of the Wannier-Mott exciton E_B

on the nuclear mass requires the simultaneous consideration of the exchange of LO phonons between the electron and hole in the exciton, and the separate interactions of carriers with LO phonons. The experimental dependence $E_B \sim f(x)$ for LiH_xD_{1-x} crystals fits in well enough with the calculation according to the model of large-radius excitonin a disordered medium; hence it follows that the fluctuation smearing of the band edges is caused by isotopic disordering of the crystal lattice.

Details analysis the process of the self-diffusion in isotope pure and heterostructures was done in Chapter 1. This chapter was organized around general principles that are applicable to all materials. There is briefly discussed very popular in nowadays the SIMS technique. As is well-known self-diffusion is the migration of constituent atoms (isotopes) in materials. The knowledge obtained in self-diffusion studies is pivotal for the understanding of many important mass transport process including impurity diffusion in solids to use in different semiconductor devices.

The new reactor technology-neutron transmutation doping (NTD) of semiconductors was described in Chapter 2. Capture of thermal neutrons by isotope nuclei followed by nuclear decay produces new elements, resulting in a very number of possibilities for isotope selective doping of solids. There are presented different facilities which use in this reactor technology. The feasibility of constructing reactors dedicated to the production of NTD silicon, germanium (and other compounds) was analyzed in terms of technical and economic viability and the practicality of such a proposal is examined. The importance of this technology for studies of the semiconductor doping (materials for different devices) as well as metal-insulator transitions and neutral impurity scattering process is underlined.

The use of the isotopes in a theory and technology of the optical fibers we considered in Chapter 3. This chapter is addressed to readers who wish to learn about fiber communications systems and, particular, about the properties of optical fibers. Very briefly in this chapter we describe the Maxwell equations as well as wave electromagnetic equation. In this chapters we describe not only the properties of optical fibres but also the materials for optical fiber and fiber technology. It was shown also the influence of the isotopes on properties of the optical fibers.

Chapter 4 is devoted the application of isotope effect in laser physics. There is short description of theory and practice of semiconductor lasers. The discovery of the linear luminescence of free excitons observed over a wide temperature range has placed lithium hydride [1], as well as crystals of diamond in line as prospective sources of coherent radiation in the UV spectral range. For LiH isotope tuning of the exciton emission has also been shown.

The last chapter of this book is devoted to description of the other unexplored applications of isotopic engineering. In the first place we considered the materials for information storage in modern personal computers as well as in biology. Large perspective has the isotope-base quantum computers. We should add here that the strength of the hyperfine interaction is proportional to the probability density of the electron wavefunction at the nucleus. In semiconductors, the electron wavefunction extends over large distances through the crystal lattice. Two nuclear spins can consequently interact with the same electron, leading to electron-mediated or indirect nuclear spin coupling. Because the electron is sensitive to externally applied electric fields, the hyperfine interaction and electron-mediated nuclear spin interaction can be controlled by voltages applied to metallic gates in a semiconductor device, enabling the external manipulation of nuclear spin dynamics that is necessary for quantum computation in quantum computers (details see [177, 383]).

The wide possibilities of isotopic engineering discussed in this review hold the greatest promise for application in solid-state and quantum electronics, biology, human memory, optoelectronics, different electronic devices, electronic and quantum computers, and many other modern and new technologies that are even now difficult to imaginate. And we should repeat that the main aim of this review is to familiarize readers with present and some future applications in isotope science and engineering.

Acknowledgment

I would like to express my deep thanks to many authors and publishers whose Figures and Tables I used in my review. Many thanks are due to my son N. V. Plekhanov for technical assistance, especially in graphical materials. Many thanks are due to A. DePina and K. Costello for their editorial assistance.

References

1. V. G. PLEKHANOV, "Isotope Effects in Solid State Physics" (Academic Press, New York, London, 2001).
2. M. CARDONA, in "Festkorperprobleme/Adv. in Solid State Physics," edited by R. Helbig (Vieweg, Braunschweig, Wiesbaden, 1994) p. 35.
3. E. E. HALLER, *J. Appl. Phys.* **77** (1995) 2857.
4. V. G. PLEKHANOV, *Mater. Sci. and Eng.* **35** (2001) 139.
5. *Idem.*, *Physics-Uspekhi (Moscow)* **40** (1997) 553.
6. *Idem.*, *Rep. Progr. Phys.* **61** (1998) 1045.
7. G. LEIBFRIED and W. LUDWIG, in "Theory of Anharmonic Effects in Crystals," edited by F. Seitz and D. Turnbull (Academic Press, New York, 1961) Vol. 12.
8. W. C. ROBERTS-AUSTEN, *Philos. Trans. R. Soc. (London)* A **187** (1896) 404.
9. R. F. MEHL, *Trans. AIME* **122** (1936) 11.
10. W. JOST, "Diffusion in Solids, Liquids, Gass" (Academic Press, New York, 1952).
11. W. SEITH, "Diffusion in Metallen" (Springer–Verlag, Berlin, 1955).
12. L. SLIFKIN, D. LAZARUS and C. T. TOMIZUKA, *J. Appl. Phys.* **23** (1952) 1032.
13. C. T. TOMIZUKA, in "Methods of Experimental Physics," edited by K. Larl-Horrowitz and V. A. Johnson (Academic Press, New York, 1959) Vol. 6, Part A, p. 364.
14. A. S. NOWICK, *J. Appl. Phys.* **22** (1951) 1182.
15. N. F. MOTT and R. F. GURNEY, "Electronic Processes in Ionic Crystals" (Clarendon Press, Oxford, 1948).
16. J. H. CRAWFORD, JR and L. M. SLIFKIN (eds.), "Point Defects in Solids" (Plenum Press, London, 1975) Vol. 1/2.
17. R. E. HOFFMAN and D. TURNBULL, *J. Appl. Phys.* **22** (1951) 634.
18. J. CADEK and J. JANDA, *Hung. Listy* **12** 1008 (1957) (English transl.: AERE Transl. 840, Harwell).

19. Y. ADDA and J. PHILBERT, La Diffusion dans les Solides (Presse Universitaires de France, Paris, 1966).
20. T. S. LUNDY, in "Techniques of Metals Research," edited by R. A. Rapp (Wiley, New York, 1970) Vol. 4, Part 2, Chap. 9A.
21. F. BENIERE, in "Mass Transport in Solids" edited by F. Beniere and C. R. A. Catlow (Plenum Press, London, 1983).
22. A. S. NOWICK and J. J. BURTON (eds.), "Diffusion in Solids: Recent Developments" (Academic Press, New York, 1975).
23. G. E. MURCH and A. S. NOWICK (eds.), "Diffusion in Crystalline Solids" (Academic Press, New York, 1984).
23a. P. SIGMUND, in "Sputtering by Particle Bombardment," edited by R. Behrisch (Springer–Verlag, Berlin, 1981) Chap. 2.
24. High-Power Semiconductor Devices, IEEE Transactions on Electron Devices ED-**23** (1976).
25. D. A. ANTONIADIS and I. MOSKOWITZ, *J. Appl. Phys.* **53** (1982) 9214.
26. R. A. CRAVEN, in "Semiconductor Silicon'81," edited by H. R. Huff, R. J. Kriegr and Y. Takeishi (The Electrochem. Soc. Pennington, NJ, 1981).
27. N. L. PETERSON, in "Diffusion in Solids-Recent Developments," edited by A. S. Nowick and J. J. Burton (Academic Press, New York, 1975) p. 115.
28. A. FICK, *Poggy Ann.* **94** (1855) 59.
29. J. R. MANNING, "Diffusion Kinetics of Atoms in Crystals" (Van Nostrand, Princeton, NJ, 1968).
30. A. S. NOWICK and B. S BERRY, "Inelastic Relaxation in Crystalline Solids" (Academic Press, New York, 1987) Chaps. 7 and 10.
31. D. WOLF, "Spin Temperature and Nuclear-Spin Relaxation in Matter: Basic Principles and Applications" (Oxford University Press, London–New York, 1979).
32. J. CRANK, "The Mathematics of Diffusion" (Oxford University Press, London–New York, 1975).
33. H. S. CARSLAW and J. C. JAEGER, "Conduction of Heat in Solids" (Oxford University Press, London, New York, 1959).
34. A. D. LE CLAIRE, *British J. Appl. Phys.* **14** (1963) 35.
35. P. C. CARMAN and R. A. W. HAUL, *Proc. R. Soc. (London)* A **222** (1954) 109.
36. S. J. ROTHMAN, in "Diffusion in Crystalline Solids," edited by G. E. Murch and A. S. Nowick (Academic Press, New York, 1984) Chap. 1.
37. M. L. SWANSON, R. F. MEHL, G. M. POUND and J. P. HIRTH, *Trans AIME* **224** (1962) 742.
38. W. SEITH and A. KOTTMAN, *Angew. Chem.* **64** (1952) 379.
39. H. J. QUEISSER, *J. Appl. Phys.* **32** (1961) 1776.
40. P. S. AYRES and P. G. WINCHELL, *ibid.* **34** (1968) 4820.
41. *Idem.*, *ibid.* **43** (1972) 816.
42. E. S. WAIDA, *Acta Metall.* **2** (1954) 184.
43. W. FRANK, U. GOSELE, H. MEHRER and A. SEEGER, in "Diffusion in Crystalline Solids," edited by G. E. Murch and A. S. Nowick (Academic Press, New York, 1984) Chap. 2.
44. D. E. ABOLTYN, YA. E. KARIS and V. G. PLEKHANOV, *Sov. Phys. Solid State* **22** (1980) 510.
45. R. N. GHOSHTAGORE, *Phys. Rev.* **155** (1967) 598.
46. M. WERNER, H. MEHRER and H. SIETHOFF, *J. Phys. C: Solid State Phys.* **16** (1983) 6185.
47. P. BRATTER and H. GOBRECHT, *Phys. Stat. Solidi* **37** (1970) 869.
48. A. SEEGER and K. P. CHIK, *ibid.* **29** (1968) 455.
49. D. GUPTA, *Thin Solid Films* **25** (1975) 231.
50. M. WUTTING, *Scr. Metall.* **3** (1969) 175.
51. L. W. BARR, D. A. BLACKBURN and A. F. BROWN, "Radioisotopes in the Physical Sciences and Industry" (Int. At. Energy Ag., Vienna, 1962) p. 137.
52. D. L. STYRIS and C. T. TOMIZUKA, *J. Appl. Phys.* **34** (1963) 1001.
53. L. W. BARR, J. N. MUNDY and A. H. ROWE, in "Amorphous Materials," edited by R. W. Douglas and B. Ellis (Wiley, New York, 1972) p. 243.
54. J. E. KELLY and N. TOMOZAWA, *J. Amer. Ceram.* **63** (1980) 478.
55. D. M. HOLLOWAY, *J. Vac. Sci. Technol.* **12** (1975) 392.
56. J. W. MAYER and J. M. POATE, in "Thin Film Interdiffusion and Reactions," edited by N. B. Urli and J. W. Corbett (Wiley, New York, 1978) p. 119.
57. D. F. STEIN and A. JOSHI, *Annu. Rev. Mater. Sci.* **11** (1981) 485.
58. C. LEA, *Met. Sci.* **17** (1983) 357.
59. H. LUTZ and R. SIZMAN, *Z. Naturforsch.* **19A** (1964) 1079.
60. R. BEHRISCH, in "Sputtering by Particle Bombardment I," edited by R. Behrisch (Springer-Verlag, Berlin, New York, 1981).
61. B. CHAPMAN, "Glow Discharge Processes" (Wiley, New York, 1980).
62. R. D. READER and H. R. KAUFFMAN, *J. Vac. Sci. Technol.* **12** (1975) 1344.
63. J. N. MUNDY and S. J. TOTHMAN, *ibid.* A **1** (1983) 74.
64. D. GUPTA and R. R. C. TSUI, *Appl. Phys. Lett.* **17** (1970) 294.
65. A. ATKINSON and R. I. TAYLOR, *Thin Solid Films* **46** (1977) 291.
66. P. DORNER, W. GUST, A. LODDING, H. OCKLINS, B. PREDEL and U. ROLL, *Z. Metallk.* **73** (1982) 325.
67. H. LIEBL, *J. Vac. Sci. Technol.* **12** (1975) 385.
68. W. REUTER and J. E. E. BAGLIN, *ibid.* **18** (1981) 282.
69. J.-L. SERAN, Thesis, Universite de Paris, 1976.
70. M.-P. MACHT and V. NAUNDORF, *J. Appl. Phys.* **53** (1982) 7551.
71. H. BRACHT, M. NORSENG, E. E. HALLER, K. EBERL and M. CARDONA, *Solid State Commun.* **112** (1999) 301.
72. G. J. SLUSSER and J. S. SLATTERY, *J. Vac. Sci. Technol.* **18** (1981) 301.
73. K. E. WILZBACH and L. KAPLAN, *J. Amer. Chem. Soc.* **72** (1950) 5795.
74. D. S. CATLETT, J. N. SPENCER and G. J. VOGT, *J. Chem. Phys.* **58** (1973) 3432.
75. J. N. SPENCER, D. S. CATLETT, G. J. VOGT, *ibid.* **59** (1973) 1314.
76. V. B. PTASHNIK and T. YU. DUNAEVA, *Fiz. Tverd. Tela* **19** (1977) 1643 (in Russian).
77. K. FUNKE and H. RICHTERING, *Ber. Bunsenges. Phys. Chem.* **72** (1968) 619.
78. T. A. DELLIN, G. J. DIENES, C. R. RISCHER, R. D. HATCHER and W. D. WILSON, *Phys. Rev.* B **1** (1970) 1745.
79. F. E. PRETZEL, D. T. VIER, E. G. SZKLARZ and W. B. LEWIS, Los Alamos Scientific Lab. Rep. LA 2463 (1961).
80. H. LETAW, JR., W. M. PORTNOY and L. SLIFKIN, *Phys. Rev.* **102** (1956) 636.
81. M. W. VALENTA and C. RAMASASTRY, *ibid.* **106** (1957) 73.
82. H. WIDMER and G. R. GRUNTHER-MOR, *Helv. Phys. Acta* **34** (1961) 635.
83. D. R. CAMPBELL, *Phys. Rev.* B **12** (1975) 2318.
84. G. VOGEL, G. HETTICH and H. MEHRER, *J. Phys. C: Solid State Phys.* **16** (1983) 6197.
85. M. WERNER, Ph. D. Thesis, University of Stutgart, Stutgart, FRG, 1984.
86. P. L. GRUZIN, *Dokl. Akad. Nauk SSSR* **15** (1952) 108 (in Russian).
87. J. STEIGMANN, W. SHOCKLEY and F. C. NIX, *Phys. Rev.* **56** (1939) 13.
88. A. SEEGER and W. FRANK, *Appl. Phys.* A **27** (1982) 171.
89. R. F. PEART, *Phys. Stat. Solidi* **15** (1966) K119.
90. R. N. GHOSHTAGORE, *Phys. Rev. Lett.* **16** (1966) 890.
91. B. J. MASTERS and J. M. FAIRFIELD, *Appl. Phys. Lett.* **8** (1966) 280.
92. J. M. FAIRFIELD and B. J. MASTERS, *ibid.* **38** (1967) 3148.
93. H. J. MAYER, H. MEHRER and K. MAIER, *Inst. Phys. Conf. Ser.* **31** (1977) 186.
94. J. HIRVONEN and A. ANTHLA, *Appl. Phys. Lett.* **35** (1979) 703.
95. L. KALINOWSKI and R. SEGUIN, *ibid.* **35** (1979) 211.
96. F. J. DEMOND, S. KALBITZER, H. MANNSPERGER and H. DAMJANTSCHITSCH, *Phys. Lett.* A **93** (1983) 503.
97. H. D. FUCHS, W. WALUKIEWICZ, E. E. HALLER, W. DONDL, R. SCHORER and G. ABSTREITE, *Phys. Rev.* B **51** (1995) 16817.
98. E. E. HALLER, *Semicond. Sci. Technol.* **5** (1990) 319.

99. J. SPITZER, T. RUF, M. CARDONA, W. DONDL, R. SCHORER, G. ABSTREITER and E. E HALLER, *Phys. Rev. Lett.* **72** (1994) 1565.
100. A. A. BEREZIN, *J. Phys. Chem. Solids* **48** (1987) 853; **50** (1989) 5.
101. K. COMPAAN and Y. HAVEN, *Trans. Faraday Soc.* **52** (1956) 786.
102. J. C. BOURGOIN and M. LANNO, *Radiat. Effects* **46** (1980) 157.
103. J. G. MULLENN, *Phys. Rev.* **121** (1961) 1649.
104. A. D. LECLAIRE, *Philos. Mag.* **14** (1966) 1271.
105. K. ITOH, W. L. HANSEN, E. E. HALLER, J. W. FARMER, V. I. OZHOGIN, in Proc. 5th Int. Conf. on Shallow Levels in Semicond., Kobe, Japan, 1992; H. D. Fuchs, K. M. Itoh and E. E. Haller, *Philos. Mag.* **B70** (1994) 661.
106. N. L. PETERSON, *J. Nucl. Mater.* **69/70** (1978) 3.
107. Diffusion in "Solid Metals and Alloys," edited by H. Mehrer (Landolt-Bornstein New Series, Group III, Springer–Verlag, Berlin, 1990) Vol. 26.
108. "Intrinsic Properties of Group IV Elements and III–V and I–VII Compounds," edited by O. Madelung (Landolt-Bornstein, New Series, Group III, Springer-Verlag, Berlin, 1989) Vol. 22, Pt. a.
109. T. Y. TAN, U. M. GÖSELE and S. YU, *Cri. Rev. Solid State Phys.* **17** (1991) 47.
110. B. GOLDSTEIN, *Phys. Rev.* **121** (1960) 1305.
111. H. D. PALFREY, M. BROWN and A. F. W. WILLOUGHBY, *J. Electrochem. Soc.* **128** (1981) 2224.
112. T. Y. TAN, H. M. YOU, S. YU. U. M. GÖSELE, *J. Appl. Phys.* **72** (1992) 5206.
113. D. G. DEPPE and N. HOLONYAK, JR., *ibid.* **64** (1988) R93.
114. L. WANG, L. HSU, E. E. HALLER, J. W. ERICKSON, A. FISCHER, K. EBERL and M. CARDONA, *Phys. Rev. Lett.* **76** (1996) 2342.
115. H. BRACHT, E. E. HALLER, K. EBERL and M. CARDONA, *Appl. Phys. Lett.* **74** (1999) 49.
116. L. WANG, J. A. WOLK, L. HSU and E. E. HALLER, *ibid.* **70** (1997) 1831.
117. S. O'BRIEN, B. P. BOUR and J. R. SHEALY, *ibid.* **53** (1988) 1859.
118. R. KRAUSE-REHBERG, A. POLITY, W. SIEGEL and G. KUHNEL, *Semicond. Sci. Technol.* **8** (1993) 290; K. J. BEERNIK, D. SUN, D. W. TREAT and B. P. BOUR, *Appl. Phys. Lett.* **66** (1995) 3597.
119. S. F. WEE, M. K. CHAI, K. P. HOMEWOOD and W. P. GILIN, *ibid.* **82** (1997) 4842.
120. B. L. OLMSTED and S. N. HOUDE-WALTER, *ibid.* **63** (1993) 530 and references therein.
121. W. WALUKIEWICZ, *Mater. Res. Soc. Symp.* **300** (1993) 421.
122. W. SCHOKLEY and J. L. MOLL, *Phys. Rev.* **119** (1960) 1480.
123. M. KAWABE, N. MATZUURA, N. SHIMIZU, F. HASEGAWA and Y. NANNICHI, *Jap. J. Appl. Phys.* **23** (1984) L632.
124. K. J BEERNIK, R. L. THORNTON, G. B. ANDERSON and M. A. EMANUEL, *Appl. Phys. Lett.* **66** (1995) 2422.
125. J. J. COLEMAN, P. D. DAPKUS, G. G. KIRKPATRIC, M. D. CAMRAS and N. HOLONYAK, JR., *ibid.* **40** (1982) 904.
126. J. KOBAYASHI, M. NAKAJIMA, Y. BAMBA, T. FUKUNAGA, *Jap. J. Appl. Phys.* **2** (1986) L385.
127. M. KAWABE, N. SHIMIZU, F. HASEGAWA and Y. NANNICHI, *Appl. Phys. Lett.* **46** (1985) 849.
128. J. KOBAYASHI, M. NAKAJIMA, T. FUKUNAGA, T. TAKAMORI, K. ISHIDA, and H. NAKASHIMA, *Jap. J. Appl. Phys.* **25** (1986) L376.
129. T. Y. TAN and U. GÖSELE, *J. Appl. Phys.* **61** (1987) 1841.
130. G. A. BARAFF and M. SCHLÜTTER, *Phys. Rev. Lett.* **55** (1988) 1327.
131. T. Y. TAN and U. GÖSELE, *Appl. Phys. Lett.* **52** (1988) 1240.
132. P. MEI, S. A. SCHWARZ, T. VENKATESAN, C. L. SCHWARTZ and E. COLAS, *ibid.* **50** (1987) 1823.
133. P. MEI, S. A. SCHWARZ, T. VENKATESAN, C. L. SCHWARTZ and E. COLAS, *J. Appl. Phys.* **65** (1989) 2165.
134. W. M. LI, R. M. COHEN, D. S. SIMONS and P. H. CHI, *Appl. Phys. Lett.* **70** (1997) 3392.
135. K. MURAKI and Y. HORIKOSHI, *J. Crystal Growth* **175/176** (1997) 162.
136. *idem. Inst. Phys. Conf. Ser.* **145** (1996) 547.
137. H. BRACHT, M. NORSENG, E. E. HALLER, K. EBERL and M. CARDONA, *Solid State Commun.* **112** (1999) 301.
138. H. BRACHT, E. E. HALLER, K. EBERL, M. CARDONA and R. CLARK-PHELPS, *Mater. Res. Soc. Symp. Proc.* **527** (1998) 335.
139. M. BOCKSTEDTE and M. SCHEFFER, *Zs. Phys. Chemie* **200** (1997) 195.
140. J. S. BLAKEMORE, *J. Appl. Phys.* **53** (1982) R123.
141. T. Y. TAN, *Mater. Sci. and Eng.* B **10** (1991) 227.
142. T. Y. TAN, H.-M. YOU and U. M. GÖSELE, *Appl. Phys.* A **56** (1993) 249.
143. I. LAHIRI, D. D. NOLTE, M. R. MELLOCH, J. M. WOODALL and W. WALUKIEWICZ, *Appl. Phys. Lett.* **6** (1996) 236.
144. J. M. MEESE, in "Neutron Transmutation Doping in Semiconductors," edited by J. M. Meese (Plenum Press, New York–London, 1979) p. 1.
145. "Neutron Transmutation Doping in Semiconductors," edited by J. M. Meese (Plenum Press, New York–London, 1979).
146. M. A. KRIVOGLAZ, "Theory of Scattering X-rays and Thermal Neutrons by Real Crystals" (Science, Moscow, 1967) (in Russian).
147. G. DOLLING, in "Dynamical Properties of Solids," edited by G. K. Horton and A. A. Maradudin (North-Holand, Amsterdam, 1974).
148. B. DORNER, "Inelastic Neutron Scattering in Lattice Dynamics" (Springer Tracts in Modern Physics, Berlin, 1982) Vol. 93.
149. V. G. PLEKHANOV, *Uspekhi Fizicheskikh Nauk* (Moscow) **173** (2003) 711.
150. D. E. CULLEN and P. J. HLAVAC, ENDF/B Cross Sections, Brookhaven National Laboratory, Upton, New York, 1972.
151. "Semiconductors Doped by Nuclear Reactions," edited by L. S. Smirnov (Science, Novosibirsk, 1981) (in Russian).
152. D. DE SOETE, "Neutron Activation Analysis" (John Wiley, New York, 1971).
153. K. N. MUKHIN, "Introduction in Nuclear Physics" (Atomizdat, Moscow, 1965) (in Russian).
154. M. TANENBAUM and A. D. MILS, *J. Electrochem. Soc.* **108** (1961) 171.
155. M. S. SNÖLLER, *IEEE Trans. Electron Devices* **ED-21** (1974) 313; W. HAAS and M SNÖLLER, *J. Electron. Mater.* **5** (1976) 57.
156. D. S. BILLINGTON and J. H. CRAWFORD, JR., "Radiation Damage in Solids" (Princeton University Press, Prnceton, NJ, 1961) Chap. 2.
157. M. V. CHUKICHEV and V. S. VAVILOV, *Sov. Phys. Solid State* **3** (1961) 1103; V. S. VAVILOV, "Influence of Radiation on the Semiconductors" (Science, Moscow, 1963) (in Russian).
158. T. G. G. SMITH, in [145] p. 157.
159. *idem.* in "Neutron Transmutation Doping of Semiconductor Material," edited by R. D. Larrabee (Plenum Press, New York–London, 1984) p. 83.
160. J. L. BOURDON and G. RESTELLI, in [145] p. 181.
161. J. E. MORRISEY, T. TILLINGHAST, A. P. PERRO and B. J. BALIGA, in [145] p. 171.
162. B. J. BALIGA, in "Neutron Transmutation Doping of Semiconductor Materials," edited by R. D. Larrabeeed (Plenum Press, New York–London, 1984) p. 167.
163. N. A. BICKFORD and R. F. FLEMING, in [145] p. 165.
164. D. A. BECKER and P. D. LAFLEUR, *J. Radioanal. Chem.* **19** (1974) 149.
165. M. A. KIRK and L. R. GREENWOOD, in [145] p. 143.
166. P. BREANT, in "Neutron Transmutation Doping of Semiconductor Materials," edited by Larrabee (Plenum Press, New York–London, 1984) p. 103.
167. M. T. ROBINSON, in "Nuclear Fision Reactors," edited by J. L. Hall and J. H. C. Maple (British Nuclear Energy Society, London, 1970) p. 364.
168. J. LINDHARD, V. NIELSEN, M. SCHARFF and P. V. THOMSEN, *Mat. Fys. Medd. Dan. Vid. Selsk.* **33** (1963) N10.
169. N. KALTENBORN and O. MALMROS, in "Neutron Transmutation Doping of Semiconductor Materials," edited by

R. D. Larrabee (Plenum Press, New York–London, 1984) p. 139.
170. K. HANSEN, K. STENDAL, K. ANDERSEN and K. HEYDORN, in [167] p. 91.
171. A. D. ROSSIN, published by T. H. Blewitt and T. J. Koppenaal, in "Radiation Effects," edited by W. F. Sheely (Gordon and Breach, New York, 1966) p. 561.
172. J. A HORAK and T. H. BLEWITT, *Phys. Stat. Solidi* **9** (1972) 721.
173. B. S. BROWN, T. H. BLEWITT, T. L. SCOTT and A. C. KLANK, *J. Nucl. Mat.* **52** (1974) 215.
174. E. J. PARMA, JR. and R. R. HART, in "Neutron Transmutation Doping of Semiconductor Materials," edited by R. D. Larrabee (Plenum Press, New York–London, 1984) p. 127.
175. N. A. VLASOV, "Neutrons" (Science, Moscow, 1971) (in Russian).
176. U. A. ARIFOV, V. A. SINUKOV and YU. A. KOROSTELEV, in "Crystallization of Thin Films" (Fan, Tashkent, 1970) p. 133 (in Russian).
177. V. G. PLEKHANOV, *Physics-Uspekhi (Moscow)* **43** (2000) 1147.
178. P. J. CLAIRON and J. M. MEESE, in [145], p. 291.
179. R. M. BRUGGER and W. YELLOW, in Proc. Conf. on Neutron Scattering, edited by R. M. Moon (1976) Vol. 11, p. 1117.
180. K. LARK-HOROWITZ, in Proc. Conf. OnSemicond. Materials, edited by H. K. Henish (Butterworth, London, 1951) p. 47.
181. H. C. SCHWEINLER, *J. Appl. Phys.* **30** (1959) 1125.
182. E. E. HALLER, *ibid.* **77** (1995) 2857.
183. SH. M. MIRIANASHWILI and D. I. NANOBASHWILI, *Fiz. i Techn. Poluprovodnikov* **10** (1970) 1879 (in Russian).
184. M. H. YOUNG, A. T. HUNTER, R. BARON and O. I. MARSCH, in "Neutron Transmutation Doping of Semiconductor Materials," edited by R. D. Larrabee (Plenum Press, New York–London, 1984) p. 1.
185. M. UEMATSU and K. WADA, *Appl. Phys. Lett.* **58** (1991) 2015.
186. A. P. GALUSHKA and I. D. KONOSEMKO, *Atom Energy* **13** (1991) 277.
187. J. M. MEESE and P. J. CLAIRON, in [145], p. 109.
188. C. N. KLAHR and M. COHEN, *Nucleonics* **22** (1964) 62.
189. V. A. KHARCHENKO and S. P. SOLOV'EV, *Fiz. i Techn. Poluprovodnikov* **11** (1971) 1641 (in Russian).
190. "Neutron-Transmutation—Doped Silicon," edited by J. Guldberg (Plenum Press, New York, 1981).
191. L. CERTIS, "Introduction in Neutron Physics" (Atomizdat, Moscow, 1965) (in Russian).
192. "Handbook of Physical Constants," edited by I. K. Kikoin (Atomizdat, Moscow, 1976) (in Russian).
193. H. M. JAMES and O. MALMROS, *IEEE Trans. on Electron Devices*, **ED-23** (1976) 797.
194. J. W. CLELAND, in Proc. Inter. School of Physics, Enrico Fermi - Course XVIII - Radiation Damage in Solids, edited by D. S. Billington (Academic Press, New York, 1962) p. 384.
195. H. J. STEIN, in "Neutron Transmutations Doping in Semiconductors," edited by J. M. Meese (Plenum Press, New York–London, 1979) p. 229.
196. V. A. J. VAN LINT, R. E. LEADON and J. F. COLWELL, *IEEE Trans. Nucl. Sci.* **NS-19** (1972) 18.
197. M. YOSHIDA, *J. Phys. Soc. Japan* **16** (1961) 44.
198. D. K. BRICE, *Radiat Effects* **11** (1971) 227.
199. M. BERTOLOTTI, T. PAPA, D. SETTE, V. GRASSO and G. VITALI, *IL Nuovo Cimento* **29** (1963) 4310.
200. M. BERTOLOTTI, D. SETTE and G. VITALI, *J. Appl. Phys.* **38** (1967) 2645.
201. M. L. SWANSON, J. R. PARSONS and C. W. HOELKE, in "Albany Conf. on Radiation Effects in Semiconductors," edited by J. W. Corbett and G. D. Watkins (Plenum Press, New York, 1971) p. 359.
202. G. DEN ONDEN, *Phil. Mag.* **19** (1969) 321.
203. R. S. NELSON, *Radiat. Effects* **32** (1977) 19.
204. B. R. GOSSICK, *J. Appl. Phys.* **30** (1959) 1214.
205. J. H. CRAWFORD, JR. and J. W. CLELAND, *ibid.* **30** (1959) 1204.
206. B. L. GREGORY, *IEEE Trans. on Nucl. Sci.* **NS-16** (1969) 53.
207. H. J. STEIN, *ibid.* **NS-15** (1968) 69.
208. *idem.*, *Appl. Phys. Lett.* **15** (1969) 61.
209. E. E. HALLER, N. P. PALAIO, M. RODDER, W. L. HANSEN and E. KREYSA, in "Neutron Transmutation Doping of Semiconductor Materials," edited by R. L. Larrabee (Plenum Press, 1984) p. 21; E. E. HALLER, *Infrared Phys. Technol.* **35** (1994) 127.
210. Y. OOTUKA, H. MATSUOKA and S. KOBAYASHI, in "Anderson Localization," edited by T. Ando and H. Fukuyama (Springer–Verlag, Berlin, 1988) p. 40.
211. K. M. ITOH, E. E. HALLER, J. W. BEEMAN and W. L. HANSEN, *Phys. Rev. Lett.* **77** (1996) 4058.
212. I. SCHLIMAK, M. KAVEH and R. USSYSHKIN, *ibid.* **77** (1996) 1103.
213. I. SCHLIMAK, *Fiz. Tverd. Tela* **41** (1999) 794 (in Russian).
214. R. RENTZSCH, A. N. IONOV, CH. REICH, V. GINODMAN and I. SCHLIMAK, *ibid.* **41** (1999) 837 (in Russian).
215. A. N. IONOV, M. N. MATVEEV and D. V. SHMIK, *J. Techn. Phys.* (*St-Petersburg*) **59** (1989) 169 (in Russian).
216. I. SCHLIMAK, A. N. IONOV, R. RENTZSCH and J. M. LAZEBNIK, *Semicond. Sci. Technol.* **11** (1996) 1826.
217. "Table of Isotopes," 6th ed., edited by C. M. Lederer, J. M. Hollander and I. Perlman (Wiley and Sons Inc., New York, 1967).
218. B. I. SHKLOVSKII and A. L. EFROS, in "Electronic Properties of Doped Semiconductors" (Solid State Series, Springer–Verlag, Berlin, 1984) Vol. 45.
219. E. E. HALLER, W. L. HANSEN and F. S. GOULDING, *Adv. Phys.* **30** (1981) 93.
220. H. FRITZSCHE, "The Metal Non-Metal Transition in Disordered Systems," edited by L. R. Friedman and D. P. Tunstall (Scotish Universities Summer School in Physics, St. Andrews, Scotland, 1978).
221. N. F. MOTT, "Metal-Insulator Transition," 2nd ed. (Taylor and Francis, London, 1990).
222. P. A. LEE and T. V. RAMAKRISHNAN, *Rev. Mod. Phys.* **57** (1985) 287.
223. R. RENTZSCH, A. N. IONOV, CH. REICH and A. MÜLLER, *Phys. Stat. Solidi* (b) **205** (1998) 269.
224. T. F. ROSENBAUM, K. ANDRES, G. A. THOMAS and F. N. BHATT, *Phys. Rev. Lett.* **45** (1980) 1723.
225. P. F. NEWMAN and D. F. HOLCOMB, *Phys. Rev.* B **28**, (1983) 638; *Phys. Rev. Lett.* **51** (1983) 2144.
226. W. N. SHAFARMAN, D. W. KOON and T. G. CASTNER, *Phys. Rev.* B **40** (1989) 1216.
227. A. N. IONOV, M. J. LEA and R. RENTZSCH, *JETP Lett. (Moscow)* **54** (1991) 473.
228. PEIHUA DAI, Y. ZHANG and M. P. SARACHIK, *Phys. Rev. Lett.* **66** (1991) 1914.
229. G. A. THOMAS, Y. OOTUKA, S. KATSUMOTO, S. KOBAYASHI and W. SASAKI, *Phys. Rev.* B **25** (1982) 4288.
230. M. J. HIRSCH, U. THOMANSCHEFSKY and D. F. HOLCOMB, *ibid.* **37** (1988) 8257.
231. A. G. ZABRODSKII and K. N. ZINOV'EVA, *Sov. Phys. JETP* **59** (1984) 425.
232. M. ROHDE and H. MICKLITZ, *Phys. Rev.* B **36** (1987) 7572.
233. G. HERTEL, D. J. BISHOP, E. G. SPENCER and R. C. DYNES, *Phys. Rev. Lett.* **50** (1983) 743.
234. W. L. MCMILLAN and J. MOCHEL, *ibid.* **46** (1981) 556.
235. TH. ZINT, M. ROHDE and H. MICKLITZ, *Phys. Rev.* B **41** (1990) 4831.
236. E. W. HASS and M. S. SCHNÖLLER, in [24] p. 803.
237. N. F. MOTT, *Proc. Phys. Soc. (London)* A **62** (1949) 416.
238. P. W. ANDERSON, *Phys. Rev.* **109** (1958) 1492.
239. A. MACKINNON and B. KRAMER, *Phys. Rev. Lett.* **47** (1981) 1546; M. HENNEKE, B. KRAMER and T. OHTSUKI, *Europhys. Lett.* **27** (1994) 389.
240. E. HAFSTETTER and M. SCHREIBER, *Phys. Rev. Lett.* **73** (1994) 3137.
241. T. KAWARABAYASHI, T. OHTSUKI, K. SLEVIN and Y. ONO, *ibid.* **77** (1996) 3593.
242. J. CHAYES, L. CHAYES, D. S. FISHER and T. SPENCER, *ibid.* **54** (1986) 2375.
243. "Neutron Transmutation Doping of Semiconductor Materials," edited by R. D. Larrabee (Plenum Press, New York–London, 1984).

244. A. N. IONOV, I. S. SHLIMAK and M. N. MATVEEV, *Solid State Commun.* **47** (1983) 763.
245. D. CHATTOPADHYAY and H. J. QUEISSER, *Rev. Mod. Phys.* **53** (1981) 745.
246. W VON AMMON, *Nucl. Instrum. Methods* B **63** (1992) 95.
247. K. M. ITOH, W. WALUKIEWICZ, H. D. FUCHS, J. W. BEEMAN, E. E. HALLER and V. I. OZHOGIN, *Phys. Rev.* B **50** (1994) 16995.
248. K. ITOH, W. L. HANSEN, E. E. HALLER, J. W. FARMER and V. I. OZHOGIN, *Mater. Sci. Forum* **117/118** (1993) 117.
249. K. M. ITOH, W. L. HANSEN, E. E. HALLER, J. W. FARMER and V. I. OZHOGIN, *J. Mater. Res.* **8** (1993) 127.
250. C. ERGINSOY, *Phys. Rev.* **79** (1950) 1013.
251. A. I. ANSEL'M, *Zh. Eksp. Teor. Fiz. (Moscow)* **24** (1953) 85 (in Russian).
252. N. SCLAR, *Phys. Rev.* **104** (1956) 1559; **104** (1956) 1548.
253. T. C. MCGILL and R. BARON, *ibid.* B **11** (1975) 5208.
254. B. K. RIDLEY, "Quantum Process in Semiconductors," 3rd ed. (Clarendon Press, Oxford, 1993).
255. R. D. DINGLE, *Philos. Mag.* **46** (1955) 831.
256. H. BROOKS, *Adv. Electron Phys.* **7** (1955) 85.
257. W. SHOCKLEY, "Elextrons and Holes in Semiconductors" (Van Nostrand Reinhold, Princeton, 1950).
258. J. BLAKEMORE, "Semiconductor Statistics," 2nd ed. (Dover, New York, 1985).
259. V. G. DZHAKELI and Z. S. KACHLISHVILI, *Sov. Phys. Semicond.* **18** (1984) 926.
260. J. W. FARMER and J. C. NUGENT, in [159] p. 225.
261. M. J. HILL, P. M. VAN ISEGHEM and W. ZIMMERMAN, in [24] p. 809.
262. A. MUHLBAUER, F. SELDAK and P. VOSS, *J. Electrochem. Soc.* **122** (1975) 1113.
263. J. MESSIER, Y. LE CORROLER and J. M. FLORES, *IEEE Trans. Nucl. Sci.* **NS-11** (1964) 276.
264. P. V. VAN ISEGHEM, *IEEE Trans. on Electron Devices* **ED-23** (1976) 823.
265. International Atomic Energy Agency: Regulations for the Safe Transport of Radioactive Materials, Rev. Ed., 1973, p. 8.
266. H. HAMANAKA, K. KURIYAMA and M. YAHAGI, *Appl. Phys. Lett.* **45** (1984) 786.
267. D. W. KOON and T. G. CASTNER, *Phys. Rev. Lett.* **60** (1988) 1755.
268. P. DAI, Y. ZHANG and M. P. SARCHIK, *Phys. Rev.* B **49** (1994) 14039.
269. H. GRUSSBACH and M. SCHREIBER, *ibid.* B **51** (1995) 663.
270. P. DAI, Y. ZHANG and M. P. SARCHIK, *Phys. Rev. Lett.* **70** (1993) 1968.
271. S. B. FIELD and T. F. ROSENBAUM, *ibid.* **55** (1985) 522.
272. R. RENTZSCH, K. J. FRIEDLAND and A. N. IONOV, *Phys. Stat. Solidi* (b) **146** (1988) 199.
273. R. MAGERLE, A. BURCHARD, M. DEICHER and T. KERLE, *Phys. Rev. Lett.* **75** (1995) 1594.
274. K. KURIYAMA and K. SAKAI, *Phys. Rev.* B **53** (1996) 987.
275. K. KURIYAMA, Y. MIYAMOTO, T. KOYAMA and O. OGAWA, *J. Appl. Phys.* **86** (1999) 2352.
276. K. KURIYAMA, K. OHBORA and M. OKADA, *Solid State Commun.* **113** (2000) 415.
277. S. M. SZE, "Physics of Semiconductor Devices" (Wiley, New York, 1969).
278. H. WINSTON, H. KIMURA, M. H. YOUNG and O. J. MARSH, in AAGG/West Sixth Conf. on Crystal Growth, 1982, Fallen Leaf Lake, CA.
279. "Table of Isotopes," edited by C. M. Lederer and V. S. Shirley, (Wiley, New York, 1978).
280. D. PONS and J. C. BOURGOIN, *J. Phys. C: Solid State Phys.* **18** (1985) 3839.
281. Properties of Gallium Arsenide, EMIS Data Reviews Series No. 2 (INSPEC, The Institute of Electrical Engineers, London, 1990).
282. L. PAVESI and M. GUZZI, *J. Appl. Phys.* **75** (1994) 4779.
283. A. MOORADIAN and A. L. MSWORTER, *Phys. Rev. Lett.* **19** (1967) 849.
284. T. SEKINE, K. UCHINOKURA and E. MATZUURA, *J. Phys. Chem. Solids* **48** (1977) 109.
285. T. KAWAKUBO and M. OKADA, *J. Appl. Phys.* **67** (1990) 3111.
286. K. KURIYAMA, Y. MIYAMOTO, M. OKADA, *ibid.* **85** (1999) 3499.
287. P. J. DEAN, in "Progress in Solid State Chemistry," edited by J. O. McCaldin and G. Somorjai (Pergamon Press, Oxford, 1973) Vol. 8, p. 1.
288. H. ALAWADHI, R. VOGELGESANG, T. P. CHIN and J. M. WOODALL, *J. Appl. Phys.* **82** (1997) 4331.
289. C. C. KLICK and J. H. SHULMAN, in "Solid State Physics," edited by F. Seitz and D. Turnbull (Academic Press, New York, 1957) Vol. 5, p. 100.
290. H. KRESSEL (ed.), "Semiconductor Devices for Optical Communications," Vol. 39 (Topics in Applied Physice, Springer-Verlag, Berlin-Heidelberg–New York, 1982).
291. P. C. BECKER and M. R. X. DE BARRAS, in "Materials for Optoelectronics," edited by M. Quilec (Kluwer Academic Publishers, Boston–Dordrecht–London, 1996).
292. A. W. SNYDER and J. D. LOVE, Optical Waveguide Theory (Chapman and Hall Medical, London–New York–Melbourne, 1996).
293. D. MARCUSE, "Light Transmission Optics" (Van Nostrand, New York, 1972).
294. W. B. ALLAN, "Fibre Optics Theory and Practice" (Plenum Press, New York, 1973).
295. N. S. KAPANY, "Fiber Optics" (Academic Press, New York–London, 1967).
296. J. A. ARNAUD, Beam and Fibre Optics (Academic Press, New York, 1976).
297. D. MARCUSE, Theory of Dielectric Optical Waveguides (Academic Press, New York, 1974).
298. J. E. MIDWINTER, Optical Fibers for Transmission (Wiley and Sons, New York–Toronto, 1979).
299. I. E. TAMM, "Theory of Electricity" (Science, Moscow, 1957) (in Russian).
300. M. BORN AND E. WOLF, "Principles of Optics" (Pergamon Press, Oxford, 1970).
301. R. E. COLLIN, "Field Theory of Guided Waves" (McGraw-Hill, New York, 1960).
302. E. SNITZER, *J. Opt. Soc. Amer.* **51** (1961) 491.
303. D. GLOGE, *Appl. Optics* **10** (1971) 2252.
304. J. A. STRATTON, "Electromagnetic Theory" (McGraw-Hill, New York, 1941).
305. P. MOON and D. E. SPENCER, "Field Theory Handbook" (Springer–Verlag, Berlin, 1961).
306. S. D. PERSONICK, *Bell Syst. Techn. J.* **52** (1973) 843.
307. R. OLSHANSKY and D. B. KECK, *Appl. Optics* **15** (1976) 483.
308. *idem.*, in "Technical Digest of OSA Topical Meeting an Optical Fiber Transmission" (Optical Society of America, Washington, DC, 1975).
309. "Materials for Optoelectronics," edited by M. Quillec (Kluwer Academic Publishers, Boston–Dordrecht–London, 1996).
310. TH. BECK, N. RENG and H. WEBER, *Opt. and Lasers Eng.* **34** (2000) 255; W. HEITMAN, *J. Opt. Commun.* **8** (1987) 3.
311. H. R. PHILIPP, *J. Phys. Chem. Solids* **32** (1971) 1935.
312. S. T. PANTELIDES and W. HARRISON, *Phys. Rev.* B **13** (1976) 2667.
313. N. F. MOTT, in "Physics of SiO_2 and Its Interfaces," edited by S. T. Pantelides (Proc. Int. Conf., New York, 1978) p. 1.
314. D. L. GRISCOM, *J. Non-Crystal. Solids* **40** (1980) 211; D. L. GRISCOM, E. L. FRIEBELE and K. J. LONG, *J. Appl. Phys.* **54** (1983) 3743.
315. A. U. GRINFELDS, D. E. ABOLTYN and V. G. PLEKHANOV, *Sov. Phys. Solid State* **26** (1984) 1075; *Phys. Stat. Solidi* (a) **81** (1984) K23.
316. R. P. GUPTA, *Phys. Rev.* B **32** (1985) 8278.
317. M. HORIGUCHI, *Electron. Lett.* **12** (1976) 311.
318. H. OSONAI, T. SHIODA, T. MORIYAMA, S. ARAKI, M. HORIGUCHI, T. IZAWA and H. TAKATE, *Electron. Lett.* **12** (1976) 550.
319. D. A. PINNOW, T. C. RICH, F. W. OSTERMAYER and M. DIDOMENICO, JR., *Appl. Phys. Lett.* **22** (1973) 527.
320. J. SCHROEDER, R. MOHR, P. B. MOCEDO and C. J. MONTROSE, *J. Amer. Ceram. Soc.* **56** (1973) 510.

321. K. J. BEALES, J. E. MIDWINTER, G. R. NEWNS and C. R. DAY, *Post Off. Elec. Eng. J.* **67** (1974) 80.
322. B. SCOTT and H. RAWSON, *Glass Technol.* **14** (1973) 115.
323. K. J. BEALES, C. R. DAY, W. J. DUNKAN, J. E. MIDWINTER and G. R. NEWNS, *Proc. IEEE* **123** (1976) 591.
324. T. INOUE, K. KOIZUMI and Y. IKEDA, *ibid.* **123** (1976) 577.
325. A. M. ZHELTIKOV, *Uspekhi Fiz. Nauk (Moscow)* **170** (2000) 1203 (in Russian).
326. "Lasers Theory and Applcations," edited by K. Thyagarajan and A. K. Ghatak (Plenum Press, New York, 1982).
327. O. SVELTO, "Principles of Lasers," 2nd ed. (Plenum Press, New York, 1982).
328. L. V. TARASOV, "Laser Physics" (Mir Publishers, Moscow, 1983).
329. N. V. KARLOV, "Lectures on Quantum Electronics" (Science, Moscow, 1983) (in Russian).
330. A. EINSTEIN, *Mitt. Phys. Ges. (Zurich)* **16**(18) (1916) 47.
331. F. REIF, "Statistical Thermal Physics" (McGraw-Hill, New York, 1965).
332. M. S. BRODIN and V. YA. REZNITCHENKO, in "Physics A_2B_6 Compounds," edited by A. N. Georgabiani and M. K. Sheinkman (Science, Moscow, 1986) p. 184 (in Russian).
333. A. J. TAYLOR, D. J. ERSKINE and C. L. TANG, *J. Opt. Soc. Amer.* (B) **2** (1985) 663.
334. J. A. KASH and J. C. TSANG, in "Light Scattering in Solids," edited by M. Cardona and G. Güntherodt (Springer–Verlag, Berlin, 1991) Vol. 6, p. 423.
335. W. P. DUMKE, *Phys. Rev.* **127** (1962) 1559.
336. R. E. NAHORY, K. L. SHAKLEY, R. F. LEHENY and R. A. LOGAN, *Phys. Rev. Lett.* **27** (1971) 1647.
337. H. KRESSEL, in "Laser Handbook," edited by F. T. Arechi and E. O. Schulz-Dubois (North-Holland, Amsterdam, 1972) Chap. B5.
338. F. STERN, *ibid*, Chap. B4.
339. C. KLINGSHIRN, in "Spectr. Soli-State Laser Type Matter" (Proc. Course Enrico Fermi, Erice, New York–London, 1987).
340. C. A. KLEIN, *Appl. Optics* **5** (1966) 1922; *IEEE* **QE-4** (1968) 186.
341. G. LASHER and F. STERN, *Phys. Rev.* **133** (1964) A553.
342. H. HAUG, *J. Appl. Phys.* **39** (1968) 4687; *Adv. Solid State Phys.* **XXII** (1982) 149.
343. W. VAN ROOSBROECK and W. SHOKLEY, *Phys. Rev.* **94** (1954) 1558.
344. C. J. HWANG, *ibid.* B **2** (1970) 4126.
345. H. HAUG and S. KOCH, *Phys. Stat. Solidi* (b) **82** (1977) 531.
346. K. C. LIU and R. L. LIBOFF, *J. Appl. Phys.* **54** (1983) 5633.
347. P. W. ANDERSON, "Concepts in Solids" (Benjamin, Reading, MA, 1963).
348. H. HAKEN, "Handbuch der Physik" (Springer–Verlag, Berlin, 1970) Vol. XXV/2c.
349. *idem*., in "Quantum Optics," edited by S. M. Kay and A. Maitland (Academic Press, New York, 1970).
350. R. L. LIBOFF, *Int. J. Theor. Phys.* **18** (1979) 185.
351. A. W. SCHAWLOW and C. H. TOWNES, *Phys. Rev.* **112** (1958) 1940.
352. R. S. KNOX, "Theory of Excitons" (Academic Press, New York, 1963).
353. E. F. GROSS, "Selected Papers" (Science, Leningrad, 1976) (in Russian).
354. D. G. THOMAS, (ed.), "II–VI Semiconducting Compounds" (Benjamin, New York, 1967).
355. N. G. BASOV, O. V. BOGDANKEVICH and A. G. DEVYATKOV, *Sov. Phys. JETP* **20** (1964) 1902; *Sov. Phys. Solid State* **8** (1966) 1221.
356. J. R. PACKARD, D. A. CAMPBELL and W. C. TAIT, *J. Appl. Phys.* **38** (1967) 5255.
357. C. BENOIT A LA GUILAUME, J. M. DEBEVER and F. SALVAN, in [354] p. 609; *Phys. Rev.* **169** (1967) 567.
358. L. A. KULEVSKY and A. M. PROKHOROV, *IEEE* **QE-2** (1966) 584.
359. M. S. BRODIN, K. A. DMITRENKO, S. G. SHEVEL and L. V. TARANENKO, in Proc. Int. Conf. Lasers'82 (STS Press, USA, 1983) p. 287.
360. M. S. BRODIN, S. V. ZAKREVSKI, V. S. MASHKEVICH and V. YA. REZNITCHENKO, *Sov. Phys.-Semicond.* **1** (1967) 495.
361. V. G. PLEKHANOV, in Proc. Int. Conf. LASERS'80 (McClean, VA:STS, 1981) p. 91.
362. K. TAKIYAMA, M. I. ABD-ELRAHMAN, T. FUJITA and T. OKADA, *Solid State Commun.* **99** (1996) 793.
363. V. G. PLEKHANOV and V. I. ALTUKHOV, in Proc. Int. Conf. LASERS'82 (McClean, VA:STS, 1983) p. 292.
364. C. KLINGSHIRN and H. HAUG, *Phys. Reports* **70** (1981) 315.
365. N. G. BASOV, V. A. DANILYCHEV and YU. M. POPOV, *JETP Lett.* **12** (1970) 473 (in Russian).
366. N. SCHWENTER, O. DOSSEL and N. NAHME, in "Laser Techniques for Extreme Ultraviolet Spectroscopy" (APTP, New York, 1982) p. 163.
367. V. G. PLEKHANOV, in Proc. Int. Conf. NLO'88, Dublin, 1988; Prog. in Solid State Chem. **29**(2001) 71.
368. T. RUF, M. CARDONA, K. THONKE, P. PAVONE and T. R. ANTHONY, *Solid State Commun.* **105** (1988) 311.
369. V. G. PLEKHANOV, *Quan. Electron. (Moscow)* **16** (1989) 2156 (in Russian).
370. V. G. PLEKHANOV, T. A. BETENEKOVA and F. F. GAVRILOV, *Sov. Phys, Solid State* **25** (1983) 159.
371. R. W. ROESSLER and W. C. WALKER, *J. Phys. Chem. Solids* **28** (1967) 1507.
372. M. PIACENTINI, *Solid State Commun.* **17** (1975) 697.
373. V. G. PLEKHANOV, in Proc. Int. Conf. Advances Solid-State Lasers, March, 1990, (Salt Lake City, UT, SOQUE, 1990).
374. V. F. AGEKYAN, B. G. ALEXANDROV and YU. A. STEPANOV, *Fiz. Tekh. Poluprovod.* **22** (1988) 2240 (in Russian).
375. A. A. BEREZIN, *Solid State Commun.* **65** (1988) 819.
376. A. A. BEREZIN, J. S. CHANG and A. IBRAHIM, *Chemtronix (UK)* **3** (1988) 116.
377. H. WERHEIT, in "Landolt-Börnstein," edited by O. Madelung (Springer-Verlag, Berlin, 1983) Vol. 17e, p. 9.
378. A. A. BEREZIN, *J. Chem. Phys.* **80** (1984) 1241; *Kybernetes* **15** (1986) 15.
379. H. W. KNOFF, M. H. MUELLER and L. HEATON, *Acta Cryst.* **23** (1967) 549.
380. T. HIDAKA and K. OKA, *Phys. Rev.* B **35** (1987) 8502.
381. A. V. ANTONOV, N. V. GALANOV and A. I. ISSAKOV, *Sov. Phys. Tech. Phys.* **31** (1986) 942.
382. V. A. BELYAKOV and S. V. SEMENOV, *JETP* **90** (2000) 290.
383. B. E. KANE, *Nature* **393** (1998) 133.
384. M. A. NIELSEN and I. L. CHUANG, "Quantum Computation and Quantum Information" (Cambridge University Press, Cambridge, 2000).

Received 28 January
and accepted 13 May 2003