ADVANCES IN PHYSICS, 1978, VOL. 27, NO. 6, 799–892

Heavily doped semiconductors and devices

By R. A. ABRAM†, G. J. REES and B. L. H. WILSON
Plessey Research (Caswell) Ltd., Allen Clark Research Centre,
Caswell, Towcester, Northamptonshire, England

[Received 9 August 1978]

ABSTRACT

High carrier concentrations and the fluctuations in random potential resulting from ionized impurities alter the density of states and electron wavefunctions in heavily doped semiconductors. The theoretical and experimental work on these effects is reviewed, special attention being paid to the consequences for the optical and transport properties. Semiconductor devices are often heavily doped and the influence of heavy doping is illustrated in the injection laser and bipolar transistor, which also serve as investigative tools.

CONTENTS

† Present address: Department of Applied Physics and Electronics, University of Durham, Durham, England.

§ 1. Introduction

Most semiconductor devices use material deliberately doped with impurities which in isolation are known to form shallow states in the band gap for electrons or holes. For the purpose of device design and development a thorough understanding of the basic physical processes is clearly desirable and the current use of high impurity concentrations in many devices has presented interesting and challenging problems. We have used the rather imprecise term ' heavily doped ' to describe semiconductors in which the doping is so high that the simple picture of individual impurity states is inadequate to describe relevant properties. In particular we consider in some detail the doping levels at which the impurity band and the parent band have merged. This is sometimes called the very heavily doped regime.

The first part of the review is concerned with the fundamental theory of heavily doped semiconductors. During the past 20 years considerable progress has been made in the theory. Using a number of seminal papers as a framework we discuss the theory, with emphasis on work which promises to be the most useful for quantitative calculations on devices. Our aim is to make clear the basic assumptions, simplifying approximations, range of validity and general relevance of various theories. We concentrate on the physical ideas and results with the view that the fine detail is best obtained from the original papers. Most attention is given to three semiconductors : gallium arsenide, silicon and germanium. The purpose of the second part of the paper is to demonstrate how the basic theory can be used to understand device performance and, conversely, to show how the device can be used as a tool to study the physical processes.

As well as serving as a general review of heavily doped semiconductors, we hope that this paper will help the device physicist to increase his understanding of heavy doping effects, while drawing the attention of other physicists to some interesting and technologically important semiconductor problems. As the review involves a number of recurrent themes, several of which are usually involved in any particular physical situation, we now briefly indicate the modifications to semiconductor properties brought about by heavy doping, and some of the phenomena which may be interpreted in terms of these modifications.

As the doping concentration is increased the effects of the fluctuating potential arising from the randomly situated impurities becomes apparent. An impurity band, formed by the overlap of wavefunctions on single impurities, has an asymmetrical low-energy tail of states bound to deep potential fluctuations from close clusters of impurities. At higher densities the impurity band merges with the parent band and eventually the picture becomes one of a tail of localized states joined to the parent band of extended states which are scattered by the ionized impurity potential. The absence of order in the impurity potential means that $\mathbf{k}$ is no longer a good quantum number. At energies high in the band this spread in $\mathbf{k}$ values can be described by a mean free path, but deep in the tail localization is complete. The doping concentrations N over which these regimes occur are determined by the extent to which impurities or the carriers they donate can ' see ' one another, an extent often described by the ratio of the effective Bohr radius a to the inter-impurity separation $N^{-1/3}$.

Even in the absence of the disordered impurity potential, the extra carriers would change the band structure. The exchange energy with the extra carriers and the modified screening of the valence electron exchange energy both reduce the energy gap and alter the band edge effective masses, although the mass change is small and often only the rigid band gap narrowing is discussed.

Heavy doping can also give rise to the incorporation of species other than the simple impurity, often involving vacancies or vacancy complexes, and to structural disorder. Strain is often generated by heavy doping and may determine the solubility limit of the impurity. However, the detailed discussion of strain and its effect on the electronic states is beyond the scope of this review.

Optical emission and absorption in heavily doped semiconductors involve all the above effects, and measurement of spectral distribution and transients can give much useful information. The changes in the distribution of states in energy alter the spectral range of inter-band optical transitions. Also, the transition probabilities are affected by modifications to the wavefunction : the loss of $\mathbf{k}$ as a good quantum number and the limited spatial overlap of states localized at the conduction and valence band edges are important here. The classical assumption of low occupation of band states may be invalid, particularly for bands of low effective mass, giving rise to a shift of the absorption edge (the Moss–Burstein effect). A high density of both types of carrier may give stimulated emission (and lasing) in suitable materials, perturbing the thermal distribution of carriers. In any case, the tail states may not be in equilibrium with the rest of the band. These radiative processes, and others such as the dielectric response of the carrier plasma, also affect the refractive

index and have important applications to light-guides and lasers where guiding is determined by local variations of carrier concentration. Non-radiative competing processes, for example the Auger effect and processes associated with structural defects, increase with doping.

Modifications to the density of states will also affect the minority carrier concentration, and hence the current–voltage characteristic of p–n junction diodes. There is often a gradation of impurity concentration, causing the minority carriers to be subjected to fields associated with the slope of the minority band edge and the changing density of states. The injection ratio of holes and electrons in p–n junctions, important for bipolar transistors, lasers and light emitting diodes, may be greatly affected.

Section 2 gives a general description of impurity states and the impurity band which remains separate from the host band at low doping levels. Sections 3 and 4 describe the effects on the band structure of the free carriers and randomly distributed impurity centres in heavily doped material. This work lays the foundation for the sections on the optical properties including radiative recombination and a brief discussion of Auger recombination (§ 5) and transport properties (§ 6). We concentrate on two representative devices: the GaAs laser (§ 7) and the silicon bipolar transistor (§ 8). For both devices a simple introduction to the basic principles of operation is given; but more information on these devices and others influenced by heavy doping can be obtained from Carroll (1974) and Sze (1969).

In common with the majority of the original papers we use c.g.s. units in this review, and the electronic charge, wherever it appears, should be expressed in e.s.u. The Appendix shows how each numbered equation may be translated into SI units.

§ 2. Impurity states

2.1. *Preliminary comments*

Since it is widely assumed in the literature discussed here that impurity states may be described in terms of the effective mass of a parabolic, isotropic band and a Coulombic potential, we now consider to what extent this approach is justified. Effective Hamiltonian theory is expected to be valid only if the perturbing potential is slowly varying, on the scale of the lattice, in the region where the wavefunction is appreciable (see, for example, Ziman 1972, p. 177). This condition is met for the excited states of many substitutional impurities in zinc-blende semiconductors where calculated energy levels agree with experiment (Kohn 1957, Faulkner 1969), but the ground states need special treatment. However, if proper consideration is given to the degeneracy of the valence band edge, the spin split-off band and the wave-vector dependence of the dielectric function, then the theory can account for the ionization energies of acceptors which are isocoric with the host semiconductor (from the same row of the periodic table), even when doubly and triply ionized (Baldereschi and Lipari 1976, and private communication).

For non-isocoric impurities it is known that central cell corrections can produce chemical shifts of the ground-state energy. The effects increase with effective mass. Light masses, for example the conduction band edge in GaAs, produce shallow, loosely bound ground states of a Coulomb potential and the

central cell part, considered as a first-order perturbation, is expected to have little effect on the energy because the probability that the bound particle occupies the central cell is small. The reverse is true in the case of the heavy masses at the band edges of Si, Ge, and GaP, for example, and the chemical shifts are marked. In this case the assumption of a Coulomb potential is clearly invalid and because of the rapid variation of the core potential the validity of the effective Hamiltonian approach is in doubt.

Pantelides and Sah (1974) have shown in an extension of this theory how the central cell corrections can be described in the language of pseudo-potentials. The effective Hamiltonian has a kinetic energy part, which represents the pseudo-band-structure of the host lattice, identical to the real band structure, and a potential energy part equal to the difference in pseudo-potentials of the impurity and host atom which it replaces. The latter can be further split into a long-range Coulomb term, which appears in the old treatment, and the central cell correction which now appears as a difference of the core potentials, each projected onto the space of their valence core states. Numerical results (Pantelides 1974) obtained with this scheme are encouraging.

It seems that with suitable refinements effective Hamiltonian theory can describe 'shallow' states adequately and the basic assumption that the potential is slowly varying appears to be generally satisfied. It is not expected that these methods will be suitable for describing the 'deep states' arising from, for instance, transition metal impurities, which are more characteristic of the impurity than the host, nor those of complexes involving vacancies and interstitials which give rise to strong lattice relaxation. Fortunately, the impurities which are used to dope semiconductors heavily for the purpose of research or device manufacture are generally those which give rise to shallow levels at a light doping concentration, although the act of heavy doping may introduce complexes.

We can now assess the consequences of the general simplifying assumptions described earlier. The replacement of the anisotropic or multiply degenerate band structure by a single isotropic effective mass adjusted to give the correct ionization energy ignores the anisotropy of the wavefunction or its vector nature. Berggren (1973) has considered the tight-binding impurity band formed by donors in Si and Ge, taking account of the anisotropy in the effective mass. The results differ only quantitatively from those obtained by assuming spherical bands. Baldareschi and Lipari (1973) have calculated the binding energies in zinc-blende semiconductors, taking full account of the degenerate band edge structure. The binding energies are more representative of the mass of the heavy hole bands than of the light hole bands. We conclude that the replacement of an anisotropic or degenerate band edge structure by an isotropic effective mass is unlikely to produce a change of a qualitative nature.

Ignoring the central cell correction may be more serious. For example, the states deep in the band tail produced by disorder (see § 4) are strongly localized and will feel the effects of the central cell. Moreover, their short-range wavefunction is likely to follow from a short-range potential whose rapid variation cannot be treated by effective Hamiltonian theory. However, the treatment of levels deep in the band tails is beset by other difficulties concerning the statistics of the impurity distribution and the free carrier screening, so the shortcomings discussed above may not be the most serious.

In the simple scheme of an isotropic effective mass m and static dielectric constant ϵ, the wavefunction of an independent Coulombic donor has the hydrogenic form

$$\phi(\mathbf{r}) = (\pi a^3)^{-1/2} \exp(-r/a), \tag{2.1}$$

with a Bohr radius

$$a = \hbar^2\epsilon/me^2 \quad (= 0{\cdot}53\epsilon(m_0/m)\ \text{Å}), \tag{2.2}$$

where m_0 is the free electron mass. The state lies at an energy

$$E_d = me^4/2\hbar^2\epsilon^2 \quad (= (13{\cdot}6/\epsilon^2)(m/m_0)\ \text{eV}), \tag{2.3}$$

below the conduction band edge. Typically, a is tens of angstroms and E_d tens of meV. As the donor concentration is increased the coupling of the individual degenerate donor states must be considered. To obtain a unified view of different semiconductors it is convenient to specify the concentration N by the dimensionless quantity $N^{1/3}a$. (The parameter r_s commonly used in many-body theory is given by $r_s = (3/4\pi)^{1/3}(N^{1/3}a)^{-1}$.

Most of the theoretical work on disorder effects has assumed that the impurity centres are distributed randomly. Because the average impurity separation is generally several lattice constants, it is a good approximation to ignore the discreteness of the lattice. However, at high doping densities the range of the screened impurity potential can exceed the impurity separation. If the material is prepared by some high-temperature process that allows the impurity ions to move, it is possible that some correlation of the position of the impurities occurs which is frozen in at lower temperatures. This problem has been raised periodically in the literature (see, for example, Kane 1963, Volkov and Matveev 1966, Rogachev and Sablina 1966, Wilson 1977), but has not received much detailed attention. Keldysh and Proshko (1964) and Stern (1971 b) have suggested the use of an ionic contribution to the potential screening as a simple method of describing the effect. In this paper we will assume a random distribution unless the contrary is stated.

We look first at the problem of an impurity band which is separate from the host conduction band and initially ignore electron–electron and electron–phonon effects.

2.2. *One-electron tight-binding model*

The Schrödinger equation for the one-electron problem is

$$-\frac{\hbar^2}{2m}\nabla^2\psi(\mathbf{r}) + \sum_i V(\mathbf{r}-\mathbf{R}_i)\psi(\mathbf{r}) = E\psi(\mathbf{r}), \tag{2.4}$$

where $\sum_i V(\mathbf{r}-\mathbf{R}_i)$ is the combined potential of the donors at positions $\mathbf{R}_i$. Equation (2.4) describes an electron moving in a spatially random potential and, in principle, a solution could be obtained starting from this general viewpoint. However, at low densities an apposite approach is the tight-binding method (Callaway 1964).

We try a solution which is a linear combination of donor orbitals :

$$\psi(\mathbf{r}) = \sum_i a_i\phi(\mathbf{r}-\mathbf{R}_i), \tag{2.5}$$

where $\phi(\mathbf{r})$ is given by eqn. (2.1). If, for the dilute system, the $\phi(\mathbf{r}-\mathbf{R}_i)$ for different centres are assumed to be orthogonal† on account of the large donor separation, then eqns. (2.4) and (2.5) lead to the eigenvalue equation

$$(E_d - E)a_j - \sum_k V_{jk}a_k = 0, \tag{2.6}$$

where, in the two-centre approximation,

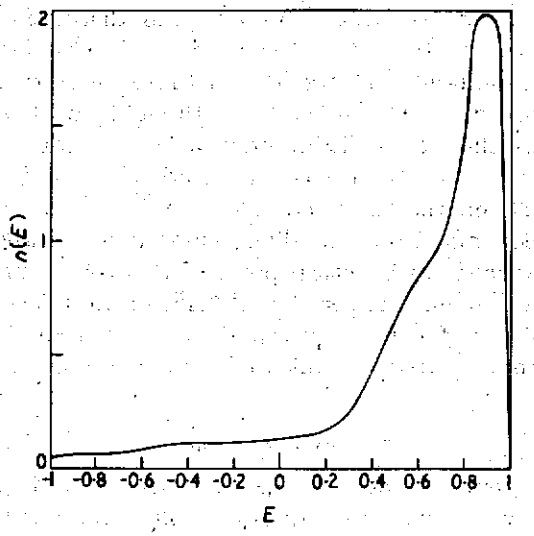
$$V_{jk} = -2E_d\left(1+\frac{r_{jk}}{a}\right)\exp\left(-\frac{r_{jk}}{a}\right),$$

with

$$r_{jk} = |\mathbf{R}_j - \mathbf{R}_k|.$$

Both analytical and numerical techniques have been used to obtain the distribution of eigenvalues—the impurity band density of states (see, for example, Matsubara and Toyozawa 1961, Gaspard and Cyrot Lackmann 1973, 1974, Debney 1977). The density of states obtained by Gaspard and Cyrot Lackmann for donors distributed randomly and $N^{1/3}a = 0{\cdot}54$ is shown in fig. 1. There are significant differences from the band shape given by a periodic array of centres : there are no Van Hove singularities and there is a long low-energy tail which is due to states centred on donor clusters with the coefficients a_j having approximately the same phase at each centre. For $N^{1/3}a \ll 1$ the main structural feature is a peak near the donor binding energy—E_d, but with

Fig. 1.



The density of states per atom for the impurity band when $N^{1/3}a = 0{\cdot}54$. The unit of energy is $2E_d$ (after Gaspard and Cyrot-Lackmann 1973).

† The donor orbitals are not strictly orthogonal, and the assumption breaks down completely for donors which are particularly close. Because of the disorder it is not easy to estimate the limiting concentration at which this approach loses validity. The problem has been discussed in some detail by Debney (1976).

increasing donor concentration the peak moves up in energy and the tail lengthens.

Our present understanding of the impurity band wavefunctions is due largely to the work of Anderson (1958) and several papers by Mott†. In the absence of spatial periodicity the wavenumber cannot be used to label the states and only the energy remains a useful quantum number. For $N^{1/3}a \ll 1$ the states are localized ; that is, each eigenfunction has an appreciable amplitude only in a finite region of space. At sufficiently low concentrations the states closely resemble individual donor states ; at higher concentrations the amplitudes a_j of a particular eigensolution of eqn. (2.6) are appreciable for a number of centres in a bounded region.

As $N^{1/3}a$ is increased further, the localized states at the centre of the band grow in spatial extent until, at approximately 0·25–0·33 (Debney 1977) some of the states have a finite amplitude throughout the system. Because of the disorder, these extended states have wavefunctions with spatial fluctuations of amplitude and phase. With increasing concentration the extended states represent a greater proportion of the states in the band, but localized states always exist at the periphery. Two energies E_c and E'_c separate the localized and extended states. As either of these energies is approached from the localized side, the states grow in spatial extent, becoming infinite at the transition energy. In contrast to the description presented above, Economou and Antoniou (1977) have suggested recently that not all the states are localized, even in the dilute limit. However, their view is based on a model related, but not directly applicable, to the impurity band problem (see also Weaire and Srivastava 1977).

The d.c. conductivity provides a fundamental distinction between the localized and extended states. An electron in a localized state has zero mobility, whereas the extended states have a finite mobility limited by the scattering due to disorder. The mobility is thought to change sharply at E_c and E'_c, which are therefore called mobility edges. Indeed, Mott has long held the view that the change is discontinuous and has concluded that if the Fermi energy lies just on the extended side of an edge, the conductivity is given by the so-called minimum metallic conductivity of about $0{\cdot}1e^2N^{1/3}/\hbar$ (Mott 1974). No conclusive theoretical proof of this conjecture has yet been derived, although there is much experimental evidence in its favour (Mott *et al.* 1975, Pepper 1976) and, for the equivalent expression in two-dimensional systems, evidence from computer studies (see, for example, Licciardello and Thouless 1975).

2.3. *Impurity band*

It has been convenient to neglect the electron–electron and electron–phonon interactions, but they do play an important role ; for example, transport between localized states becomes possible by phonon activation. Furthermore, we have treated the impurity states in a model which does not explicitly include the conduction band, and in due course we consider the possible overlapping of these bands. For the moment we assume that they remain separate.

† See, for example, Mott and Twose (1961), Mott (1970, 1972, 1973 a, 1974), Mott and Davis (1971) and references therein.

At room temperature the impurity band does not play a significant role in the conductivity. Even the extended states have low mobility compared to states in the conduction band to which the electrons are readily excited. However, the impurity states may influence other semiconductor properties, such as optical properties (see § 5). At lower temperatures, when activation to the conduction band does not dominate, impurity conduction can be seen and is a valuable investigative technique.

For $N^{1/3}a \ll 1$ the zero-temperature impurity wavefunctions resemble those of the individual donor states. The conclusions of the one-electron treatment are not changed by including many electron effects ; electron–electron repulsion encourages the electrons to avoid each other and remain on the donor sites. Indeed, for $N^{1/3}a \ll 1$ the electron repulsion would produce this sort of localization even in a crystalline array of centres (Mott 1974). This description is supported by a range of experiments including E.P.R., Knight shift and transport (Mott 1974). There is always some overlap of the donor wavefunctions which allows transport to take place if energy is supplied by phonons. Energy is required to place the electron on an already occupied centre, and it follows that the conductivity has the activated form proportional to $\exp(-\Delta E/kT)$, where ΔE is roughly the difference between E_d and the electron affinity of the occupied donor (Mott and Davis 1971). The doubly occupied donor states have been studied using far infra-red photoconductivity techniques (Norton 1976 a, b).

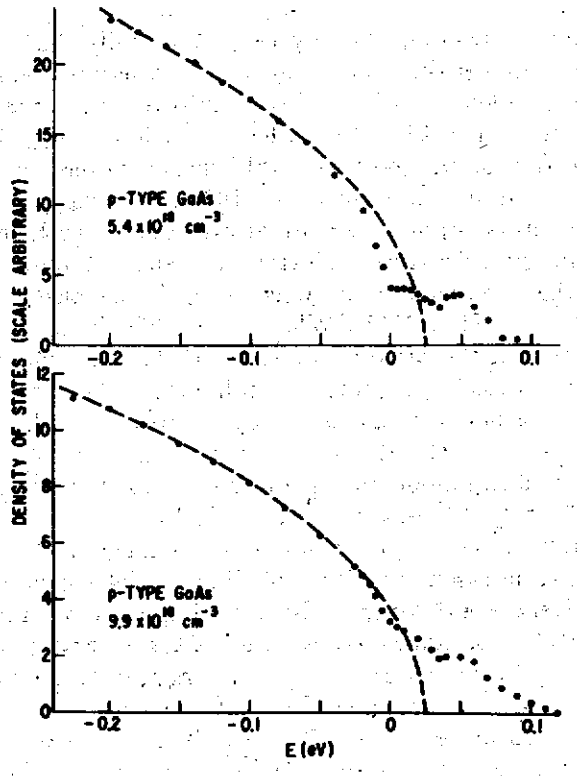
The electron–electron effects may be substantially removed by compensation which depopulates the donor states. The spatially varying field of the acceptors also broadens the donor levels into a band, the bandwidth from overlap being small for $N^{1/3}a \ll 1$. The activation energy for conduction is then representative of the most probable energy difference between nearest-neighbour states (Miller and Abrahams 1960). At very low temperatures a more likely process is hopping to more distant donors with small state overlap but smaller energy difference ; Mott (1969) showed how this leads to the logarithm of the conductivity varying as $T^{1/4}$ (see also Ambegaokar *et al.* 1971, Shklovskii and Éfros 1971, Pollak 1972). The band shape is particularly easy to predict when the random potential of the acceptors is slowly varying over the extent of each donor state. It follows that the band shape is a direct replica of the local potential at each donor. However, this situation is only likely to pertain when the donor states have a very small spatial extent. Morgan (1965) used this model to explain the shapes of optical spectra involving deep levels in heavily doped, compensated GaAs.

We now return to the uncompensated semiconductor. As the donor concentration is increased, the tight-binding model suggests localized states which extend over several donors. By spreading its wavefunction the electron can, in the one-electron approximation, lower its energy, but this leads to an overlap of states which implies a higher energy contribution from electron–electron interactions. However, it does not necessarily follow that the many-body effects act as a localizing influence. The problem is one involving a subtle interplay of many electron and disorder effects. The system can be investigated experimentally by the magnetic and activated conductivity methods mentioned earlier (Mott 1974). Above a value of $N^{1/3}a$ estimated to be between 0·25 and 0·33, the impurity band exhibits conductivity at zero

temperature because the states at the Fermi level are extended. This estimate is based on a combination of experimental results (Alexander and Holcomb 1968, Mott 1974), metal–insulator theory in ordered arrays (Berggren 1973, Mott 1974) and one-electron models of disorder (Debney 1977). Mott (1974) has discussed the behaviour of the impurity states near the critical concentration in some detail, and the interested reader is referred to this work and the references therein.

The one-electron model gives the impurity band shown in fig. 1 which is derived from a calculation based on the assumption that there exist impurity states which do not mix with the conduction band states. However, the calculations of Matsubara and Toyozawa (1961), also based on this assumption, show that the upper edge of the impurity band enters the unperturbed conduction band at $N^{1/3}a \simeq 0{\cdot}18$ and the Fermi level (no compensation) enters at $N^{1/3}a \simeq 0{\cdot}43$; hence a separate band model is not tenable at these doping levels. In any case, the approximation of orthogonal basis wavefunctions loses validity as the impurity concentration increases. Furthermore, the many-body effects have been neglected ; in particular, the electrons screen the donor potentials and a self-consistent calculation is necessary. The original basis states of the

Fig. 2



The density of states of p-type GaAs deduced from Schottky junction tunnelling experiments. The dashed curves are fitted parabolae (after Mahan and Conley 1967).

tight-binding calculation become more weakly bound, and eventually become unbound, as a result of screening. The physical picture is of merging impurity and conduction bands, and this is exemplified in the Schottky barrier tunnelling experiments of Mahan and Conley (1967). Figure 2 shows the derived valence–impurity band shape of p-type GaAs at doping levels of $5{\cdot}4\times10^{18}$ cm^{-3} ($N^{1/3}a\simeq0{\cdot}23$) and $9{\cdot}9\times10^{18}$ cm^{-3} ($N^{1/3}a\simeq0{\cdot}28$). The impurity band hump is clearly seen at the lower concentration and a band tail is seen to have developed at the higher concentration.

When the remnant impurity band still makes a significant contribution to the density-of-states function of the merged bands, a satisfactory quantitative theory does not appear to exist, and some workers have used the expedient of a superposed impurity band and conduction band, each obtained in separate calculations (see § 8). The absence of a suitable theory is unfortunate since the relevant doping levels are found in some semiconductor devices. However, progress has been made, when $N^{1/3}a$ is close to or greater than unity, by adopting the more direct approach of considering each electron to interact with the random impurity potential $\sum_i V(\mathbf{r}-\mathbf{R}_i)$ and the other electrons from the beginning. It is interesting to note that, by virtue of the semiconductor dielectric constant and the effective mass, values of $N^{1/3}a$ exceeding those typical of the constituent ions of metals (0·1–0·3) are attainable in many materials. This fact suggests the use of many-body theories developed for the electron gas problem. Wolff (1962), realizing this, showed how many-body and disorder effects could be treated simultaneously in a perturbation calculation. However, it is now apparent that the disorder effects cannot be treated satisfactorily by perturbation theory and more recent, improved theories of disorder effects have not been able to incorporate the many-body aspects fully. In §§ 3 and 4 we describe the present state of the theory for high impurity concentrations.

§ 3. Many-body effects at high impurity concentrations

3.1. *Introduction*

The many-body interactions alone appear, from theory and experiment, to have important effects on the band structure, including shrinking the band-gap and perturbing the carrier effective mass. This is particularly clear from the work on electron–hole droplets (Rice 1977, Hensel *et al.* 1977) which can be produced in some semiconductors by optical excitation. However, because the droplets contain large concentrations of both electrons and holes in the absence of a random impurity potential, the theory of droplets is not directly relevant for our purposes. For the heavily doped semiconductor we start by describing the many-body effects by ignoring disorder in a jellium model of a direct-gap, parabolic band semiconductor. As a basis we use the recent theory of Inkson (1976) which embraces its antecedents as well as introducing a new effect. The methods of treating disorder are then considered, with the many-body effects included, where possible, in § 4.

3.2. *Basic method*

Inkson (1976) has calculated the effect on the semiconductor band-gap of an electron gas at zero temperature which occupies part of the conduction band.

The calculation employs the self-energy formalism of many-body theory (Hedin and Lundqvist 1969) and is the most general approach to the problem so far attempted. The method may be used to examine other effects on the band structure, such as the many-body changes in the Fermi level or effective masses. For the intrinsic semiconductor, the one-electron Green's function is given by solving

$$(\hbar\omega - H)G(\mathbf{r}, \mathbf{r}', \omega) - \int \Sigma(\mathbf{r}, \mathbf{r}'', \omega)G(\mathbf{r}'', \mathbf{r}', \omega)\, \mathrm{d}\mathbf{r}'' = \hbar\delta(\mathbf{r} - \mathbf{r}'), \qquad (3.1)$$

where H represents the one-electron Hamiltonian and the direct average electron–electron interaction. The exchange and correlation effects of the electrons in the system are represented by the non-local, energy-dependent self-energy operator $\Sigma(\mathbf{r}, \mathbf{r}', \omega)$. To first order,

$$\Sigma(\mathbf{r}, \mathbf{r}', \omega) = (i/2\pi) \int G(\mathbf{r}, \mathbf{r}', \omega - \omega')W(\mathbf{r}, \mathbf{r}', \omega') \exp(-i\delta\omega')\, \mathrm{d}\omega', \qquad (3.2)$$

where $W(\mathbf{r}, \mathbf{r}', \omega)$ is the dynamically screened Coulomb interaction whose spatial Fourier transform may be written in terms of the longitudinal dielectric function as

$$W(\mathbf{q}, \omega) = 4\pi e^2/q^2\epsilon(\mathbf{q}, \omega) \qquad (3.3)$$

(ignoring local field effects (Adler 1962)). With the introduction of electrons into the conduction band (together with a neutralizing uniform positive background charge to represent the ionized donors), the operator $\Sigma(\mathbf{r}, \mathbf{r}', \omega)$ changes, leading to shifts in the energies of the electron states. To obtain the change in the band-gap, Inkson examines the relative movement in energy of states at the bottom of the conduction band and the top of the valence band.

The ω' integral in eqn. (3.2) gives a term from the poles in the Green's function and a term from the poles in the interaction. The latter term is essentially electrostatic in origin, deriving from the polarization of the electron gas or ' Coulomb hole ' around a given electron. Inkson claims that this term affects both bands equally, and so has no influence on the band-gap (Inkson 1973, Brinkman and Goodman 1966, Kane 1972), which is not surprising in view of its electrostatic nature. The term from the poles of the Green's function gives the so-called ' screened dynamic exchange ' contribution

$$\Sigma_{ex}(\mathbf{r}, \mathbf{r}', \omega) = -\sum_{j \text{ occupied}} \phi_j(\mathbf{r})\phi_j^*(\mathbf{r}')W(\mathbf{r}, \mathbf{r}', \omega_j - \omega). \qquad (3.4)$$

Here the $\phi_j(\mathbf{r})$s are strictly the quasi-particle wavefunctions of the semiconductor, but in a first approximation may be taken as the Bloch eigenfunctions obtained in a suitable one-electron calculation. It is worth noting that if the ω dependence of the interaction is ignored and $W(\mathbf{r}, \mathbf{r}', 0)$ used for all ω in eqn. (3.4), then Σ_{ex} reduces to the statically screened Hartree–Fock exchange potential.

An examination of eqn. (3.4) shows that, when electrons are added to the conduction band, $\Sigma_{ex}(\mathbf{r}, \mathbf{r}', \omega)$ changes for two reasons. Firstly, the screening of the Coulomb potential is changed because the dielectric function is altered by the electron gas. Secondly, more states are occupied and the sum over j extends to the occupied states in the conduction band. In a first-order calculation these two changes may be separated to provide additive changes to Σ_{ex}, and consequently to the state energies.

If the change in Σ_{ex} is given by $\Delta\Sigma_{ex}(\mathbf{r}, \mathbf{r}', \omega)$, then the first-order energy shift in the lowest conduction band state at $E_{c0} = \hbar\omega_{c0}$ is

$$\Delta E_{c0} = \int \phi_{c0}^*(\mathbf{r})\Delta\Sigma_{ex}(\mathbf{r}, \mathbf{r}', \omega_{c0})\phi_{c0}(\mathbf{r}')\, d\mathbf{r}\, d\mathbf{r}', \tag{3.5}$$

where $\phi_{c0}(\mathbf{r})$ is the state wavefunction. In similar notation the valence band edge state is shifted by

$$\Delta E_{v0} = \int \phi_{v0}^*(\mathbf{r})\Delta\Sigma_{ex}(\mathbf{r}, \mathbf{r}', \omega_{v0})\phi_{v0}(\mathbf{r}')\, d\mathbf{r}\, d\mathbf{r}'. \tag{3.6}$$

The change in the band-gap is

$$\Delta E_g = \Delta E_{c0} - \Delta E_{v0}. \tag{3.7}$$

In line with the discussion above, Inkson calculates ΔE_{c0}, for example, in two parts :

$\Delta E_{c0}^{(1)}$, the change in E_{c0} due to the change in the interaction with the summation in eqn. (3.4) remaining over the intrinsic occupancy, and

$\Delta E_{c0}^{(2)}$, the change in E_{c0} due to the summation over the newly occupied states using the new interaction.

For the valence band, ΔE_{v0} may be divided in a similar fashion.

3.3. *Calculation of* $\Delta E_{c0}^{(1)}$ *and* $\Delta E_{v0}^{(1)}$

The change in the interaction is determined by the modification to the dielectric function due to the electron gas. For the intrinsic material (Pines 1963, Adler 1962) the random phase approximation to the longitudinal dielectric function is

$$\epsilon_{int}(\mathbf{q}, \omega) = 1 + \frac{4\pi e^2}{q^2\Omega} \sum_{ijk,\,spin} \frac{|\Omega_{cell}^{-1} \int_{cell} u_{i\mathbf{k}+\mathbf{q}}^*(\mathbf{r}) u_{j\mathbf{k}}(\mathbf{r})\, d\mathbf{r}|^2 (f(E_i(\mathbf{k}+\mathbf{q})) - f(E_j(\mathbf{k})))}{\hbar\omega - (E_j(\mathbf{k}) - E_i(\mathbf{k}+\mathbf{q}))}, \tag{3.8}$$

where Ω_{cell} and Ω are the unit cell and crystal volumes respectively and $u_{j\mathbf{k}}(\mathbf{r})$ is the periodic part of the Bloch function $\Omega^{-1/2}u_{j\mathbf{k}}(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r})$ of band j with energy $E_j(\mathbf{k})$ and occupancy $f(E_j(\mathbf{k}))$ for each spin state. When the electrons are present in the conduction band, Inkson approximates the dielectric function by

$$\epsilon(\mathbf{q}, \omega) = \epsilon_{int}(\mathbf{q}, \omega) + (\epsilon_g(\mathbf{q}, \omega) - 1), \tag{3.9}$$

where $\epsilon_g(\mathbf{q}, \omega)$ is the dielectric function of a free electron gas with the conduction band effective mass. There should be another part to $\epsilon(\mathbf{q}, \omega)$ in eqn. (3.9) which removes the terms in the sum for ϵ_{int} which correspond to the occupied states in the conduction band. Inkson ignores this contribution, judging it to be insignificant provided the electrons occupy a small part of the Brillouin zone.

The change in interaction, which follows from eqns. (3.3) and (3.9), is

$$\Delta W(\mathbf{q}, \omega) = \frac{4\pi e^2}{q^2} \left\{ \frac{1}{\epsilon_{int}(\mathbf{q}, \omega)} - \frac{1}{\epsilon_{int}(\mathbf{q}, \omega) + \epsilon_g(\mathbf{q}, \omega) - 1} \right\}. \tag{3.10}$$

$\Delta E_{v0}^{(1)}$ is obtained, by using this expression in conjunction with eqns. (3.4) and (3.6), as

$$\Delta E_{v0}^{(1)} = -\frac{\Omega}{(2\pi)^3}\int\left\{\int \phi_{v0}^*(\mathbf{r})\phi_{v\mathbf{k}}(\mathbf{r})\Delta W(\mathbf{q}, \omega_v(\mathbf{k})-\omega_v(\mathbf{0})) \times \exp(i\mathbf{q}\,.\,(\mathbf{r}-\mathbf{r}'))\,\phi_{v0}(\mathbf{r}')\phi_{v\mathbf{k}}^*(\mathbf{r}')\,d\mathbf{r}\,d\mathbf{r}'\,\frac{d\mathbf{q}}{(2\pi)^3}\right\} d\mathbf{k}. \quad (3.11)$$

The sum over j in eqn. (3.4) has been replaced by the integral over $\mathbf{k}$ for the valence band.

To enable an analytical evaluation of eqn. (3.11), Inkson uses the Thomas–Fermi approximation

$$\epsilon_{int}(\mathbf{q}, \omega)+\epsilon_g(\mathbf{q}, \omega)-1=\epsilon(1+\kappa^2/q^2), \quad (3.12)$$

where $\epsilon=\epsilon_{int}(\mathbf{0}, 0)$ and $\kappa^2=(4\pi e^2\rho(E_F)/\epsilon)$, $\rho(E_F)$ being the density of states at the Fermi level. This approximation leads to substantial errors, and later we will improve it. In eqn. (3.11) the integrals over $\mathbf{r}$ and $\mathbf{r}'$ are complex conjugates of each other. If $\mathbf{q}$ is very much smaller than a reciprocal lattice vector, the periodic part of the Bloch functions is not important and ϕ_{vk} may be replaced by plane waves (Callaway 1964, p. 234). Completing the integrals, Inkson obtains

$$\Delta E_{v0}^{(1)}=e^2\kappa/\epsilon. \quad (3.13)$$

To find $\Delta E_{c0}^{(1)}$ it is necessary to consider a test electron at the bottom of the conduction band and its interaction with the valence band electrons. The expression for $\Delta E_{c0}^{(1)}$ is obtained from eqn. (3.11) by changing ϕ_{v0} to ϕ_{c0} and $\omega_v(\mathbf{0})$ to $\omega_c(\mathbf{0})$. Here the change in the interaction at frequencies close to the band-gap frequency is important. In most semiconductors the band-gap is sufficiently large to make $\epsilon_g(\mathbf{q}, \omega_g)\simeq 1$ at practicable doping levels. Therefore, from eqn. (3.11), $\Delta W\simeq 0$ and

$$\Delta E_{c0}^{(1)}=0. \quad (3.14)$$

It can be argued that the term omitted from eqn. (3.9)—the modification of the intrinsic dielectric function due to occupied states—does contribute a change to the interaction at energies around the band-gap. However, the change in ϵ_{int} is confined to small q, corresponding to the occupied states. For small q the integrals over $\mathbf{r}$ and $\mathbf{r}'$ in the modified form of eqn. (3.11) are small because the periodic part of the Bloch function for the valence band is orthogonal to that for the conduction band (Callaway 1964, p. 234).

3.4. *Calculation of* $\Delta E_{c0}^{(2)}$ *and* $\Delta E_{v0}^{(2)}$

To determine $\Delta E_{c0}^{(2)}$ the interaction of the electron at the bottom of the conduction band with the rest of the electron gas is considered. Using eqns. (3.4) and (3.5),

$$\Delta E_{c0}^{(2)} = -\frac{\Omega}{(2\pi)^3}\int\left\{\int \phi_{c0}^*(\mathbf{r})\phi_{c\mathbf{k}}(\mathbf{r})W(\mathbf{q}, \omega_c(\mathbf{k})-\omega_c(\mathbf{0})) \times \exp(i\mathbf{q}\,.\,\mathbf{r}-\mathbf{r}')\,\phi_{c0}(\mathbf{r}')\phi_{c\mathbf{k}}^*(\mathbf{r}')\,d\mathbf{r}\,d\mathbf{r}'\,\frac{d\mathbf{q}}{(2\pi)^3}\right\} d\mathbf{k}, \quad (3.15)$$

where W is the screened interaction of eqns. (3.3) and (3.9). Again, Inkson uses the Thomas–Fermi approximation to evaluate the integrals which in this context turns out, for reasons given later, to be quite accurate. Dealing with the Bloch functions in the same way as before, $\Delta E_{c0}^{(2)}$ reduces to the statically screened Hartree–Fock energy of the $k=0$ state of a free electron gas :

$$\Delta E_{c0}^{(2)} = -(2e^2/\pi\epsilon)[k_F - \kappa \arctan(k_F/\kappa)], \tag{3.16}$$

where k_F, the Fermi wave-vector, is given in terms of the electron density n by $k_F=(3\pi^2 n)^{1/3}$. $\Delta E_{v0}^{(2)}$ is taken to be negligibly small since the newly occupied states are at small k and the orthogonality argument used after eqn. (3.14) may be applied :

$$\Delta E_{v0}^{(2)} = 0. \tag{3.17}$$

Combining eqns. (3.13), (3.14), (3.16), (3.17) and (3.7), and indicating the Thomas–Fermi approximation by the superscript TF,

$$\Delta E_g^{TF} = -\frac{2e^2 k_F}{\pi\epsilon}\left[\frac{\pi}{2}\frac{\kappa}{k_F} + 1 - \frac{\kappa}{k_F}\arctan\left(\frac{k_F}{\kappa}\right)\right], \tag{3.18}$$

which is always negative.

The effects of the many-body interactions on other aspects of the band structure may also be determined by calculating the energy shifts for general values of $\mathbf{k}$. Indeed, the generalization of eqns. (3.15) and (3.16) has been derived by a number of authors (see, for example, Wolff 1962). For example, the use of $\Delta E_{ck}^{(2)}$ and $\Delta E_{v0}^{(1)}$ gives the valence-band-edge–Fermi-level-separation change. Also, the many-body modifications of the band effective masses may be determined. An examination of the relevant equations shows that the valence band mass is unaffected but the conduction band mass becomes m^*, given by

$$\frac{m^*}{m} = \left\{1 + \left(\frac{(4/3\pi)(k_F/\kappa)^4}{k_F a(1+(k_F/\kappa)^2)^2}\right)\right\} = \left\{1 + \frac{1}{7{\cdot}29 n^{1/3} a(1+0{\cdot}41/n^{1/3}a)^2}\right\}. \tag{3.19}$$

For $n^{1/3}a \geqslant 1$, the change is quite small. The $\kappa=0$ limit of eqn. (3.19) has been derived by Wolff, but there is a small numerical discrepancy between the results.

Naturally, similar calculations can be carried out for holes in the valence band. For the hole gas, $\Delta E_{v0}^{(1)}$ and $\Delta E_{v0}^{(2)}$ are the significant terms and the expression for ΔE_g^{TF} remains the same, with k_F and κ applying to the hole gas. The change in valence band effective mass is given by eqn. (3.19), n now being the hole concentration.

3.5. *Results and discussion*

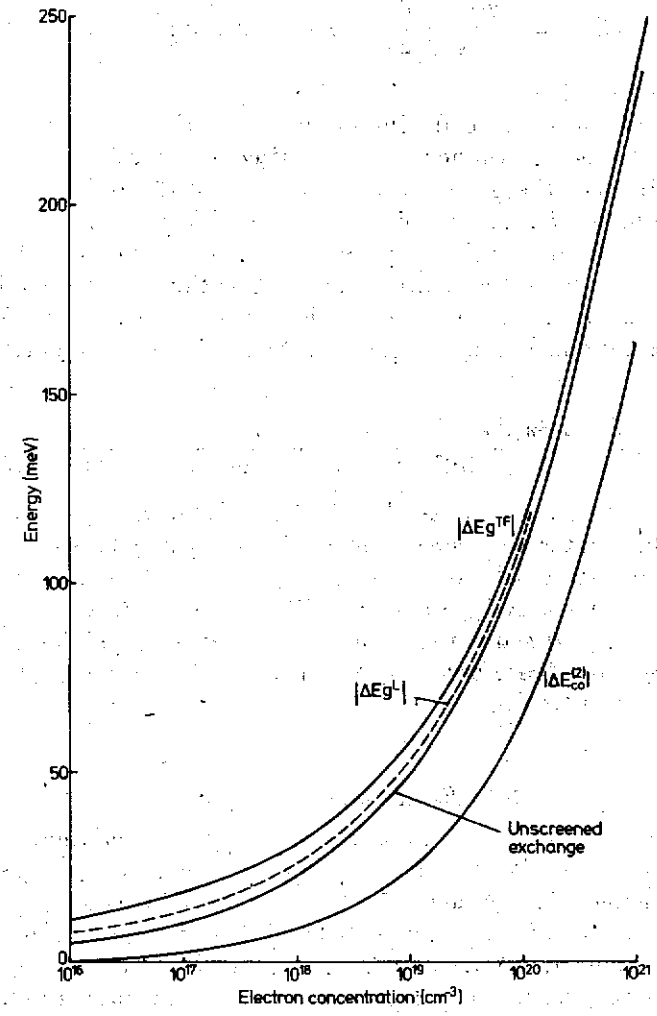
For definiteness we return to the electron gas calculation. We have described Inkson's calculation in some detail since it includes, and goes beyond, previous treatments of many-body effects. Earlier workers simply calculated the exchange lowering of the conduction band due to the electron gas and so obtained the component $\Delta E_{c0}^{(2)}$ or the unscreened limit ($\kappa=0$) of this. Haas (1962), for example, considering degenerate germanium, used the unscreened interaction for exchange in an anisotropic band. Wolff (1962) and Bonch-Bruevich (1966 a, b) obtained eqn. (3.16) when studying the many-body

changes to the E–k relation. These authors did not consider the Coulomb-hole term. In most calculations $\Delta E_{v0}{}^{(2)}$ is tacitly assumed to be zero, and indeed all effects on the valence band are generally ignored.

Inkson's new contribution to the change in band-gap is $\Delta E_{v0}{}^{(1)}$. This shift occurs because the intrinsic band-gap energy has a component derived from the exchange interaction between the valence band electrons. The electron gas changes the electron–electron interaction and hence the gap. The new term is a significant part of $\Delta E_g{}^{TF}$; indeed, for $(k_F/\kappa) \lesssim 2{\cdot}8$, $|\Delta E_{v0}{}^{(1)}|$ is greater than the electron gas exchange term $|\Delta E_{c0}{}^{(2)}|$. We have that

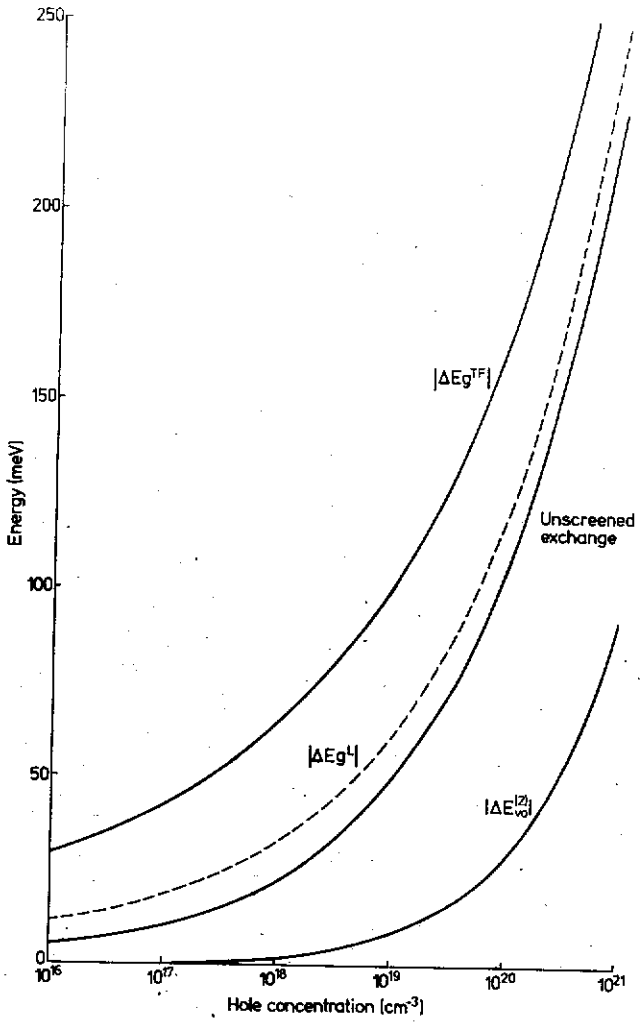
$$\left.\begin{aligned} &\kappa = 2{\cdot}73 \times 10^4 (m/m_0)^{1/2} (n^{1/6}/\epsilon^{1/2}) \text{ cm}^{-1}, \\ \text{and} \quad &k_F = 3{\cdot}094 n^{1/3} \text{ cm}^{-1}, \end{aligned}\right\} \tag{3.20}$$

Fig. 3



(*a*)

Fig. 3 (*continued*)



(*b*)

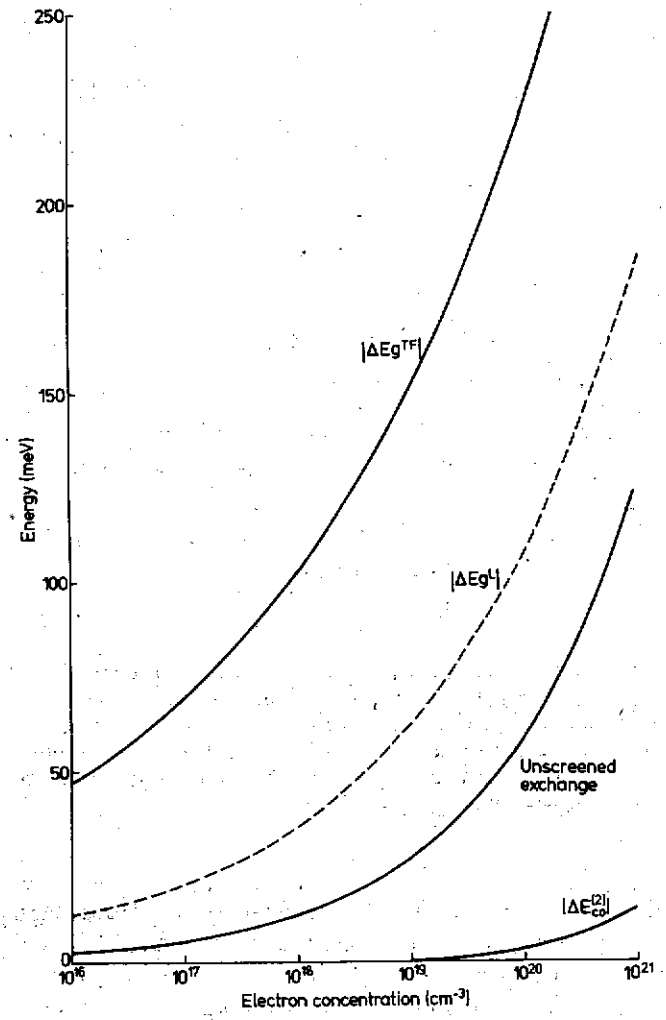
The theoretical carrier concentration dependence of band-gap narrowing for GaAs at 0 K ; (*a*) *n*-type, (*b*) *p*-type. $|\Delta E_g^{TF}|$, Thomas–Fermi calculation ; $|\Delta E_g^{L}|$, Lindhard calculation ; $|\Delta E_{c0}^{(2)}|$ or $|\Delta E_{v0}^{(2)}|$, screened exchange contribution. Also shown is the ' unscreened exchange ', the shift of the band edge due to the exchange interaction with free carriers in that band calculated without screening. $m_e/m_0=0{\cdot}066$, $m_h/m_0=0{\cdot}5$, $\epsilon=12{\cdot}5$.

so $\Delta E_v^{(1)}$ dominates at sufficiently low concentrations and high effective masses. The change with electron concentration of $|\Delta E_g^{TF}|$, $|\Delta E_{c0}^{(2)}|$ and $|\Delta E|$ for unscreened exchange are shown for GaAs in fig. 3 (*a*). For holes in GaAs the heavier hole mass leads to greater screening, which reduces $|\Delta E_{v0}^{(2)}|$ but increases $|\Delta E_g^{TF}|$, as shown in fig. 3 (*b*).

The calculation may be extended straightforwardly to indirect gap semiconductors. Numerical results for electrons and holes in silicon are presented in fig. 4. The many-valley conduction band with a fairly large mass makes the screening in n-type material particularly strong, and this is reflected in the disparity between $|\Delta E_g{}^{TF}|$ and $|\Delta E_{c0}{}^{(2)}|$. As expected, the results for p-type silicon are similar to those for gallium arsenide.

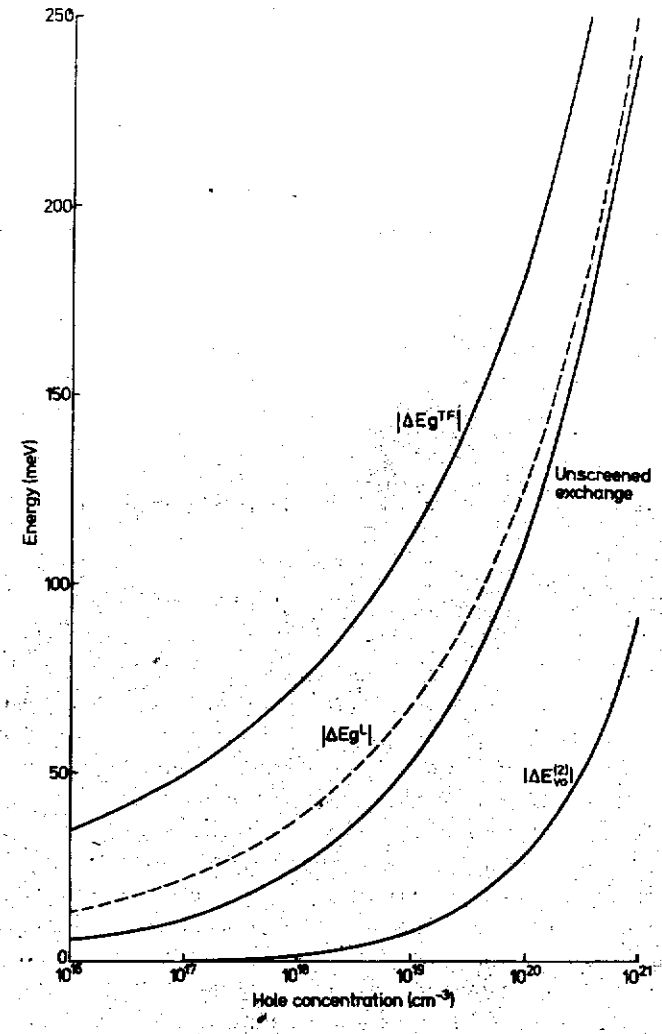
In fact, the values of $|\Delta E_g|$ obtained in these calculations are too large because the Thomas–Fermi approximation has been used for the dielectric function $\epsilon(\mathbf{k}, \omega)$. We have improved the accuracy by using the Lindhard dielectric function (Stern 1963, Hedin and Lundqvist 1969) for $\epsilon_g(\mathbf{k}, \omega)$ and calculating the integrals (3.11) and (3.15) numerically. It turns out that the Thomas–Fermi expression overestimates the screening for $k \gtrsim k_F$. This is not

Fig. 4



(*a*)

Fig. 4 (*continued*)



(*b*)

The theoretical carrier concentration dependence of band-gap narrowing for Si at 0 K ; (*a*) *n*-type, (*b*) *p*-type. $|\Delta E_g{}^{TF}|$, Thomas–Fermi calculation ; $|\Delta E_g{}^{L}|$, Lindhard calculation ; $|\Delta E_{c0}{}^{(2)}|$ or $|\Delta E_{v0}{}^{(2)}|$, screened exchange contribution. Also shown is the ' unscreened exchange ', the shift of the band edge due to the exchange interaction with free carriers in that band calculated without screening. $m_{el}/m_0=0{\cdot}97$, $m_{et}/m_0=0{\cdot}19$, $m_h/m_0=0{\cdot}55$, $\epsilon=11{\cdot}8$.

important for the integral (3.15), which involves only $k<k_F$, but the integral (3.11) is over all k space and large errors result from the Thomas–Fermi form for all practical values of k_F/κ. The results of our numerical calculations using the Lindhard dielectric function are shown in figs. 3 and 4 and show considerably smaller gap reductions than the Thomas–Fermi results in all but *n*-type GaAs. The curves of figs. 3 and 4 are derived using a single, isotropic valence

band. In reality exchange coupling of the light and heavy hole bands will change the valence band shift (Combescot and Nozières 1972), probably by a factor of approximately 0·75.

At least in an heuristic sense, the theory may be adapted to the finite temperature problem. The change in interaction used in eqn. (3.11) can again be written in the Thomas–Fermi form, with κ taking on the finite temperature form (see eqn. (4.4)). With this replacement $\Delta E_{v0}^{(1)}$ is given by eqn. (3.13) again. The integral for $\Delta E_{c0}^{(2)}$, eqn. (3.15), remains the same with the occupied states determined by the Fermi–Dirac distribution. In the limit of Boltzmann statistics, we obtain

$$\Delta E_{g(\mathrm{Boltz})}{}^{\mathrm{TF}} = -\frac{2e^2 k_{\mathrm{th}}}{\pi\epsilon}\left\{\frac{\pi}{2}\frac{\kappa_{\mathrm{B}}}{k_{\mathrm{th}}} + \frac{2\pi^2 n}{k_{\mathrm{th}}{}^3}\left[1-\sqrt{\pi}\,\frac{\kappa_{\mathrm{B}}}{k_{\mathrm{th}}}\,\exp\left(\kappa_{\mathrm{B}}{}^2/k_{\mathrm{th}}{}^2\right)\left[1-\mathrm{erf}\left(\frac{\kappa_{\mathrm{B}}}{k_{\mathrm{th}}}\right)\right]\right]\right\}, \quad (3.21)$$

where

$$k_{\mathrm{th}} = \left(\frac{2mk_{\mathrm{B}}T}{\hbar^2}\right)^{1/2} = 8{\cdot}24\times 10^6 (T/300)^{1/2}(m/m_0)^{1/2}\ \mathrm{cm}^{-1},$$

and

$$\kappa_{\mathrm{B}} = \left(\frac{4\pi e^2 n}{\epsilon k_{\mathrm{B}} T}\right)^{1/2} = 8{\cdot}36\times 10^{-3}(n/\epsilon)^{1/2}(300/T)^{1/2}\ \mathrm{cm}^{-1}. \quad (3.22)$$

Although eqn. (3.21) is a useful guide to the behaviour of ΔE_g in the Boltzmann limit, accurate quantitative results require a better description of the dielectric function, just as was necessary in the degenerate case. Hedin and Lundqvist (1969) have discussed the formal extension of the self-energy method to finite temperatures, and the reader is referred to this review and the references therein for a more rigorous approach to the problem.

Inkson's work introduces some new ideas which deserve further consideration. The expression for ΔE_g is obtained in a first-order calculation and could be improved in a numerical calculation. Our use of the Lindhard dielectric function has increased the accuracy of the results considerably, but more refined treatments of the dielectric function are possible. The coupling of the light and heavy hole valence bands by the exchange interaction is expected to reduce the valence band shift slightly and should be included in any accurate calculation.

The whole calculation is based on the self-energy formalism and in particular on the first-order expression, eqn. (3.2), for Σ. This approximation improves with increasing density. However, Hedin and Lundqvist (1969) suggest a qualitative accuracy down to metallic electron densities of $n^{1/3}a \simeq 0{\cdot}1$. The quantitative accuracy at $n^{1/3}a \simeq 1$ is not known, but Hedin and Lundqvist point out that the self-energy technique using eqn. (3.2) is an improvement over the conventional Hartree–Fock approximation.

However, even with improvements in the many-body theory, one large difficulty remains which is common to most of the many-body theory of heavily doped semiconductors. The application of this many-body model is complicated by the effects of the impurity centres, which have been conveniently smeared into a background charge for the purposes of the calculation, and the theory can be sensibly applied only to doping densities at which the impurity

band and host band have merged. Furthermore, in this doping regime the random positioning of the impurity centres profoundly affects the one-particle states near the band edges (see § 4). Intuitively, it seems unlikely that disorder and many-body effects can be separated in any simple way. We have also neglected the influence of the carriers on any polaron contribution to the band structure (Mahan 1972). No such effects are expected in elements, and in the common polar semiconductors the ' large polaron ' effect is small (see, for example, Casey and Stern (1976) for a brief consideration of GaAs).

Because of disorder, the effects of many-body interactions cannot be unambiguously determined in most experiments. However, most of the results to be presented in subsequent sections suggest a band-gap narrowing that is a little less, but in fair agreement with, our improved evaluation of Inkson's theory.

§ 4. Disorder effects at high impurity concentrations

4.1. *General considerations*

The effects of disorder are now considered by replacing the uniform positive background charge of the electron gas with the true array of impurity centres. The electrons respond to screen the random impurity potential. In lightly doped materials at finite temperatures the electrons occupying the conduction band screen by adjusting their density, while those in donor states screen by preferentially occupying certain centres (see, for example, Falicov and Cuevas 1967). As the impurity band widens and begins to merge with the conduction band, the difficulties involved in a quantitative description become severe. However, a quantitative theory is possible at higher impurity densities. The theory is usually based on the approximation of linear screening : the screened potential at any point is then given by a superposition of individually screened impurity potentials. This approximation is valid provided the root-mean-square potential fluctuation is small compared to the Fermi level (or thermal energy when classical statistics are valid) (Halperin and Lax 1966, Hwang 1970 a).

Assuming linear screening, the electron–impurity potential may be written as

$$V(\mathbf{r}) = \sum_i \frac{-e^2}{\epsilon |\mathbf{r} - \mathbf{R}_i|} \exp\{-\kappa |\mathbf{r} - \mathbf{R}_i|\}, \tag{4.1}$$

where $\mathbf{R}_i$ are the impurity locations. The reciprocal screening length is given in the Thomas–Fermi approximation by

$$\kappa^2 = \frac{4\pi e^2}{\epsilon} \int \rho(E) \left[-\frac{\partial f(E)}{\partial E} \right] \mathrm{d}E, \tag{4.2}$$

where $\rho(E)$ is the band density of states. If the screening is due to electrons in several bands, or to electrons and holes, $\kappa^2 = \sum_i \kappa_i^2$ where κ_i^2 is calculated from eqn. (4.2) for each band. It should be noted that when $\rho(E)$ is calculated it will depend on κ through the potential $V(\mathbf{r})$, and κ should be obtained in a self-consistent calculation. However, κ is given to a good approximation by

using the unperturbed band density of states in the integral if the Fermi level is well above the band edge. Then, in degenerate conditions,

$$\kappa^2 = 4\pi e^2 \rho(E_F)/\epsilon = 4(3/\pi)^{1/3}(n^{1/3}/a) = 7{\cdot}45 \times 10^8 (m/m_0)(n^{1/3}/\epsilon) \text{ cm}^{-2}, \quad (4.3)$$

and, when Boltzmann statistics are applicable,

$$\kappa^2 = 4\pi e^2 n/\epsilon kT = 6{\cdot}99 \times 10^{-5}(300/T)(n/\epsilon) \text{ cm}^{-2}. \quad (4.4)$$

The exchange and correlation effects of the electron–electron interaction can be included only in a crude fashion. Accepting this, a suitably simplified problem is to solve the effective mass Schrödinger equation

$$-(\hbar^2/2m)\nabla^2\psi + V(\mathbf{r})\psi = E\psi, \quad (4.5)$$

with $V(\mathbf{r})$ defined by eqn. (4.1) and the effective mass m taking on the appropriate values for the conduction and valence bands. Equation (4.5) represents a one-particle problem with the many-body aspects implicit in the effective mass, band-gap, potential screening and exclusion principle. The interesting physical effects of the disorder come from the spatial fluctuation in the potential. The mean value of $V(\mathbf{r})$ is of no consequence to the band structure since it affects all states in all bands equally and is, in any case, cancelled by the mean potential of the free carriers.

The deeper fluctuations and fluctuation complexes are capable of binding electrons at energies below the intrinsic band edge. These are localized states in the sense described in § 2. Because the random potential has fluctuations of varying shape and size distributed throughout the semiconductor, a quasi-continuous distribution of the energy of localized states is expected. Large fluctuations are less likely than small ones, so a tail of states decreasing in density away from the band edge, as shown in fig. 2, is plausible. Indeed, the existence of band tails has been demonstrated in many experiments, some of which are referred to in § 4.6. Optical and tunnelling experiments give particularly direct evidence of tails.

The nature of the localized wavefunction is determined by the details of the potential $V(\mathbf{r})$ and the energy of the state. In general, we expect the states deep in the tail to be of small spatial extent, but as the energy moves up the tail the states increase in size. At some energy E_c the states become extended and for energies just above E_c the wavefunctions are expected to be wildly fluctuating in amplitude and phase on a length-scale given roughly by the correlation length of the potential. Much higher in the band the states are only weakly scattered by the disorder and resemble slightly perturbed conduction band states. The physical picture is essentially that proposed by Mott (1970); we will now describe some attempts to obtain a quantitative understanding of the heavily doped semiconductor.

Our earlier discussion of the tight-binding method demonstrated that it was inadequate for handling the high-density regime and so an alternative approach must be found. The use of various forms of perturbation theory using the intrinsic band states as a basis has been attempted by a number of workers; see, for example, Parmenter (1955), Wolff (1962), Bonch-Bruevich and Mironov (1962), Zvyagin (1963), and Rimbey and Mahan (1974). For states high in the conduction and valence bands the perturbation methods have proved useful and able to predict changes in the state parameters and band shapes. However,

for band-edge states the perturbation theories run into difficulties because the change in the states produced by the disorder cannot, in any sense, be regarded as small. The low-momentum particles are strongly scattered by the simultaneous interaction with several impurities. Indeed, it is such processes which produce the localized states of the tail.

To obtain a better description of the states near the band-edges it is necessary to allow the carriers to respond more completely to the random potential than appears to be possible with perturbation methods. To this end Kane (1963) and Bonch-Bruevich (1963) independently developed a semi-classical theory of the bands. Bonch-Bruevich (1963, 1966 a, b) has given a formal mathematical description of the method, but in the next section we adopt the physically intuitive view of Kane.

4.2. *Semi-classical theory*

If the potential $V(\mathbf{r})$ is sufficiently slowly varying, a semi-classical approximation to the band density-of-states function can be made. For definiteness we consider an electron in the conduction band of n-type material. In the semi-classical method a local density of states is assigned to each point $\mathbf{r}$ and is given by the free-electron density of states, assuming the potential to have the local value throughout the system. This approximation appears to be reasonable for states high in the band, but its validity is less clear for localized states. In these the electron is bound to negative fluctuations; the lowest level bound to a fluctuation will, in general, have a significant kinetic energy. Excited states bound to the fluctuation, if they exist, will be separated by finite intervals of energy. The semi-classical method will be valid if the kinetic energy of the lowest state and the energy separations of higher states are small compared to the magnitude of the potential energy. To estimate these quantities we examine a typical negative potential fluctuation.

From eqn. (4.1) it is clear that $V(\mathbf{r})$ has large contributions from impurities within a screening length of $\mathbf{r}$ and weaker contributions from more distant centres. The average number of impurities within a screening length of any point is

$$N_0 = 4\pi\kappa^{-3}N/3, \tag{4.6}$$

for a donor density N. If N_0 is large, it is reasonable to expect a typical negative fluctuation to be of radius $\simeq\kappa^{-1}$ and depth $N_0^{1/2}\,e^2\kappa/\epsilon$, the latter quantity being derived from a fluctuation of $N_0^{1/2}$ donors with charge smeared uniformly over a spherical volume of radius κ^{-1}. The kinetic energy of the lowest state in such a potential well is roughly $\hbar^2\kappa^2/2m$, and this also gives a rough value for the excited state separation. If the semi-classical approximation is to be good the potential energy must be larger than this kinetic energy. This is true if $4N^{1/2}/a\kappa^{5/2} \gg 1$ which, for degenerate conditions, reduces to $(n^{1/3}a)^{1/4} \gg 1$. The above condition does not guarantee the accuracy of the method because fluctuations differing greatly from our typical fluctuation occur, in some of which the kinetic energy of localization cannot be neglected, but $4N^{1/2}/a\kappa^{5/2}$ does act as a guide to the theory's general validity.

In the semi-classical approximation the local density of states is obtained at each elemental volume $\Delta\Omega$ by taking the unperturbed band density of states to be shifted in energy by the deviation of the local potential $V(\mathbf{r})$ from its mean

value V_0. (The reason for ignoring the mean potential V_0 was explained after eqn. (4.5).) So at $\mathbf{r}$ the number of states per unit energy range in $\Delta\Omega$ is $\rho(E, \mathbf{r})\Delta\Omega$, where

$$\rho(E, \mathbf{r}) = (\sqrt{2}m^{3/2}/\pi^2\hbar^3)(E - (V(\mathbf{r}) - V_0))^{1/2}, \tag{4.7}$$

with

$$V_0 = -4\pi e^2 N/\epsilon\kappa^2.$$

Each elemental volume is considered independently so the density of states for the whole system is obtained by summing over them. The procedure is equivalent to averaging $\rho(E, \mathbf{r})$ over the probability distribution of the local potential fluctuations :

$$\rho(E) = (\sqrt{2}m^{3/2}/\pi^2\hbar^3) \int_{-\infty}^{E} [E - (V - V_0)]^{1/2} P(V - V_0)\, \mathrm{d}(V - V_0). \tag{4.8}$$

The distribution $P(V - V_0)$ for the potential of eqn. (4.1) in the high-density limit is shown by Kane to be

$$P(V - V_0) = (\pi\eta^2)^{-1/2} \exp(-(V - V_0)^2/\eta^2), \tag{4.9}$$

where

$$\eta = (e^2/\epsilon)(4\pi N/\kappa)^{1/2}. \tag{4.10}$$

The density of states may then be written as

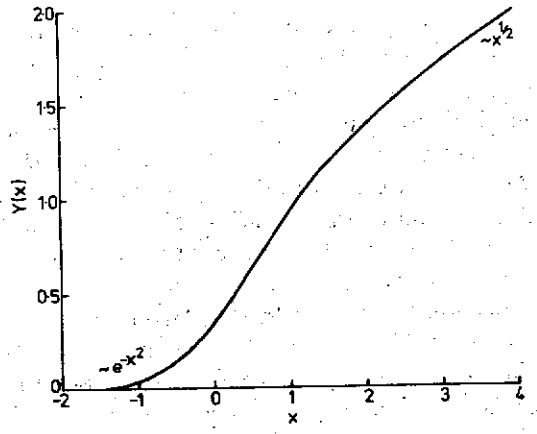
$$\rho(E) = (1/2\pi^2)(2m/\hbar^2)^{3/2}\eta^{1/2} Y(E/\eta), \tag{4.11}$$

where

$$Y(x) = \pi^{-1/2} \int_{-\infty}^{x} (x - \zeta)^{1/2} \exp(-\zeta^2)\, \mathrm{d}\zeta. \tag{4.12}$$

The function $Y(x)$ is plotted in fig. 5. For $E \to \infty$, $\rho(E) \sim E^{1/2}$, as in the unperturbed band, whereas for $E \to -\infty$, $\rho(E) \sim \exp(-E^2/\eta^2)$, the density of states reflecting the Gaussian statistics of the potential in the low-energy states which form a band tail.

Fig. 5



The function $Y(x)$ of eqn. (4.11).

The discussion has so far dealt with the conduction band, but the valence band is also affected by the donor centres. The analysis is essentially the same as for the conduction band. The random potential for holes is of opposite sign to the electron potential, but because of the symmetry of $P(V)$ the expression for the valence band density of states is identical to eqn. (4.11), except that the hole effective mass must be used. The extension to p-type material is also straightforward—the interaction now being with acceptors. In compensated semiconductors the random field of both impurity species must be considered ; however, if there is a low density of carriers the linear screening approximation will fail.

The condition $4N^{1/2}/a\kappa^{5/2} \gg 1$ is a basic requirement for the theory to be valid. For degenerate uncompensated material this implies the high-density limit $(n^{1/3}a)^{1/4} \gg 1$. However, it is to be expected that the theory remains useful for $n^{1/3}a \simeq 1$, which is more frequently encountered in practice. One point, quite apart from the semi-classical approximation, which must be considered at lower doping levels is the Gaussian approximation for $P(V)$. Equation (4.7) is valid for $|V - V_0| \ll |V_0|$ being essentially a Gaussian approximation to a Poisson distribution. For the same reason the true distribution is not symmetrical around V_0 and this can lead to a noticeable difference in the conduction and valence band tails.

To be more explicit, suppose that at some point there is a higher than average density of donors in n-type material. They will produce an electron potential energy more negative than the average V_0 (which is negative). There is no limit to the excess density and the magnitude of the negative fluctuations, except that demanded by the host lattice. However, at no point in the system will the electron potential energy exceed zero due to a smaller than average density of donors ; the maximum positive fluctuation is given by $|V_0|$. Using the correct potential distribution does not greatly affect the conduction band result, eqn. (4.11), since it is the negative fluctuations which produce the tail. These are well described by eqn. (4.9), except when $V \ll V_0$, and then the density of states is very low anyway. The positive fluctuations do not have much effect on the main body of the band. However, the positive fluctuations in electron potential energy are negative fluctuations of potential energy for holes and produce the valence band tail ; the above argument shows that the tail can only be of finite length. We expect these considerations to be important when N_0 in eqn. (4.6) drops to the order of unity. For degenerate material this implies that $(n^{1/3}a)^{3/2} \simeq 2$. Both Bonch-Bruevich and Kane have discussed the failure of Gaussian statistics ; the former author derived some analytic approximations to $\rho(E)$.

The corresponding arguments for p-type material can be developed in the same way. In compensated semiconductors the presence of both positive and negative centres can produce large potential fluctuations of both signs, but the precise details depend on the degree of compensation.

The semi-classical approximation is a fundamental limitation on the accuracy of the theory ; the method does not take account of the kinetic energy of localization. Intuitively this seems a reasonable omission for states above or just below the average potential of the system, but it cannot be justified for states deep in the tail. We envisage the deep states as having relatively large kinetic energies due to localization in small volumes of negative potential.

Because the semi-classical method neglects the kinetic energy effect, it places states at energies which are too low and tends to exaggerate the tail length.

Despite the inherent limitations of the method, it has proved to be a useful description of the band shape in disordered semiconductors. In a historical context it provided a germinal description from which a quantum-mechanical theory of the tail states could be developed. Furthermore, the theory still remains one of the simplest to apply and it provides a tolerably accurate single-function expression for $\rho(E)$ throughout the band.

4.3. *Quantum-mechanical theory of tail states*

Halperin and Lax (1966) presented a quantum-mechanical theory of the tail states. Although their method includes the kinetic energy properly, it is only valid for deep tail states. Furthermore, Halperin and Lax obtain explicit results only for a potential obeying Gaussian statistics. Again, we consider electrons in the conduction band of n-type material to illustrate the method.

A tail state at energy E can be produced by an infinite variety of negative potential fluctuations. If we imagine the state wavefunction to be significant only in a small volume of space, the electron has a high kinetic energy, and to produce a state at energy E the expectation value of the potential energy must be sufficiently negative, i.e. a deep fluctuation over the small volume is required. Alternatively, the wavefunction can be well spread out, so having associated with it a small kinetic energy, and then the expectation value of the potential energy must be close to E, i.e. a fluctuation close to E over a large volume is required. The statistics of the impurity potential determine a most probable shape for the wavefunction at energy E. Halperin and Lax argue that in an ensemble of semiconductors large deviations from this most probable shape are small. Furthermore, the probability of getting an impurity fluctuation of the necessary magnitude in a region the size of the optimum wavefunction is the dominant factor in the density of states as a function of energy.

The Gaussian potential has its probability distribution specified by its mean value, which is taken as zero, and by its two-point correlation function $\langle V(\mathbf{r})V(\mathbf{r}')\rangle$. Calling the latter quantity $W(\mathbf{r}-\mathbf{r}')$, we have that

$$W(\mathbf{r}-\mathbf{r}') = (2\pi e^4 N/\kappa\epsilon^2)\exp(-\kappa|\mathbf{r}-\mathbf{r}'|). \tag{4.13}$$

To treat the states quantum-mechanically, Halperin and Lax use a variational method. If the real, normalized function $f(\mathbf{r}-\mathbf{r}_0)$ is taken as a trial wavefunction, then the associated energy is

$$E(\mathbf{r}_0) = \int f(\mathbf{r}-\mathbf{r}_0)(-(\hbar^2/2m)\nabla^2 + V(\mathbf{r}))f(\mathbf{r}-\mathbf{r}_0)\,\mathrm{d}\mathbf{r}. \tag{4.14}$$

The $\mathbf{r}_0$ dependence of $E(\mathbf{r}_0)$ is contained in the potential energy part

$$V_s(\mathbf{r}_0) = \int f^2(\mathbf{r}-\mathbf{r}_0)V(\mathbf{r})\,\mathrm{d}\mathbf{r}\,; \tag{4.15}$$

$V_s(\mathbf{r}_0)$ fluctuates about a zero mean value as $\mathbf{r}_0$ varies. Halperin and Lax assume that an eigenstate of the system may be associated with each local minimum of $E(\mathbf{r}_0)$. In the spirit of the variational calculation, at these minima $E(\mathbf{r}_0)$ is close to but greater than the energy of the true eigenstate if $f(\mathbf{r}-\mathbf{r}_0)$ is a good choice. Furthermore, because $f(\mathbf{r}-\mathbf{r}_0)$ gives a higher energy than the true eigenvalue and $\rho(E)$ is expected to decrease rapidly and monotonically in

the tail, any density of states calculated will be smaller than the true value†. Therefore the best $f(\mathbf{r})$ may be chosen by maximizing the density of states. It is assumed that each state can be occupied by only one electron (Halperin, private communication) and the density of states is given by the number of local minima of $E(\mathbf{r}_0)$ which depends on the potential and $f(\mathbf{r})$ according to eqn. (4.14). Maximizing $\rho(E)$ with respect to the wavefunction gives

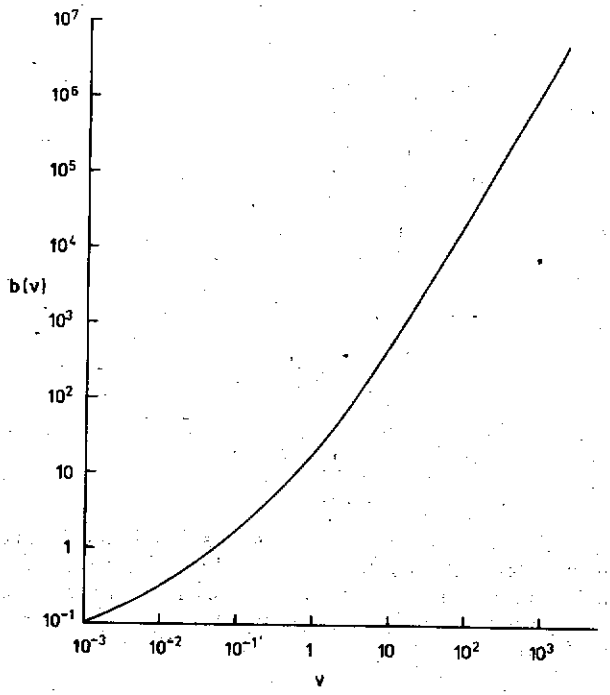
$$-(\hbar^2/2m)\nabla^2 f(r) - \mu f(r) \int f^2(r') W(r-r')\, dr' = Ef(r). \tag{4.16}$$

For given E, this non-linear integral equation may be solved numerically for $f(r)$; μ is a Lagrange multiplier which plays the role of an eigenvalue. Once $f(r)$ is known, the density of states can be calculated and is found to be

$$\rho(E) = C(\nu) \exp\{-(a^2\kappa^5/16\pi N) b(\nu)\}, \tag{4.17}$$

where $\nu = 2m|E|/\hbar^2\kappa^2$, a is the Bohr radius for the conduction band and $C(\nu)$ is a function of ν and the other material parameters which is slowly varying compared to the exponential. The dimensionless function $b(\nu)$ is shown in fig. 6, and behaves like $\nu^{1/2}$ for $\nu \ll 1$, but like ν^2 for large ν. It is interesting to note that the factor $a^2\kappa^5/16\pi N$ in the exponential is precisely the square of the quantity which should be small for the semi-classical theory to be valid.

Fig. 6

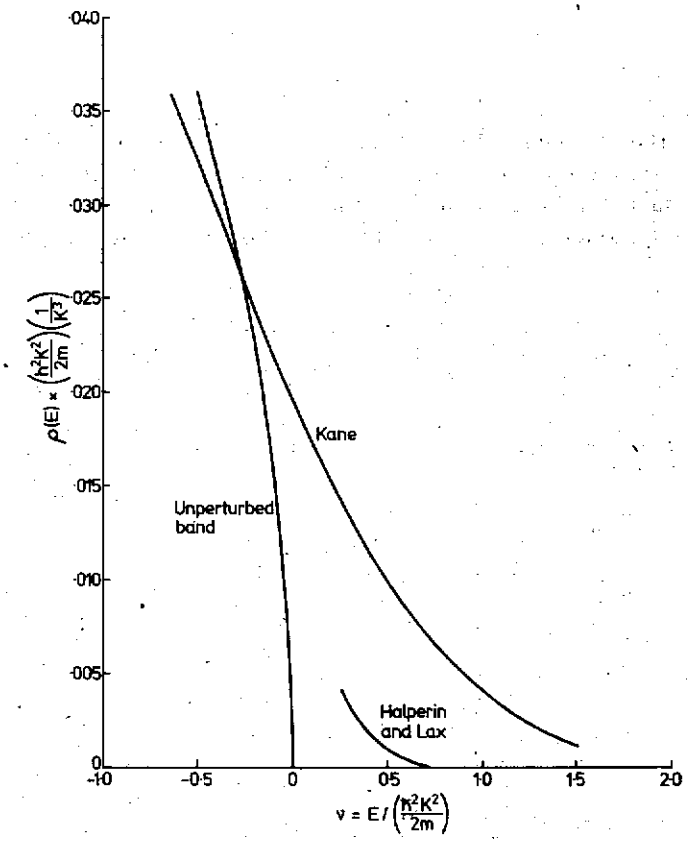


The function $b(\nu)$ obtained from the tabulated values of Halperin and Lax (1966).

† The use of this argument, the neglect of excited states and the choice of a localized trial function restrict the validity of the theory to tail states.

The tail is considered to begin for a value of ν such that the argument of the exponential is roughly 3. The theory is valid for a range of ν in the tail limited by an upper value of ν at which the approximation of Gaussian statistics fails. Halperin and Lax suggest that the range of ν for a useful theory is generally no more than a decade. For a given semiconductor, doping level and temperature, $a^2\kappa^5/N$ is fixed and the relevant range of $b(\nu)$, and hence ν, may be deduced with the help of fig. 6. For $16\pi N/a^2\kappa^5 \gg 1$, the valid tail region corresponds to $\nu \gg 1$ and the energy dependence is Gaussian, just as in the semi-classical theory.

Fig. 7



A comparison of the tail density of states given by a semi-classical calculation (Kane) and a quantum calculation (Halperin and Lax) for $n=N$, $N^{1/3}a=1$. Note that the latter curve does not include a factor of two for spin degeneracy because it is assumed that only single occupancy is possible. Also shown is the unperturbed band.

In a degenerate semiconductor this behaviour is achieved at high densities. It follows from eqns. (4.3) and (4.17) that

$$\rho(E) \propto \exp\left\{-\left(\frac{2}{\pi}\right)\left(\frac{3}{\pi}\right)^{5/6} \frac{b(\nu)}{(n^{1/3}a)^{1/2}}\right\}. \tag{4.18}$$

For $(n^{1/3}a)^{1/2} \gg 1$, $b(\nu) \sim \nu^2$ in the range of interest, but for lower values of n the power of ν decreases towards 0·5. In fig. 7 the semi-classical and quantum results for the tail density of states are compared for $n^{1/3}a = 1$.

Equation (4.17) refers to the conduction band of n-type material, and the expression for the valence band tail follows by replacing the conduction band Bohr radius with that appropriate to the valence band. The formula (4.17) is easily extended to cover p-type and compensated semiconductors. In the compensated case, N is the total impurity concentration, but care must be exercised in using the theory for compensated material since the linear screening approximation is likely to fail. In a given semiconductor with a given doping level and type, the relative lengths of the valence and conduction band tails is determined in eqn. (4.17) by the respective Bohr radii ; a larger Bohr radius gives a shorter tail.

Recently Sa-yakanit (unpublished) has used the path integral formulation of quantum mechanics to carry through an analytical calculation based on the same general principles as the work of Halperin and Lax. This work gives results in good agreement with the numerical calculations.

Although the work on tail states has not been extended rigorously beyond the simple parabolic band model, it is anticipated that the theory can be applied, at least semi-quantitatively, to more general cases such as multiple-valley band structures (see, for example, Stern 1971 a, b).

4.4. *Fermi level and screening*

To obtain an accurate value for the reciprocal screening length κ, a self-consistent solution of eqn. (4.2) is required. For heavily doped, uncompensated material the Halperin–Lax theory shows that the number of states in the tail is small compared to the total carrier concentration, and that eqn. (4.3) provides a good approximation to K. However, in compensated material or at low doping levels a better calculation is necessary. The Fermi level and screening length may be determined self-consistently for either the Halperin–Lax or the semi-classical method. Such calculations have been reported by Hwang (1970 a) and by Hwang and Brews (1971). Similar calculations with injected carriers have been carried out by Stern (1966) and by Hwang (1970 b) in the context of the GaAs laser (see § 7).

Another problem that arises when disorder is large and carrier concentration is small is the failure of linear screening and the approximation of a spatially constant screening length. Herbert *et al.* (1975) have examined this problem in heavily doped, compensated GaAs. They use Kane's theory as a basis, which would appear to be justified because of the large screening lengths expected in compensated material. The carrier density is allowed to respond to the potential fluctuations. At any point a local screening length may be defined in terms of the local electron and hole concentrations. The Fermi level is determined self-consistently from the condition of overall charge neutrality for the whole semiconductor. The numerical calculations show that variable screening considerably reduces the band tailing predicted by average linear screening theory.

4.5. *Nature of the wavefunctions*

Halperin and Lax also give details of the optimum wavefunction $f(r)$ of a tail state. A measure of the extent of a tail state is given by the radius at which $rf(r)$ peaks. This occurs for $\nu=0{\cdot}1$ at $r\simeq 3/\kappa$, for $\nu=10$ at $r\simeq 0{\cdot}8/\kappa$ and for $\nu=1000$ at $r\simeq 0{\cdot}2/\kappa$. Far from the relevant potential fluctuation the asymptotic decay of the wavefunction is expected to be

$$\psi\simeq\exp(-\alpha r),$$

with

$$\alpha\simeq|E-E_c|^{1/2},$$

since the surrounding shallower fluctuations have little influence on a state well below the mobility edge E_c.

In contrast, the states just below E_c are bound to large complexes of potential fluctuations and the wavefunction is strongly affected by the ambient disorder. The early analytical models (Abram and Edward 1972 a, Anderson 1972, Lukes 1972, Freed 1972) suggest that as E_c is approached the extent of the localized states increases to infinity. The wavefunction of a state near E_c has strong spatial fluctuations, but a finite extent defined by some energy-dependent radius of localization $R_L(E)\propto|E-E_c|^{-S}$. For example, if the overall decay is described by the model of an exponential wavefunction $\exp(-\alpha r)$, then $\alpha\sim R_L^{-1}(E)$. An alternative model, due to Last and Thouless (1974), is of a simple power-law decay of the wavefunction. Abram and Edwards, Anderson, and Lukes all obtained $S=\frac{2}{3}$ from their calculation, but Freed obtained $S=\frac{4}{5}$. Both these values contrast with the stronger decay, $S=\frac{1}{2}$, of a wavefunction bound to a single short-range potential in a system without disorder. The extension of the analytical models to two dimensions and a comparison with the numerical studies of Licciardello and Thouless (1976) and Yoshino and Okazaki (1977) indicate that $S=\frac{2}{3}$ is too small. Indeed, Mott (1976 a) has suggested that unless $S\geqslant\frac{4}{5}$ there will be no discontinuity in the zero-temperature mobility at E_c. Finally, a very recent calculation by Edwards (unpublished) gives a value of $S=\frac{4}{5}$. The mobility edge itself has not been fixed with any definiteness, but a plausible position is close to the unperturbed band edge.

The states just above the mobility edge are expected to resemble those originally described by Mott (1970) for Anderson's model of cellular disorder (Anderson 1958). The wavefunction, representing an extended state, has a fluctuating phase and amplitude on the scale of the potential correlation length. The connection of these states to those below the edge has been studied by Abram and Edwards (1972 b). States higher up the band gradually go over to the slightly perturbed intrinsic band states expected at high energies.

We expect the semi-classical approximation to the density of states to be quite good throughout the energy range in question. Indeed, it is one of the few methods available since the perturbation results are unreliable for the tail states. Recently Lloyd and Best (1975) reported an interesting technique for obtaining $\rho(E)$. They show that in a variational model for the density of states the best choice is the one which maximizes the quantity

$$P(E)=\int_{-\infty}^{E}\int_{-\infty}^{E'}\rho(E'')\,\mathrm{d}E''\,\mathrm{d}E'.$$

This is a rigorously derived maximum principle for the density of states which in the limit $E \rightarrow -\infty$ reduces to the variational method used by Halperin and Lax (1966). The important point is that this variational method is valid throughout the band.

4.6. *Some experimental methods*

Evidence for band tails was provided by early experiments on the tunnel diode. The simple model of this device is unable to explain the large current level beyond the peak in the current–voltage characteristic. Kane (1963) suggested that the so-called excess current came from band tailing which prevented a sharp uncrossing of the bands. By using the semi-classical model to describe the tails, Kane produced modified characteristics. This explanation received support from the experiments of Logan and Chynoweth (1963) whose silicon diode characteristics were in close agreement with the theory. Evidence for tail states has also come from photon assisted tunnelling work, particularly in GaAs (see, for example, Pankove 1962, Archer *et al.* 1963).

Attempts have been made to obtain the band density-of-states distribution from tunnelling measurements, but in general the extraction of the relevant information is far from straightforward. However, a number of workers claim to have done this: see, for example, the work of Shepherd *et al.* (1965) on germanium diodes and Fistul and Agaev (1966) on gallium arsenide diodes. The Schottky barrier tunnelling results of Mahan and Conley (1967) were mentioned in § 2.3 and are shown in fig. 2.

There is a vast body of experimental work relating to the optical and transport properties of heavily doped semiconductors; some of this is described in the subsequent sections. Suffice it to say at this stage that these experiments have provided a considerable amount of information on the band structure.

Quantities such as the low-temperature electronic specific heat and Pauli susceptibility are directly proportional to the density of states at the Fermi energy in the degenerate limit and are capable of giving direct information on the band structure. Kobayashi *et al.* (1977) have measured the specific heat of degenerate silicon and find the results to agree with the theoretical value for a parabolic band, with the effective mass of the intrinsic material when the electron concentration approaches 10^{20} cm^{-3}. At lower concentrations, down to 3×10^{18} cm^{-3}, the measured specific heat is rather higher than theory predicts. Enhancement of the specific heat at these doping levels was also noticed much earlier by Keesom and Seidel (1959). It appears that at high concentrations the foregoing theory is vindicated and that the effects of disorder and electron–electron interactions are small. At the lower concentrations these perturbations play a more significant role. Similar conclusions can be drawn from the magnetic susceptibility measurements of Sasaki and Kinoshita (1968). In germanium the specific heat experiments of Keesom and Seidel (1959) and Bryant and Keesom (1962), and the susceptibility measurements of Bowers (1957), show that the parabolic-band–intrinsic-mass model describes these properties well in the heavily doped material.

Magnetic resonance studies have proved valuable for investigating the states of individual impurities and for gaining insight into the nature of the

states at impurity concentrations close to the Mott transition. Reviews of the latter work have been given by Alexander and Holcomb (1968) and Mott (1974). Studies of individual centres can have relevance to heavily doped material. For example, complexes known to occur in heavily doped material, such as the E centre in silicon, can be studied in this way. E.P.R. and ENDOR experiments on this particular centre have been reported by Watkins and Corbett (1964). In degenerate material, E.P.R. measurements may also be used to obtain the Pauli susceptibility (see, for example, Ue and Maekawa 1971).

Devices such as the silicon transistor and gallium arsenide injection laser have given information on their constituent materials, and some of this work is described in the sections on devices.

§ 5. Optical properties

5.1. *Dielectric function*

The transmission of electromagnetic radiation through a medium may be described in terms of the transverse dielectric function. At optical frequencies the zero wavenumber limit $\epsilon(\omega)$ may be used, and we assume this to be a scalar, as indeed is the case for isotropic and cubic materials. The dielectric function† has been derived by many authors (see, for example, Aspnes and Bottka 1974) as

$$\epsilon(\omega)=1-\frac{4\pi e^2}{m_0\,\omega^2\Omega}\sum_j f(\omega_j)+\frac{4\pi e^2\hbar}{m_0^2\,\omega^2\Omega}\sum_i\sum_{j,\,\text{spin}}|\langle\psi_i|\hat{\mathbf{e}}\cdot\nabla|\psi_j\rangle|^2 \times f(\omega_j)(1-f(\omega_i))\left\{\text{princ. pt.}\,\frac{2\omega_{ij}}{\omega_{ij}^2-\omega^2}+i\pi[\delta(\omega_{ij}-\omega)-\delta(\omega_{ij}+\omega)]\right\}, \tag{5.1}$$

where $\hbar\omega_{ij}=E_i-E_j$, $\hat{\mathbf{e}}$ is the polarization vector of the radiation, i and j label the eigenstates of the bands, $f(\omega_j)$ denotes the state occupation probability and m_0 is the free electron mass.

It is convenient to introduce the frequency-dependent complex refractive index, $n_r(\omega)+in_i(\omega)$, which is defined by

$$\epsilon(\omega)=(n_r(\omega)+in_i(\omega))^2. \tag{5.2}$$

Then the electric field $\mathbf{E}$ of a plane wave propagating through the material in the x direction has the form

$$\mathbf{E}=E_0\hat{\mathbf{e}}\exp\left(-\frac{n_i\omega x}{c}\right)\exp\left(\frac{in_r\omega x}{c}-i\omega t\right).$$

The spatial dependence of the phase is determined by $n_r(\omega)$, while the attenuation is described by the absorption coefficient defined by

$$\alpha(\omega)=\frac{2\omega n_i(\omega)}{c}=\frac{\omega\epsilon_i(\omega)}{cn_r(\omega)}, \tag{5.3}$$

† We note in passing that the longitudinal and transverse dielectric functions are identical in the zero-wavenumber limit; local-field effects are ignored (Adler 1962).

where $\epsilon_i(\omega)$ is the imaginary part of the dielectric function. Here $n_r(\omega)$ is a function of the real and imaginary parts of $\epsilon(\omega)$, but if the dispersion is not too strong at the frequency of interest, the frequency dependence of $\alpha(\omega)$ is dominated by $\omega\epsilon_i(\omega)$.

5.2. *Free carrier effects*

Equation (5.1) can be written in an alternative form which explicitly shows the free carrier contribution to the dielectric function. Neglecting disorder, using Bloch functions $\Omega^{-1/2}u_{ik}\exp(i\mathbf{k}\cdot\mathbf{r})$ for the eigenstates, writing

$$\frac{\omega_{ij}}{\omega^2(\omega_{ij}^2-\omega^2)}=\frac{1}{\omega_{ij}(\omega_{ij}^2-\omega^2)}+\frac{1}{\omega_{ij}\omega^2},$$

and using the f sum rule (Callaway 1964) gives

$$\frac{m_0}{\hbar}\frac{\partial^2\omega_j(k)}{\partial k^2}=1-\sum_{i\neq j}\frac{2\hbar\left|\frac{1}{\Omega_{\text{cell}}}\int_{\text{cell}}u_{ik}^*(\mathbf{r})\hat{\mathbf{e}}\cdot\nabla u_{jk}(\mathbf{r})\,d\mathbf{r}\right|^2}{m_0\omega_{ij}(k)}.$$

Equation (5.1) can be rearranged as

$$\begin{aligned}\epsilon(\omega)=1+\frac{4\pi e^2\hbar}{m_0^2}\frac{2}{(2\pi)^3}\sum_{i,\,j\neq i}\int d\mathbf{k}\left|\frac{1}{\Omega_{\text{cell}}}\int_{\text{cell}}u_{ik}^*\,\hat{\mathbf{e}}\cdot\nabla u_{jk}\,d\mathbf{r}\right|^2 f(\omega_j(k))\\ \times(1-f(\omega_i(k)))\left\{\text{princ. pt.}\frac{2}{\omega_{ij}(k)(\omega_{ij}^2(k)-\omega^2)}\right.\\ \left.+\frac{i\pi}{\omega^2}[\delta(\omega_{ij}(k)-\omega)-\delta(\omega_{ij}(k)+\omega)]\right\}\\ -\frac{4\pi e^2}{\hbar\omega^2}\frac{2}{(2\pi)^3}\sum_j\int d\mathbf{k}\,f(\omega_j(k))\frac{\partial^2\omega_j(k)}{\partial k^2}.\end{aligned}\qquad(5.4)$$

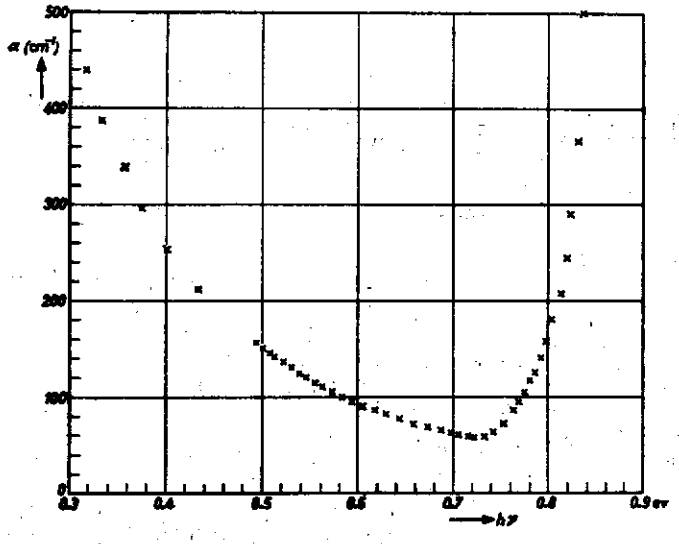
The last term is the free carrier contribution which, written as ω_p^2/ω^2, defines the plasma frequency ω_p. In a heavily doped n-type semiconductor, for example, the conduction band electrons give the only significant contribution to ω_p. The optical properties of the free carriers provide information on the structure of the occupied bands. In particular the effective mass, which appears in ω_p, can be obtained in a number of experiments.

Spitzer and Fan (1957) showed how the low frequency dependence of the real part of the dielectric function, $\epsilon_r(\omega)$, could be used to obtain the conductivity effective mass. Since then various plasma and magnetoplasma reflectivity experiments have been carried out on a large range of semiconductors (see, for example, Fan 1967, Palik and Wright 1967, Black *et al.* 1970). Also, plasma resonance can be used as a tool to measure semiconductor carrier concentrations. Schumann (1970) gives curves for Ge, Si and GaAs relating the wavelength of the reflectivity minimum to the carrier concentration. The free carrier Faraday effect has also proved to be a valuable technique for determining effective masses (Mavroides 1972, Piller 1972). These methods do show how the effective mass changes with energy above the band-edge, but most attention has been paid to comparing the change with the theory of

intrinsic band non-parabolicity (see, for example, Kane 1966). The effects of carrier–carrier and carrier–impurity interactions have been largely ignored but, since agreement between electron intrinsic band theory and experiment is generally good, this in itself is an indication that heavy doping effects are small.

The large numbers of free carriers present in uncompensated, heavily doped semiconductors can lead to strong absorption below and near the fundamental absorption edge. Figure 8 shows the absorption spectrum of heavily doped n-type germanium obtained by Haas (1962); the rise in the curve with decreasing frequency is characteristic of free carrier absorption. Free carrier

Fig. 8



The absorption coefficient at 80 K of germanium doped with $1{\cdot}95 \times 10^{19}$ phosphorus atoms per cm^3 (after Haas 1962).

absorption is due to transitions between or within the constituent bands of the valence or conduction band. The intra-band transitions are only possible when carrier scattering mechanisms are present because of energy and momentum conservation requirements. An alternative to including scattering in the derivation of eqn. (5.4) is to use

$$\epsilon(\omega) = 1 - 4\pi\sigma(\omega)/i\omega, \tag{5.5}$$

where $\sigma(\omega)$ is the conductivity. For example, the Drude $\sigma(\omega)$ leads to an $n\omega^{-2}$ dependence for the spectrum, where n is the carrier concentration. Quantum-mechanical calculations show the frequency dependence to be sensitive to the type of scattering determining the conductivity, and a review of results is given by Fan (1967). Attention has been concentrated on lightly doped non-degenerate semiconductors and there is clearly a need for more work on strong impurity scattering effects.

In experimental investigations of the fundamental edge, the free carrier absorption is usually eliminated from spectra by extrapolating the absorption measured at sub-band-gap frequencies and subtracting this contribution.

5.3. *Fundamental absorption*

The fundamental absorption coefficient is obtained from eqns. (5.1)–(5.3), after a little algebraic manipulation, as

$$\alpha(\omega)=\frac{4\pi^2e^2\hbar^2}{m_0{}^2\,\omega c n_r\Omega}\sum_n\sum_{\substack{n'\\ \text{spin}}}(f(E_n)-f(E_{n'}))\times|\langle\psi_{n'}|\hat{\mathbf{e}}\cdot\nabla|\psi_n\rangle|^2\delta(E_{n'}-E_n-\hbar\omega), \qquad (5.6)$$

where n' and n label the conduction and valence band states respectively. For a semiconductor in thermal equilibrium, $\alpha(\omega)$ is positive because $f(E_n)>f(E_{n'})$ for all n and n'. However, increasing the conduction band occupancy and decreasing the valence band electron occupancy by some external means can create a situation in which $\alpha(\omega)$ is negative and net stimulated emission occurs. Then it is common to quote the photon emission rate, which can be derived from $\alpha(\omega)$ by straightforward energy considerations. The rate of emission of photons, per unit volume, per unit solid angle between ω and $\omega+\mathrm{d}\omega$, for one photon in each of the two polarization states of each plane-wave mode of the stimulating radiation, is

$$r_{\text{stim}}(\omega)\,\mathrm{d}\omega=-(n_r{}^2\,\omega^2/\pi^2c^3)\alpha(\omega)\,\mathrm{d}\omega$$

(see, for example, Stern (1963) for more details).

We may choose any convenient representation to evaluate eqn. (5.6). For example, in perfect crystals it is usual to employ Bloch states labelled by the crystal momentum and then, in the absence of phonon interactions, only vertical transitions are possible in the reduced-zone scheme. The evaluation of $\alpha(\omega)$ for a direct-gap, parabolic band semiconductor is well known (see, for example, Callaway 1964) and

$$\alpha(\omega)\propto(\hbar\omega-E_g)^{1/2},\quad \hbar\omega\geqslant E_g. \qquad (5.7)$$

This result is obtained by neglecting excitonic effects which, if included, produce enhancement of the absorption band and an additional peak below the band-edge (Elliot 1957, Moss *et al.* 1973).

With the introduction of shallow impurities, absorption processes should be observed between the impurity level and the states in the band on the other side of the gap. Eagles (1960) and Dumke (1963) have considered the simple case of non-interacting impurities arising from a single parabolic band, allowing for level degeneracy. The absorption profile is expected to be red-shifted from the fundamental edge with a characteristic shape determined by the Fourier components of the impurity wavefunction and the k-selection rule. If the material is compensated, the random potential arising from the ionized impurities will lead to a spread of impurity levels, as discussed for deep impurities by Morgan (1965), and the absorption peak will broaden accordingly. Increasing the doping concentration of the uncompensated material will also produce a broadened absorption peak which reflects the width of the impurity band

broadened by overlap. However, profiles predicted by Eagles and Dumke for lightly doped material have not been identified unambiguously, largely because in such material the importance of intrinsic absorption, with its complicating excitonic effects, is so much greater.

5.4. *Fundamental absorption at high impurity concentrations*

In heavily doped semiconductors the translational symmetry is broken by the random potential of the impurities and **k** can nb longer be regarded as a good quantum number. Therefore it is expedient to use energy as an eigenfunction label in eqn. (5.6). The absorption coefficient, averaged over all polarization, becomes

$$\alpha(\omega)=\frac{2\pi^2e^2\hbar^2\Omega}{m_0{}^2\,\omega cn_r}\int (f(E)-f(E+\hbar\omega))P(E,\,E+\hbar\omega)\rho_c(E+\hbar\omega)\rho_v(E)\,\mathrm{d}E. \qquad (5.8)$$

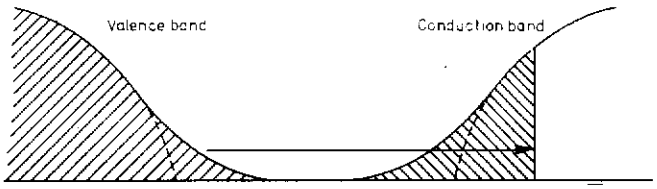
$P(E, E+\hbar\omega)$ is $\Sigma|\langle\psi_E|\nabla_i|\psi_{E+\hbar\omega}\rangle|^2/3$ averaged over states of the same spin in a small energy interval around E and $E+\hbar\omega$, and $\rho_c(E)$ and $\rho_v(E)$ are the density-of-states functions for the conduction and valence bands respectively. If there are several constituent bands in the valence and conduction bands, a sum of expressions (5.8) for $\alpha(\omega)$ should be taken for each pair.

Writing $\alpha(\omega)$ in the form of eqn. (5.8) allows us to make some qualitative predictions about the absorption spectrum, and ultimately to do quantitative calculations. From the consideration of §§ 3 and 4 we have a good idea of the density-of-states function for a heavily doped semiconductor ; the bands have tails and the band-gap is narrowed by many-body effects. The behaviour of $P(E, E')$ is a more difficult problem since it depends on the detailed nature of the eigenfunctions of the system. However, we can presume that when E' is well into the conduction band and E is well into the valence band, $P(E, E')$ will behave similarly to the corresponding quantity for a perfect crystal. In contrast, when E and E' are deep in the band tails the matrix element will resemble that for transitions between two localized impurity states. The third term in the integrand of eqn. (5.8) is the occupancy factor $f(E)-f(E+\hbar\omega)$ which shifts the absorption edge to higher energies in degenerate materials—this is the Moss–Burstein effect (Moss 1954, Burstein 1954).

The individual effects are significant and conflicting making qualitative judgements difficult except in special cases. The Moss–Burstein effect mitigates against band-gap narrowing ; band tailing brings states of the two bands closer together in energy while reducing their spatial overlap. Under the circumstances of degeneracy, certain limited information can be extracted directly from the experimental results. For example, in degenerate n (p)-type material the absorption spectrum is expected to mirror strongly the shapes of the valence (conduction) band tail. As shown in fig. 9 for n-type material, the strongest transitions will be from the valence band tail states to the Fermi level which is assumed to be well into the conduction band, and not significantly affected by the band tailing.

Reliable interpretation of experimental results requires a detailed quantitative model, which some workers have attempted to produce. Lasher and Stern (1964) devised a scheme to model the spontaneous and stimulated emission profiles of heavily doped p-GaAs. The **k**-selection rule for optical

Fig. 9



An illustration of electronic transitions from the valence band tail to the Fermi level in degenerate n-type material (the shading represents occupied states).

processes is weakened by the random potential of the impurities. As a measure of the spread in k space of a state in the valence band, it is reasonable to take the Fourier transform of an acceptor envelope function, and so it was assumed that the matrix element is constant and equal to that for an electron at $k=0$ in the conduction band passing to a shallow acceptor. This procedure clearly overestimates the effect of impurities on transitions between states high in the bands and underestimates the localization of states deep in the tail. It also ignores the effect of free carrier screening on localization.

Casey and Stern (1976) have remedied these defects somewhat. For the band density-of-states functions, they use the expression derived by Kane (see § 4.2), which has the advantage of being able to give a fair representation of the density of states, both well into the band and deep into the tail. To give the Kane expression further accuracy, it is fitted to the Halperin–Lax band tail function (see § 4.3) in the tail region and to the intrinsic density of states well into the band. This procedure is adopted individually for the conduction band and the light and heavy hole valence bands.

The calculation of $P(E, E')$ is the weakest part of the method. In the absence of a theory for the wavefunctions of the states throughout the band, Casey and Stern propose wavefunctions of the form

$$\psi(r)=u(r)(\beta^3/\pi)^{1/2}\exp(i\mathbf{k}\,.\,\mathbf{r})\exp(-\beta|r-r_i|). \tag{5.9}$$

The Bloch periodic function $u(r)$ is modulated by an envelope function involving the energy-dependent parameters k and β. These are determined in a rather arbitrary manner for each band. We take the conduction band as an example, which in the pure material has its edge at E_c and a density of states $\rho_{pc}(E)$. For each state at energy E' in the heavily doped semiconductor, Casey and Stern associate a state in the pure crystal at energy E^* by

$$\int_{E_c}^{E^*}\rho_{pc}(E)\,dE=\int_{-\infty}^{E'}\rho_c(E)\,dE. \tag{5.10}$$

Then k is assigned the value appropriate to the pure crystal by

$$\hbar^2k^2/2m=(E^*-E_c), \tag{5.11}$$

and β is determined by the apparent lowering of the conduction band state due to disorder by

$$\hbar^2\beta^2/2m=\lambda(E^*-E'), \tag{5.12}$$

where $\hbar^2\beta^2/2m$ is the kinetic energy of localization of the deep tail states and λ is a parameter which ensures that this coincides with the kinetic energy obtained by Halperin and Lax at the energy where $\rho_0(E)$ is fitted to their density of states. This formulation provides matrix elements which are correct for $E' \to \infty$ and plausible as $E' \to -\infty$ and, although rather arbitrary, at least provides a working approximation.

Casey and Stern do not include a band-gap shrinkage due to many-body effects in the theory. However, they have fitted their results to absorption data (see fig. 11), and this leads to an empirical shrinkage law. We will discuss these results in § 5.8.

Despite the ingenuity and success of this model, it should be emphasized that certain aspects of the calculation, particularly those concerning the matrix element P, can only be regarded as intuitive. A more rigorous evaluation of eqn. (5.8) requires a direct evaluation of the matrix element without recourse to obtaining individual state wavefunctions. Furthermore, the many-body effects have not been fully incorporated with, for example, band-gap shrinkage only included empirically and excitons neglected altogether. At sufficiently high carrier densities excitons will be precluded by screening, but the effects of the electron–hole interaction should still be considered. In pure indirect-gap semiconductors phonons make possible absorption at the indirect-gap energy, and it is reasonable to assume that they still play a significant role in heavily doped indirect-gap semiconductors.

5.5. *Emission*

In degenerate material the absorption studies are not able to give direct information on the energy separation of states at the bottom of each band because the occupancy of the states at both band-edges precludes edge-to-edge transitions. One way to circumvent the problem is to study absorption in compensated material. However, as we pointed out in § 4.4, the density of states in the band tails is changed by the much reduced screening in compensated material. An alternative approach is to observe the recombination radiation when the equilibrium carrier distribution has been disturbed by some external stimulus. For example, emission may be observed in photo-, cathodo- or electroluminescence experiments.

It is usually assumed that the carriers in each band have come into thermal equilibrium because the ratio of recombination time to thermalization time is large. However, this may not be the case when thermalization is in deep isolated tail states or when recombination proceeds by stimulated emission (see, for example, Nishimura 1974) and ' spectral hole burning ' in the population of the joint density of states at the lasing energy may occur. The rate of photon emission per unit volume per unit solid angle into the frequency range $d\omega$ about ω due to spontaneous recombination is given by Stern (1963) as

$$r_{\text{spon}}(\omega)\,d\omega = \frac{4n_r e^2 \omega \hbar^2}{m_0^2 c^3 \Omega} \sum_{n\,n',\,\text{spin}} \sum f(E_{n'})(1-f(E_n))|\langle\psi_n|\hat{\mathbf{e}}\cdot\nabla|\psi_{n'}\rangle|^2 \times \delta(E_{n'}-E_n-\hbar\omega). \qquad (5.13)$$

It is worth pointing out the evidence from emission studies that any breakdown of the k-selection rule due to heavy doping with shallow impurities

is incomplete. The indirect band-gap material GaP used for green and yellow electroluminescent devices has a low intrinsic efficiency because the indirect phonon-assisted optical transition is improbable compared with the competing non-radiative decay process (Bergh and Dean 1976). Increasing the concentration of shallow substitutional impurities like S_P or Zn_{Ga} does not significantly improve the efficiency and it is necessary to dope with the isoelectronic impurity N_p whose short-range core potential binds an electron in a tight orbit (Faulkner 1968) and allows radiative recombination with holes at the centre of the Brillouin zone. Similarly, solid-state lasers in the visible made from alloys of GaP are confined to direct-gap compounds, or those doped with nitrogen (see, for example, Wolford *et al.* 1976).

5.6. *Radiative lifetime*

Measurement of minority carrier lifetime is not a fine tool for probing the electronic structure of heavily doped semiconductors. However, lifetime is an important parameter in several devices, and we consider here the effects of heavy doping on the radiative lifetime in a direct-gap material where radiative recombination can be fast enough to determine the minority carrier lifetime. GaAs has received the most attention because of its importance as an optical source, and we shall always have this material in mind.

If the free carrier concentration is increased by an amount n, by an external stimulus such as electron- or photon-beam excitation or injection across a p–n junction, and then allowed to return to equilibrium, we can define a carrier lifetime τ by the relation

$$-\frac{\mathrm{d}n}{\mathrm{d}t}=\frac{n}{\tau}. \tag{5.14}$$

In intrinsic GaAs where the equilibrium electron and hole concentrations at room temperature are around 10^7 cm^{-3}, the equal concentrations of generated free electrons and holes may exceed the intrinsic levels by many orders of magnitude. It has been shown (Lasher and Stern 1964) for Boltzmann distributions of carriers in either band that the spontaneous optical recombination rate is given by

$$R_{\mathrm{spon}}=B_{\mathrm{bb}}pn=-\frac{\mathrm{d}n}{\mathrm{d}t}, \tag{5.15}$$

where the inter-band recombination constant B_{bb} is a property of the intrinsic semiconductor and also depends on temperature. For GaAs at 300 K, application of the formula given by Lasher and Stern yields

$$B_{\mathrm{bb}}=2\times10^{-10}\ \mathrm{cm}^{-3}\ \mathrm{s}^{-1}.$$

In this case $p=n$, and eqn. (5.15) becomes

$$-\frac{\mathrm{d}n}{\mathrm{d}t}=B_{\mathrm{bb}}n^2;$$

decay is not exponential and the lifetime as determined from eqn. (5.14) depends on concentration and is not a useful concept.

In doped material, p-type for example, where the number of free holes exceeds the number of generated carriers, only the electron population is significantly perturbed. The recombination rate between free electrons and holes is again given by eqn. (5.15), provided the carriers are Boltzmann distributed. Since the free hole concentration p produced by the doping is independent of excitation, decay is exponential and the value of the useful concept of lifetime is given by

$$\tau_{\mathrm{bb}} = (B_{\mathrm{bb}} p)^{-1}.$$

The calculation of B_{bb} ignores the effect of ionized impurities or scattering states in the valence band.

At higher doping levels the hole population in the valence band becomes degenerate and must be described by Fermi–Dirac statistics. Eventually all states in the valence band directly below those occupied in the conduction band contain holes. No further increase in the hole concentration can provide more final states for **k**-conserving optical processes, so the recombination rate saturates, becoming independent of electron distribution in the conduction band and therefore of temperature. Lasher and Stern (1964, eqn. (5 *d*), later corrected by Stern (1966)) showed that the optical recombination rate takes the form

$$R_{\mathrm{spon}} = An, \tag{5.16}$$

so that the minimum band to band lifetime is given by

$$\tau = A^{-1},$$

which in GaAs is equal to 0·3 ns for electrons.

In lightly doped material holes will be statistically distributed between the valence band and the acceptors, and recombining electrons can find a final state on either. The recombination rate between a density of n electrons and p holes occupying the ground state of shallow hydrogenic acceptors was shown by Dumke (1963) to have the form

$$R_{\mathrm{spon}} = B_{\mathrm{ba}} np.$$

The band–acceptor recombination constant B_{ba} depends on temperature via the Boltzmann distribution of the electrons. The calculation leaves the acceptor ionization energy as a parameter from which the Bohr radius of the acceptor may be determined in a hydrogenic model. At room temperature, for an acceptor 40 meV deep in GaAs, the band–acceptor constant takes the value $B_{\mathrm{ba}} \simeq 3 \times 10^{-10}$ cm^3 s^{-1} and is larger for shallower levels and at lower temperatures. Dumke (1963) estimates a lower limit for band-impurity recombination by pointing out that the wavefunction of an acceptor is provided at the expense of states in the band and the rate cannot exceed that for electrons recombining with holes which occupy all states allowed by momentum conservation. Recombination then proceeds at the rate given by eqn. (5.16).

In the doping region where the impurity band has merged with the valence band edge we must, to calculate matrix elements, resort to such phenomenological wavefunctions as proposed by Lasher and Stern (1964), corresponding to band-edge–shallow-acceptor transitions, or by Casey and Stern (1976) as described in § 5.4. The good agreement with absorption measurements obtained by Casey and Stern lends credibility to their recombination rates

which were calculated by a detailed balance argument from absorption curves. Interpreting their rates in terms of eqn. (5.15), they find values of B ranging from $1{\cdot}9\times10^{-10}$ cm^3 s^{-1} for $p=1{\cdot}2\times10^{18}$ cm^{-3} to $0{\cdot}9\times10^{-10}$ cm^3 s^{-1} for $p=1{\cdot}6\times10^{19}$ cm^{-3}; these values are just over half those calculated by detailed balance from experimentally determined absorption coefficients.

5.7. *Lifetime measurements*

In early work (Schade *et al.* 1971, Abrahams *et al.* 1971, Kawakami and Sugiyama 1973) on GaAs, measurements of diffusion length L were used to obtain the lifetime of minority carriers, invariably electrons in p-type material, from the relation

$$\tau=L^2/D.$$

The results are in broad agreement with those of Zucker *et al.* (1976) who find a diffusion length varying as $p^{-1/2}$ at hole concentrations above 10^{18} cm^{-3}, in accordance with eqn. (5.15). Measurement of laser turn-on delay gives a value of lifetime (Konnerth and Lanza 1964) at a concentration of minority carriers comparable with the doping level. At a temperature of 300 K, Hwang and Dyment (1973) and Namizaki *et al.* (1974) used this method to obtain values of $B=1{\cdot}3\times10^{-10}$ cm^3 s^{-1} and $0{\cdot}9\times10^{-10}$ cm^3 s^{-1}. Casey and Stern (1976) obtain recombination rates from absorption measurements using detailed balance arguments to find values of B ranging from $3{\cdot}2\times10^{-10}$ cm^3 s^{-1} at $p=1{\cdot}2\times10^{18}$ cm^{-3} to $1{\cdot}7\times10^{-10}$ cm^3 s^{-1} at $p=1{\cdot}6\times10^{19}$ cm^{-3}, all at 300 K. The photoluminescence measurements of Vilms and Spicer (1965) show that at moderate doping concentrations of around $p=10^{18}$ cm^{-3} the recombination constant increases from $B=0{\cdot}7\times10^{-10}$ cm^3 s^{-1} at 300 K to an order of magnitude greater at 80 K, a result which is broadly consistent with the prediction that holes frozen out on acceptors provide a faster recombination rate than free holes.

The most direct observations are those of Ackett *et al.* (1974), who measured the phase shift of a high-frequency photoluminescence signal. In their least-compensated samples ($N_D/N_A\simeq0{\cdot}1$), they observed a recombination constant of $B=0{\cdot}6\times10^{-10}$ cm^3 s^{-1} at 300 K over the range $p=2\times10^{18}$ cm^{-3} to 10^{19} cm^{-3}. For a fixed hole concentration, as we proceed to samples with higher levels of compensation we might expect the increased density of states in the acceptor impurity band to enhance the recombination at bound holes, leading to a shorter lifetime. On the contrary, an increase in lifetime is observed for more heavily compensated samples and the effect is enhanced at 77 K where the lifetime is seen to saturate more strongly with increased hole concentration. The following model accounts for the effect by suggesting that the matrix element for the process becomes determined by the inter-impurity separation when this becomes small enough to compare with the extent of the impurity state.

At 77 K in the doping range $p=10^{18}$–10^{19} cm^{-3}, holes are frozen out into the acceptor impurity band. The rate constant B for recombination between the conduction band and the acceptors is proportional to the squared matrix element for the process: $B\sim|M|^2$, where (Dumke 1963)

$$M=\int\exp(i\mathbf{k}\,.\,\mathbf{r})\,\phi_A(\mathbf{r})\,d\mathbf{r}.$$

At low temperatures it is reasonable to set the conduction band wave-vector $\mathbf{k}$ equal to zero, and the matrix element becomes the integral of the acceptor

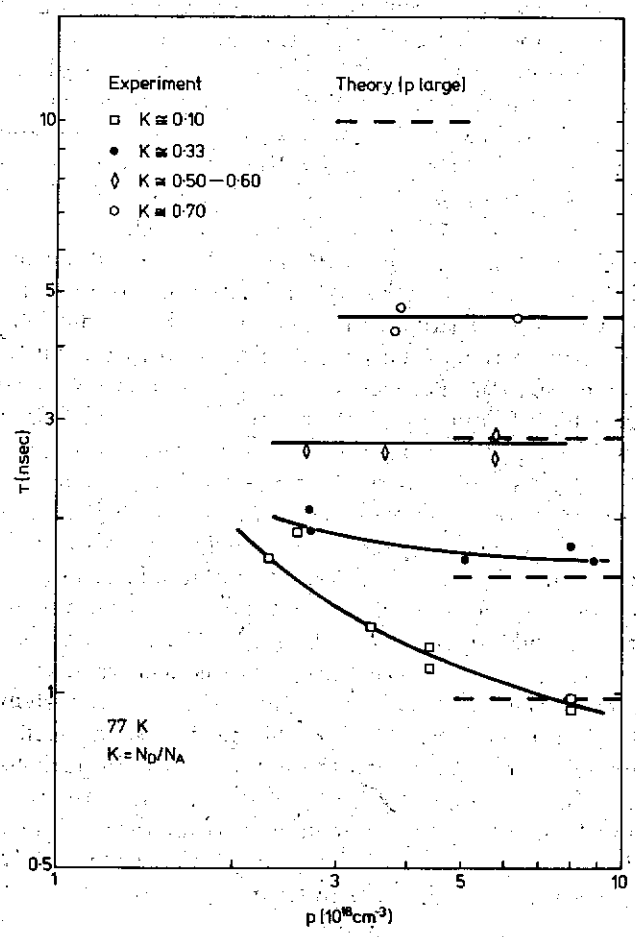
wavefunction $\phi_A = (\pi a^3)^{-1/2} \exp(-r/a)$, where a is the effective Bohr radius. The squared matrix element $|M|^2 = 64\pi a^3$ is independent of doping concentration, and the lifetime varies inversely with hole concentration, in agreement with observations at low doping levels.

At high concentrations of acceptors, N_A, and donors, N_D, following the arguments of Mott (1970 ; see also Mott and Davis 1971), the wavefunction in the acceptor impurity band suffers random changes of phase in a distance d proportional to the mean impurity separation :

$$d = \text{const.}\ (N_A + N_D)^{-1/3},$$

and M must be evaluated by adding contributions from volumes of size d^3 in a random way. In the limit when ϕ_A changes little over a volume d^3, we find

Fig. 10



Electron lifetime against hole concentration in compensated GaAs : Ge, Sn at 77 K (after Ackett *et al.* 1974). Theoretical values for large hole concentrations are plotted on a scale so as to coincide with experiment at $K = 0{\cdot}7$.

that a rough calculation gives the value

$$\bar{M} = (a^3/d^3)^{1/2} \int_{d^3} (\pi a^3)^{-1/2} \exp(-r/a)\, \mathrm{d}\mathbf{r} \sim d^{3/2}$$

(this precise result can be justified on more rigorous grounds). It follows that the squared matrix element is

$$|\bar{M}|^2 = \text{const.}\ (N_A + N_D)^{-1},$$

and since $p = N_A - N_D$, the lifetime

$$\tau = B_p^{-1} \sim \frac{1}{\text{const.}^3} \frac{N_A + N_D}{N_A - N_D},$$

becomes independent of hole concentration and depends only on the compensation, $K = N_D/N_A$ increasing with K as $(1+K)/(1-K)$.

The lifetime data of Ackett *et al.* (1974) are reproduced in fig. 10, together with values of $(1+K)/(1-K)$ on a scale adjusted to make experiment and theory coincide at $K = 0{\cdot}7$. By comparing saturated lifetimes with lifetime in the p^{-1} region, tentatively extrapolated from the $K = 0{\cdot}1$ curve, and using an effective Bohr radius of 13 Å, we can deduce that the phase coherence length d is related to the mean impurity separation $(N_A + N_D)^{-1/3}$ by

$$d = 1{\cdot}3(N_A + N_D)^{-1/3}.$$

A comparison of the curves with the above expression shows that saturation of lifetime is complete when d has reduced to about 50 Å, considerably larger than the extent of the isolated Bohr orbital. This result might appear to cast doubt on the method of evaluating the matrix element $\bar{M}$. In fact, at high doping densities the impurity states couple and the extent of the wavefunction a is more likely to represent the length of localization due to disorder in the impurity band than the Bohr radius. At the onset of localization this length is very large (see § 4.5).

5.8 *Inter-band optical experiments*

Here we describe some experimental work on the inter-band optical properties at high impurity concentrations and pay particular attention to three materials: gallium arsenide, germanium and silicon. GaAs is a direct-gap semiconductor whose properties should be described well by the foregoing theory. Silicon is an important material in the electronics industry, but its optical properties have not been studied extensively. In contrast, germanium, an indirect-gap semiconductor like silicon, has received considerable attention and its properties demonstrate the new features important in indirect-gap materials.

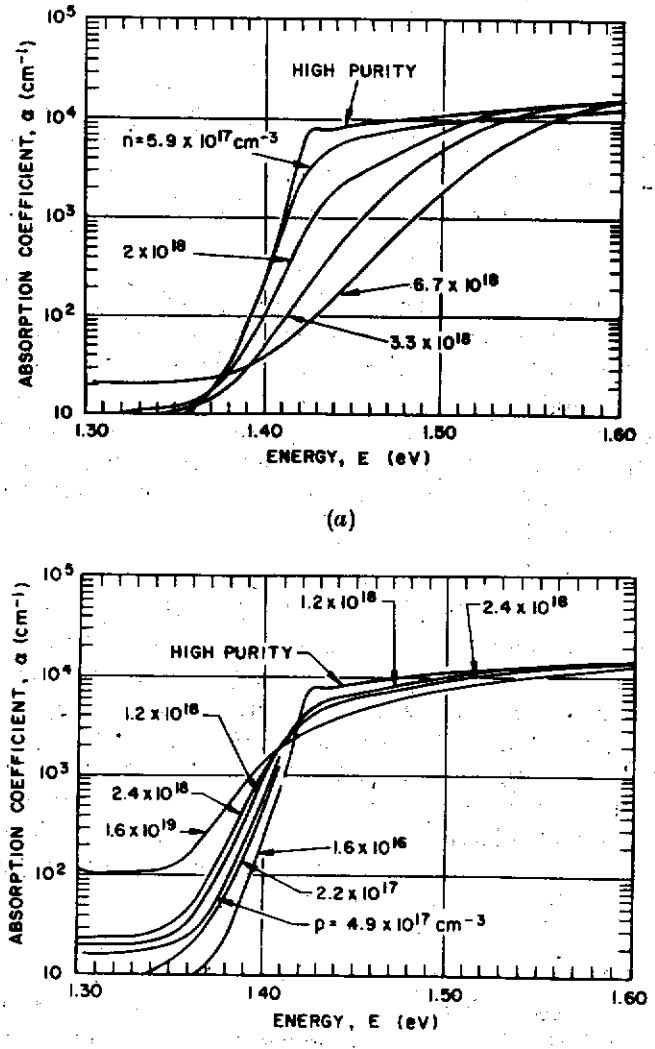
5.8.1. *Gallium arsenide*

Because of its use in infra-red sources (see § 7), GaAs has been the subject of intensive optical studies. Early experiments on optical absorption in heavily doped material at photon energies near the band-gap were carried out by Hill (1964) and Pankove (1965). Neither author was able to measure the absorption coefficient α beyond $\sim 10^3$ cm^{-1} or to obtain the full structure of the

absorption edge. Nevertheless, they were able to show significant changes in the position and shape of the edge with doping concentration at a variety of temperatures. Although the quantitative results on the band structure changes deduced by these workers cannot be regarded as reliable, they found evidence for band tails and departures from the expected Moss–Burstein shifts.

Hill's results for n-type samples ($5{\cdot}3 \times 10^{16}$–9×10^{18} cm^{-3}) showed a shift of the edge to higher energies at 77 K and a less pronounced shift at 300 K. For p-type samples ($6{\cdot}1 \times 10^{16}$–$5{\cdot}7 \times 10^{19}$ cm^{-3}), increased doping moved the edge to lower energies at 300 K but had no significant effect at 77 K. Pankove

Fig. 11

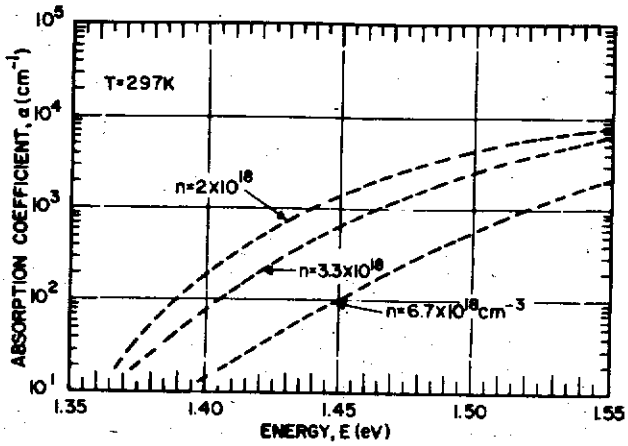


The absorption coefficient for heavily doped GaAs at room temperature; (*a*) n-type. (*b*) p-type (after Casey *et al.* 1975).

obtained similar results in his low-temperature studies. These results may be understood qualitatively in terms of impurity–host band merging at low concentrations, band tailing and gap shrinkage with the competing rise of the Fermi level at higher concentrations. The effect of the latter is not so pronounced in p-type material because of the large hole mass, or at room temperature where degeneracy may be lost.

The results of Hill and Pankove have been corroborated by the comprehensive room-temperature absorption measurements of Casey *et al.* (1975). Using a Kramers–Kronig analysis of reflectance measurements, they obtained values for $\alpha(\omega)$ right into the absorption band. Some of their results for n- and p-type material are shown in figs. 11 (*a*) and 11 (*b*) respectively. The shifts and broadening of the spectra with increased doping, intimated by earlier experiments, are now clear in these results. It is interesting to observe the exciton enhancement peak near 1·42 eV in the pure material curve is not present at high impurity concentrations, presumably because of free carrier screening.

Fig. 12



Calculated values of the absorption coefficient for n-type GaAs (after Casey and Stern 1976).

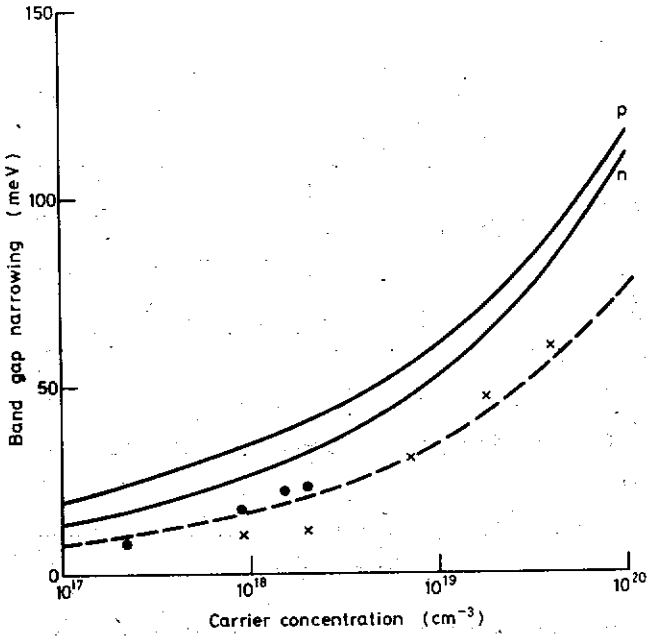
Casey and Stern (1976) have attempted a theoretical explanation of these results by using the model discussed in § 5.4. Figure 12 shows for comparison some of the n-type theory results. The many-body change in the band-gap is not included in the model, but is determined empirically by rigidly shifting the theoretical curves to give the best fit to the experimental data. The agreement is seen to be quite good ; for their p-type results Casey and Stern deduce an empirical gap shrinkage

$$|\Delta E_g| = 1{\cdot}6 \times 10^{-8} p^{1/3}\,\text{eV}. \tag{5.17}$$

This relation is plotted in fig. 13, where it is compared with the theoretical results of § 3.5. At the lower hole concentrations the assumption of degeneracy implicit in fig. 3 (*b*) is not satisfactory and a fair comparison is not possible.

Zverev *et al.* (1977) have defined an effective band-gap as the zero magnetic field limit of the separation of the Landau levels of the conduction and valence bands. They deduce this quantity from oscillatory magneto-absorption experiments. The gap narrowing has a similar dependence on carrier concentration for both n- and p-type GaAs, reaching ~ 40 meV at 10^{19} cm^{-3}. In closely compensated material the gap is only weakly dependent on the total impurity concentration, suggesting that this measurement technique is largely unaffected by the band tails. The results, shown in fig. 13, are seen to agree well with eqn. (5.17), but lie somewhat below the theoretical curves. Reasons for this have been discussed in § 3.5.

Fig. 13



Band gap narrowing versus carrier concentration for GaAs ; p, theory ($-\Delta E_g^{\Gamma}$ of § 3) for p-type GaAs at 0 K ; n, theory ($-\Delta E_g^{\Gamma}$ of § 3) for n-type GaAs at 0 K. The dashed curve is eqn. (5.17) for gap narrowing in p-type GaAs at 300 K deduced empirically by Casey and Stern (1976). The experimental results of Zverev *et al.* (1977) at 77 K are also shown : ●, n-type ; ×, p-type.

The emission spectrum of heavily doped GaAs has also received considerable attention. The band-edge photoluminescence of p- and n-type GaAs and the electroluminescence of junctions were measured by Nathan *et al.* (1963 a). The luminescence peak in p-type material decreases in energy with increased doping, implying a shrinkage of the effective gap. The n-type peak rises, indicating band filling. Essentially similar results have been obtained by Hill (1964). Leite *et al.* (1965) interpreted their electroluminescence results in terms of a band tailing model. Cathodoluminescence experiments have been

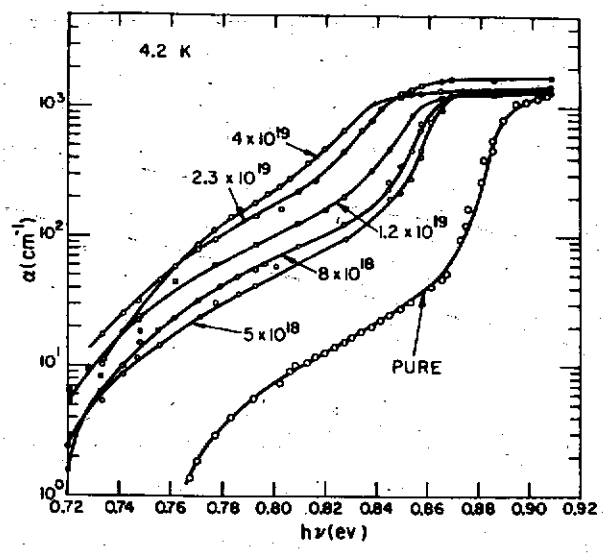
carried out by many workers, including Cusano (1964), Casey and Kaiser (1967), Tuck (1968) and Pankove (1968 a). Results are in general agreement with other techniques.

In compensated material it is possible for the important states in radiative recombination to be deep in the band tails and the luminescence peak can be shifted to energies well below the band-gap. Also, recombination times are increased considerably because the participating states are strongly localized and spatially separated, as fluctuations attracting electrons repel holes and vice versa. Furthermore, it is possible for the thermalization times of excited carriers to exceed the radiative lifetimes because, in the tails, states close in energy are likely to be well separated in space. These effects have been observed in the transient photoluminescence experiments of Redfield *et al.* (1970) on partially compensated Si doped GaAs (see also Alferov *et al.* 1973). These experiments are a potential source of information on the spatial distribution of tail states.

5.8.2. *Germanium*

The inter-band optical properties of germanium are interesting because this material provides a good example of the behaviour of an indirect-gap semiconductor. Haas (1962) and Pankove and Aigrain (1962) measured the inter-band absorption spectrum of *n*-type Ge. Figure 14 shows the results of the latter for As-doped Ge at 4·2 K with free carrier absorption removed. The absorption due to indirect transitions (low energies) is enhanced by heavy doping and can be attributed to the effect of impurity scattering. This explanation receives support from Keldysh and Proshko (1964), who showed the strength of the indirect absorption in Ge to be sensitive to the impurity type.

Fig. 14



The inter-band absorption coefficient of arsenic-doped Ge at 4·2 K (after Pankove and Aigrain 1962).

Some workers, such as Haas (1962) and Rogachev and Ryvkin (1966), have suggested electron–electron scattering to be important and this should be investigated further. At higher energies the direct absorption shoulders in the curves can be seen.

With increasing impurity concentration the curves move to lower energies. Both Haas (1962) and Pankove and Aigrain (1962) interpreted this in terms of an effective band-gap narrowing without discriminating between band tailing and quasi-rigid band movements. In reality, the existence of tails means that the band-gap cannot be defined unambiguously. Nevertheless, Pankove and Aigrain found an effective indirect-gap narrowing of ~ 40 meV at 4×10^{18} cm^{-3} and ~ 100 meV at 5×10^{19} cm^{-3}, and contributions can be expected to come from tailing and many-body effects. In experiments on compensated material Haas found the gap narrowing to be determined by the total impurity concentration, which suggests significant band tailing effects.

Rogachev and Sablina (1966) have also determined an effective band-gap shrinkage from electroluminescence experiments at 80 K to be

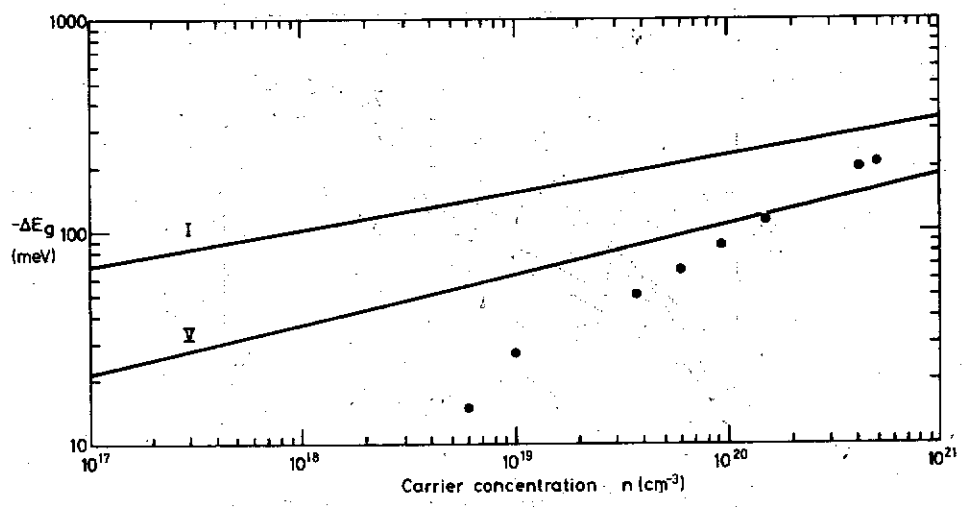
$$|\Delta E_g| = 1{\cdot}5 \times 10^{-8} n^{1/3} \text{ eV},$$

for n-type material. Just as with the absorption experiments, a more careful analysis in terms of disorder and many-body effects is required to extract accurate information. Photoluminescence work similar to that of Redfield *et al.* (1970) on GaAs has been carried out on heavily doped, compensated Ge by Rentzsch and Shlimak (1977).

5.8.3. *Silicon*

Despite the technological importance of silicon, there has been relatively little work on the inter-band optical properties of the heavily doped material.

Fig. 15



Band-gap narrowing versus electron concentration for n-type Si. I, theory ($-\Delta E_g^{TF}$ of § 3) for 0 K; V, theory ($-\Delta E_g^{L}$ of § 3) for 0 K; ●, experiment (Balkanski *et al.* 1969) at 35 K.

Balkanski *et al.* (1969) carried out transmission experiments on n-type (phosphorus doped) silicon (6×10^{18}–$4{\cdot}9\times10^{20}$ cm^{-3}) at temperatures of 35, 85 and 300 K. The results are interpreted in terms of free carrier absorption, indirect inter-band transitions and intra-conduction band transitions, neglecting band tailing effects. Within its framework a gap shrinkage is deduced amounting to approximately 100 meV at a donor density of $1{\cdot}5\times10^{20}$ cm^{-3}, irrespective of temperature. For 35 K the gap shrinkage varies with the square root of the donor density ; details are given in fig. 15 and compared with the predictions of the many-body theory. Using fig. 7 as a guide, band tails greater than 100 meV in length are expected at $1{\cdot}5\times10^{20}$ cm^{-3}. Despite the tails, there is fair agreement between the experimental gap shrinkage and the many-body theory predictions at the higher carrier concentrations.

Vol'fson and Subashiev (1967) have also carried out absorption experiments on heavily doped silicon at room temperature. They find a fundamental absorption edge which shifts with increasing impurity concentration in both n- and p-type materials. On the basis of a rather small number of data points, they explain their results in terms of a rigid band model with a gap shrinkage given by

$$|\Delta E_g| = \begin{cases} A(N^{1/3}-N_0^{1/3}), & N>N_0 \\ 0, & \text{otherwise,} \end{cases}$$

where for p-type material $A\simeq2{\cdot}5\times10^{-6}$ meV cm and $N_0\simeq1{\cdot}4\times10^{19}$ cm^{-3}, and for n-type material $A\simeq3{\cdot}4\times10^{-6}$ meV cm and $N_0\simeq1{\cdot}8\times10^{19}$ cm^{-3}.

5.9. *Auger recombination*

Despite the fact that the rate of inter-band Auger recombination increases as the square of the majority carrier concentration (at least in non-degenerate material), no attention has been paid to the effects of disorder on the wavefunctions, and hence the matrix element for the process, although these effects are expected to be significant in the region of intermediate doping. For example, the calculations on GaP (Hill and Landsberg 1976, Hill 1976) and GaAs (Zschauer 1969) all use unperturbed band wavefunctions.

In the calculations by Antončik (1967) and by Kane (1967) on the converse process of impact ionization, the **k**-selection rule is relaxed only as a simplifying artefact, while Landsberg and Robbins (1977) make reference to the breakdown of **k** selection only in connection with phonon-assisted transitions.

Clearly there is scope and need for further calculations to account for the effects of degeneracy and modifications to the wavefunctions and densities of states from high doping levels.

§ 6. Electrical conductivity

6.1. *Introduction*

In heavily doped semiconductors the scattering processes which determine the electrical conductivity are just those present in lightly doped material, although their relative importance may differ. Impurity scattering becomes

more important as the doping is increased, and we concentrate on this because some well-defined problems remain which are closely connected with the theory of band structure we have discussed already.

At zero temperature, phonon scattering is absent and the real part of the conductivity at frequency ω follows from eqns. (5.1) and (5.5) as

$$\sigma_R(\omega)=\frac{e^2\pi\hbar}{m_0^2\,\omega\Omega}\sum_i\sum_{j,\,\text{spin}} f(\omega_j)(1-f(\omega_i))|\langle\psi_i|\nabla|\psi_j\rangle|^2 \times[\delta(\omega_{ij}-\omega)-\delta(\omega_{ij}+\omega)]. \qquad (6.1)$$

In principle this formula provides a method of calculating the conductivity as a function of the position of the Fermi energy in the band, although quantitative evaluation is generally difficult. A deduction which can be made is that at zero temperature the d.c. conductivity $\sigma_R(0)$ vanishes when the Fermi energy lies among localized states. However, our main interest is in calculating the conductivity of heavily doped material at finite temperatures (e.g. room temperature) when phonon scattering must be included. As we have seen in § 4, the Fermi level of heavily doped materials can lie well above the tail states of the band. As a result the significant contributions to the conductivity can come from carriers in states only weakly pe rbed by the impurity potentials, and it seems likely that the Boltzmann eq tion may be used as the basis of a transport theory. Indeed, as was first shown by Edwards (1958), eqn. (6.1) reduces to that obtained from the Boltzmann equation in the weak scattering limit. The use of the Boltzmann equation in semiconductor transport calculations is well established, and has been remarkably successful in describing the transport properties of lightly doped semiconductors (Ziman 1960, Conwell 1967, Rode 1975). The method can be readily employed in degenerate situations and it is therefore worthwhile examining its efficacy in heavily doped semiconductors.

6.2. *Boltzmann equation*

We consider electrons in an isotropic parabolic conduction band. In doing so we have a reasonable representation of, for example, electrons in GaAs. A description of electron transport in multiple-valley conduction bands or hole transport in a degenerate valence band requires further elaborations which are pointed out where relevant.

For low fields we may expand the electron k-space distribution function about the equilibrium Fermi–Dirac form $f_0(k)$. To first order,

$$f(\mathbf{k})=f_0(k)+g(k)\cos\phi, \qquad (6.2)$$

where ϕ is the angle between $\mathbf{k}$ and the applied field $\mathbf{F}$. It follows from the definition of the conductivity that

$$\sigma=\frac{1}{3\pi^2}\frac{e\hbar}{mF}\int k^3 g(k)\,\mathrm{d}k\,; \qquad (6.3)$$

$g(k)$ is obtained by substituting eqn. (6.2) into the Boltzmann equation (see Rode (1975) for details). The resulting equation for $g(k)$ is

$$\frac{eF}{\hbar}\frac{\partial f_0(k)}{\partial k} = -g(k)\int (1-\cos\theta)S_{\mathrm{el}}(\mathbf{k},\mathbf{k}')\,\mathrm{d}\mathbf{k}'$$
$$-g(k)\int [S_{\mathrm{inel}}(\mathbf{k},\mathbf{k}')(1-f_0(k'))+S_{\mathrm{inel}}(\mathbf{k}',\mathbf{k})f_0(k')]\,\mathrm{d}\mathbf{k}'$$
$$+\int g(k')\cos\theta[S_{\mathrm{inel}}(\mathbf{k}',\mathbf{k})(1-f_0(k))+S_{\mathrm{inel}}(\mathbf{k},\mathbf{k}')f_0(k)]\,\mathrm{d}\mathbf{k}', \quad (6.4)$$

where θ is the angle between $\mathbf{k}$ and $\mathbf{k}'$, and $S_{\mathrm{el}}(\mathbf{k},\mathbf{k}')$ and $S_{\mathrm{inel}}(\mathbf{k},\mathbf{k}')$ are respectively the elastic and inelastic differential scattering rates from state $\mathbf{k}$ to $\mathbf{k}'$. The integral in the first term on the right-hand side of eqn. (6.4) is simply the reciprocal of the relaxation time for elastic scattering, $\tau_{\mathrm{el}}(k)$.

Just as in lightly doped material, scattering due to acoustic phonons and impurities may be considered to be elastic at room temperature. It is significant that $f_0(k)$ does not appear in the integral for the elastic relaxation time :

$$\tau_{\mathrm{el}}^{-1}(k) = \int (1-\cos\theta)S_{\mathrm{el}}(\mathbf{k},\mathbf{k}')\,\mathrm{d}\mathbf{k}'. \quad (6.5)$$

Hence $\tau_{\mathrm{el}}(k)$ is unaffected by arbitrary degeneracy, although the distribution function in equilibrium clearly influences the conductivity : from eqns. (6.3)–(6.5), and assuming elastic scattering only, we obtain the familiar expression

$$\sigma = \frac{1}{3\pi^2}\frac{e^2}{m}\int \tau_{\mathrm{el}}(k)k^3\left(-\frac{\partial f_0(k)}{\partial k}\right)\mathrm{d}k. \quad (6.6)$$

When inelastic scattering occurs (e.g. by optical-phonon or inter-valley scattering), all the terms in eqn. (6.4) must be retained and the equilibrium distribution $f_0(k)$ appears on the right-hand side. The solution of this integral equation has been carried out by Rode (1971) and by Rode and Knight (1971) to determine the transport properties of a number of III–V semiconductors. For other materials the reader is referred to the review by Rode (1975).

Equation (6.4) includes no term for electron–electron scattering. This form of scattering does not contribute to the resistivity in isotropic bands in the degenerate limit because conservation of electron momentum and energy near the Fermi surface implies conservation of total electron velocity. However, the argument no longer holds in the many-anisotropic-valley case where electrons in different valleys can scatter, remaining in their respective valleys, and contributing to the resistivity because of the anisotropy. This problem has been discussed by Morgan (1963), Price and Krieger (1966) and Meeks and Krieger (1969). The relevant term in the resistivity is proportional to T^2, and in § 6.3 we draw attention to its observation in degenerate silicon and germanium. In non-degenerate material the electron–electron scattering tends to equalize the energy among the electrons, and therefore affects even the resistivity of isotropic band materials through the distribution functions.

For all the scattering processes the differential scattering rates are given in the standard works on transport theory we have quoted. We now turn to impurity scattering which we discuss in more detail.

6.3. *Impurity scattering*

The theory of impurity scattering developed by Brooks (1955), Herring (unpublished) and Dingle (1955) is commonly employed in semiconductor transport calculations. Despite improving on the earlier work of Conwell and Weisskopf (1950), this approach has certain shortcomings and it is convenient to introduce these within the framework of a brief description of the method.

The electrons are assumed to be non-interacting, but are capable of screening and obey Fermi statistics. The screened potential of each centre is taken to be of the normal form with Thomas–Fermi screening,

$$V(r) = -\frac{e^2}{\epsilon r}\exp(-\kappa r), \tag{6.7}$$

where κ is given by eqn. (4.2). Brooks also considers screening due to carriers on impurities, but we will ignore that effect here (Falicov and Cuevas 1967, Rode 1975). The impurities are assumed to scatter independently and the cross-section is calculated by the Born approximation. It then follows that, for a density of N ionized centres, the scattering relaxation time for an electron with wavenumber k is given by

$$\frac{1}{\tau(k)} = \frac{2\pi N e^4 m}{\epsilon^2 \hbar^3 k^3}\left\{\ln\left(1+\frac{4k^2}{\kappa^2}\right) - \frac{4k^2/\kappa^2}{(1+4k^2/\kappa^2)}\right\}. \tag{6.8}$$

As we would expect, the scattering rate is directly proportional to the number of scattering centres. However, as N increases, the carrier density increases and κ changes. Often the dominant energy dependence is contained in the k^{-3} term of the bracket prefactor which in the Boltzmann limit gives the mobility a $T^{3/2}N^{-1}$ dependence and in the degenerate limit an $E_F{}^{3/2}N^{-1}$ dependence.

The scattering is taken to be elastic because of the comparatively large mass of an impurity centre. For the Born and Thomas–Fermi approximations to be valid, we must have $4k^2/\kappa^2 \gg 1$ (Schiff 1968). For non-degenerate material, this can be shown to imply that

$$(k_B T/E_d)^{1/2}(N_c/n) \gg 1, \tag{6.9}$$

where E_d is the donor binding energy, n the electron concentration and N_c the number of states within $3k_BT/2$ of the band edge. In degenerate semiconductors the equivalent of eqn. (6.9) is the ubiquitous

$$n^{1/3}a \gg 1. \tag{6.10}$$

Krieger and Meeks (1973) have suggested that, in degenerate semiconductors with many-valley, highly anisotropic conduction bands, the Thomas–Fermi screening approximation is inadequate. For these semiconductors the wavenumber dependence of κ, neglected in Thomas–Fermi screening, is important. Krieger and Meeks (1973) and Krieger *et al.* (1974) have shown how agreement between theory and experiment for degenerate n-type germanium and silicon can be improved by allowing for this effect. At impurity densities of practical interest the Born approximation cannot always be justified, and attempts to improve it have been made. Moore (1967 a, b) has calculated corrections to the Born approximation due to higher-order terms in the scattering series.

The corrections increase the resistivity and are significant ($\gtrsim 10\%$) in n-type GaAs (2×10^{17}–10^{19} cm^{-3}) at room temperature and below.

An alternative approach has been purused by Krieger and Strauss (1968). They consider a degenerate electron gas and assume the impurity potential to be of the Thomas–Fermi form, eqn. (6.7), but fix κ by requiring the partial wave shifts δ_l to satisfy the Friedel sum rule (Ziman 1972). For singly ionized impurities this is

$$\sum_{l=0}^{\infty}(2l+1)\delta_l = 2/\pi. \tag{6.11}$$

The relaxation time is then given by

$$\tau^{-1} = (4\pi N\hbar/m\kappa)\sum_{l=0}^{\infty}(l+1)\sin^2(\delta_l-\delta_{l+1}). \tag{6.12}$$

Krieger and Strauss find a higher resistivity than predicted by the Born approximation for $n^{1/3}a > 0{\cdot}1$; the two methods disagree by about 50% at $n^{1/3}a = 0{\cdot}2$ but give roughly the same result for $n^{1/3}a \simeq 1$ and higher values.

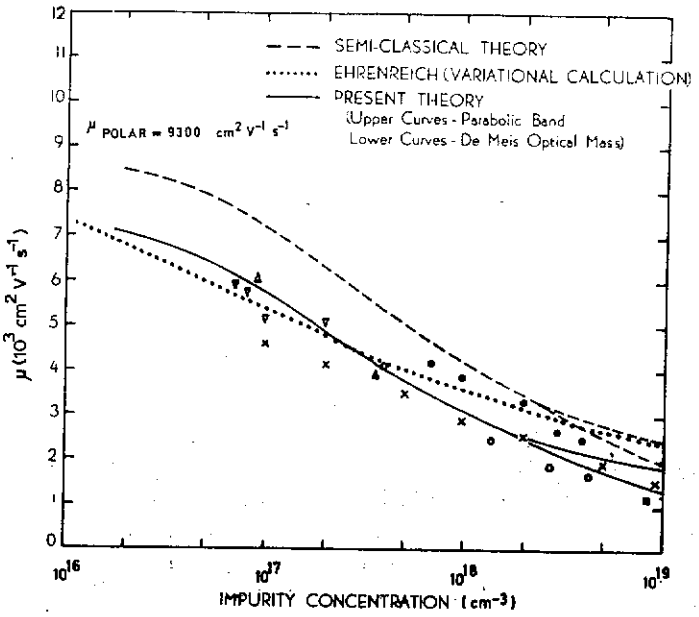
Although the inaccuracies of the Born approximation can be overcome by an improved calculation, such as one of those mentioned above, problems still remain. Equation (6.8) is calculated on the assumption that the electron is scattered separately and incoherently by the impurities. When the electron is influenced strongly by several centres simultaneously, multiple scattering occurs. A simple condition for avoiding multiple scattering is $\kappa N^{-1/3} > 1$. However, this criterion is too restrictive since we are not simply concerned with how many centres the electron interacts with at any one time, but also with the strength of the effect of the centres on the electron motion. Rode (1975) has suggested that multiple scattering is unimportant if the time required for the deflection of the electron by a centre is smaller than the mean time between collisions $\tau(k)$. This implies approximately

$$\kappa^{-1}/(\hbar k/m) < \tau(k). \tag{6.13}$$

From eqns. (6.8) and (6.13) it is clear that multiple scattering will always occur when k is sufficiently small. Another complication is the quantum-mechanical nature of the electron motion. At low energies the electron wavelength will exceed the inter-impurity distance and the quantum mechanics of the problem must be considered. Even at higher energies, coherent scattering effects can occur, and any structure in the distribution of impurities may be important. Such structural effects are well known in the theory of liquid metal (see, for example, Ziman 1961). Moore (1967 a, b), in the calculation mentioned previously, considered corrections to the Brooks–Herring theory due to multiple scattering and has obtained, in particular, numerical results for GaAs. In GaAs at room temperature optical phonon scattering is also important, and this has been included to give the results shown in fig. 16. Also shown are the results of simple Brooks–Herring theory, a variational calculation due to Ehrenreich (1963) and the data points of various other experimental investigations. Moore's results are seen to agree well with experiment.

The problems encountered with multiple scattering at low electron energies are signs of the failure of what is essentially a perturbation theory. As we have seen, the states at the bottom of the band of a heavily doped semiconductor

Fig. 16



Electron mobility of n-type GaAs versus impurity concentration at 300 K. The curves are theoretical results ; the data points are experimental measurements (after Moore 1967 b).

cannot be described as perturbed Bloch waves. Localized states occur and cannot be described by perturbation calculations ; even the extended states near E_c are strongly scattered. To describe the conductivity in these extended states it is necessary to return to eqn. (6.1). In the localized states transport is only possible by phonon excitation to the mobility edge where the states are extended, or by phonon excitation to another localized state, i.e. by hopping. For a discussion of these phenomena the reader is referred to Mott and Davis (1971).

6.4. *Some transport experiments*

Low-temperature conductivity, Hall effect and magnetoresistance experiments have played a major role in impurity band studies. We shall not discuss that work here ; the interested reader is referred to Mott and Davis (1971) and Mott (1974). It is customary to use the plot of Hall coefficient R_H against inverse temperature to determine whether the impurity band is separate from the host band at a particular doping level. If the bands are separate, a simple two-band model of the Hall effect (Mott and Twose 1961) predicts a characteristic hump in the plot. This is indeed seen experimentally and its disappearance at higher doping concentrations is associated with a merging of the bands. In degenerately doped material the Hall effect gives $R_H = 1/nec$, which is independent of temperature. This classical behaviour, characteristic of weak scattering and reminiscent of many liquid metals, is shown clearly in the result of Yamanouchi *et al.* (1967) for heavily doped n-type silicon.

When the significant carriers are in a single valley and out of the band tail, the conductivity theory described in this section gives a fair description of the experimental results. Figure 16 bears this out for GaAs ; the reader is also referred to the collection of results given by Rode (1975) and references therein. However, in many-valley semiconductors below the degeneracy temperature T_D the resistivity is found to vary quadratically with temperature in a different manner than the theory would suggest, unless inter-valley electron–electron scattering is included (Katz *et al.* 1965, Meeks and Krieger 1969). For As-doped Ge (Katz *et al.* 1965) and P-doped Si (Yamanonchi *et al.* 1967),

$$(\rho(T)-\rho(0))/\rho(0) \simeq A(T/T_D)^2, \quad T \ll T_D,$$

where $A \simeq 2$ for Ge and 10 for Si.

Sb-doped Ge shows a 3/2 power-law behaviour which cannot be explained in the same way. Nevertheless, the electron–electron scattering hypothesis received support from piezoresistance experiments performed by Katz (1965). In germanium the [111] conduction band valley may be lowered relative to the other three by uniaxial compression along [111]. When all the electrons are transferred to one valley in this way, the temperature dependence is given by the theory without electron–electron scattering. At lower stress levels the partial transfer of electrons depends on the density-of-states distribution of the valleys, and may be monitored by the change of resistivity. Fritzsche and Cuevas (1962), Cuevas and Fritzsche (1965) and Katz (1965) have found evidence for band tailing in heavily doped germanium using this method. However, the extraction of quantitative information on the tail shape is difficult.

The oscillatory behaviour in the magnetic field dependence of the resistivity (the Shubnikov–de Haas effect) has been recognized as a possible way of determining the effective mass of light mass carriers since the early work of Sladek (1958) on InAs and of Frederikse and Hosler (1957) on InSb. More recently, Vul *et al.* (1976) have studied compensated *n*-type GaAs, and found the effective mass to be unchanged from the lightly doped value. This is both surprising and interesting since the total impurity concentration is 5×10^{17} cm^{-3}, giving 10^{17} cm^{-3} electrons, and the Fermi level cannot be far into the band. Raymond *et al.* (1977) have also examined compensated GaAs but tried to analyse their results in a way which eliminated the effects of disorder. The low-field magnetoresistance in many semiconductors is negative and this behaviour, which is still not completely understood, is discussed by Alexander and Holcomb (1968).

To obtain more direct information on the tail states it is advantageous to move the Fermi level into the tail by compensation. There has not been much experimental interest in this area, but an exception is the work of Redfield (1975) who has looked at a.c. and d.c. electron conductivity in heavily doped, closely compensated GaAs. Redfield's results are qualitatively interpreted in terms of a conduction band tail containing localized states, but without a sharp mobility edge. At low fields and low temperatures the conductivity behaves as

$$\sigma \propto \exp\{-(T_0/T)^{1/2}\},$$

which is not the anticipated variable-range hopping dependence on temperature, possibly because of correlation effects (Mott 1976 b).

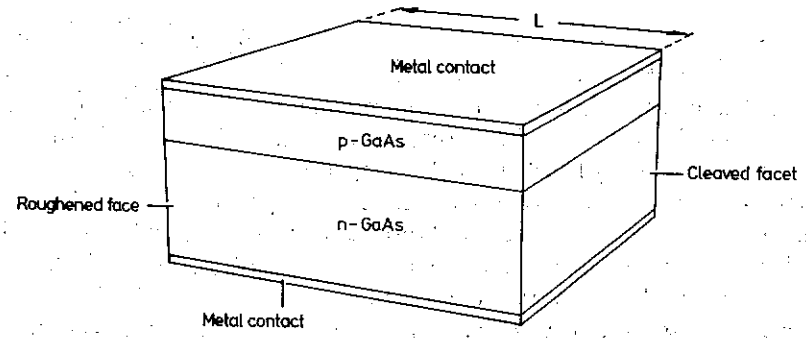
Finally, we mention the recent work on low-temperature conductivity in field-effect transistors. In the experiments on MOS and MNOS transistors, the carriers are confined to a two-dimensional inversion layer at the silicon–silicon-dioxide interface. In this layer the carriers feel the random potential of fixed charges at the interface and within the oxide (and nitride). The great advantage of this system for studying the effects of disorder is the facility with which the position of the Fermi level may be varied by the transistor gate bias. Mott (1973 b) suggested that the random potential produces mobility edges in the semiconductor bands, and the subsequent experiments of Pepper and co-workers confirmed this (see Mott *et al.* (1975) for a description). Since then, extensive experimental and theoretical work has improved our understanding of disorder effects. More recently, Pepper (1977) has shown how two-dimensional impurity band experiments can be carried out using the GaAs Schottky barrier FET.

§ 7. Injection lasers

7.1. *Introduction*

In this section we shall consider how the consequences of heavy doping affect the operation of solid-state injection lasers. Lasing has been observed in a large number of crystalline semiconductors, and was first achieved in GaAs. Because of its advanced material technology and the suitability of its

Fig. 17



Schematic diagram of a diffused homojunction laser.

emission wavelength for optical-fibre communications, GaAs has been the subject of most investigations of lasing. Several reviews exist on the subject : the most recent, by Panish (1976) deals with heterostructures. The principles of operation, discussed for example by Sze (1969), Gooch (1973) and Carrol (1974), are now briefly described.

A schematic diagram of a diffused homojunction laser is shown in fig. 17. Minority carriers are injected into the active region across a p–n junction. If the injection rate is sufficiently high to cause population inversion, then, in the presence of radiation, the absorption coefficient $\alpha(\omega)$ at frequency ω becomes negative, corresponding to stimulated emission of radiation when an electron

and hole recombine across the energy gap (see eqn. (5.6)). There occurs a net amplification of radiation traversing this active region, where the absorption coefficient is negative, to an extent described by the gain coefficient, defined as

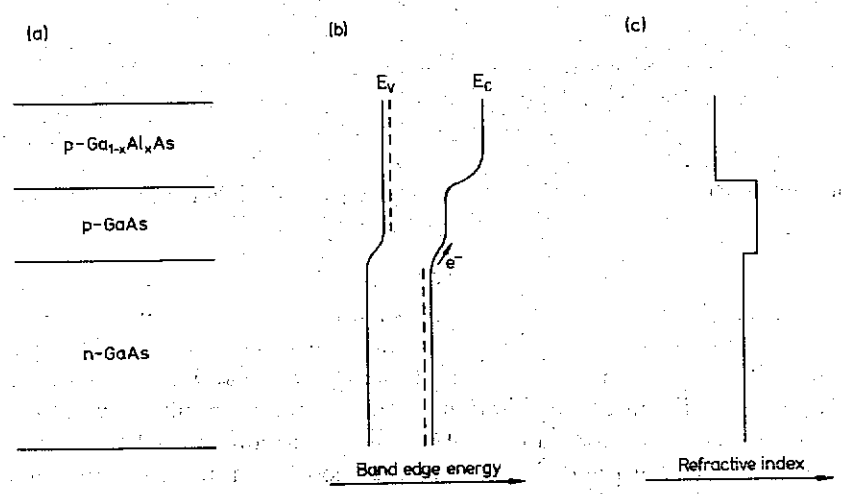
$$g(\omega) = -\alpha(\omega). \tag{7.1}$$

To produce a sharp lasing line, opposing faces in the plane perpendicular to the junction are cleaved to make them optically reflecting. If a unidirectional laser is required, the two remaining faces are roughened, leaving a Fabry–Perot structure. Stimulated emission is then greatest at the frequency at which the gain most closely approaches the cavity losses, the emission line narrows, and finally lasing takes place.

The dependence of gain on frequency, temperature and carrier population determines such direct observables as the temperature and doping dependence of the threshold current at which lasing begins and the frequency of the lasing line. The effects of heavy doping on the band-gap, density of states and optical matrix elements influence these observables via the gain, and it is this subject which we shall discuss in § 7.2.

In the early devices the junction was formed by diffusing acceptors, usually Zn, into n-type material. Mobility ratios and the Moss–Burstein shift favour electron injection into the p region, although more majority carriers are drawn in to preserve space-charge neutrality. The injected electron population is

Fig. 18



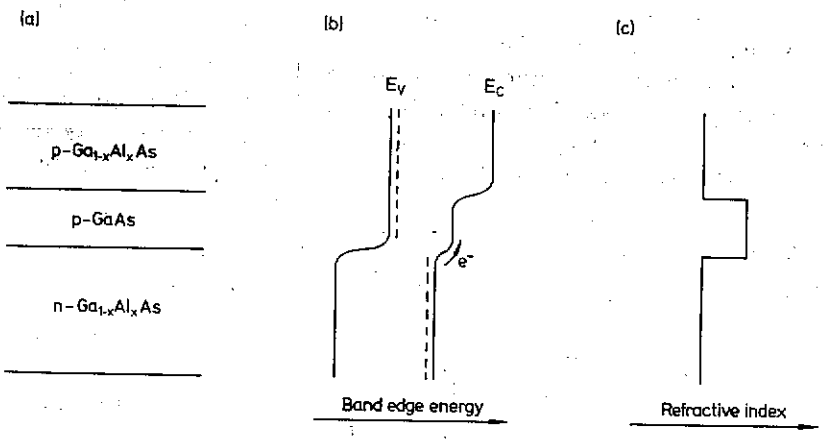
Single heterostructure. (*a*) Layer arrangement, (*b*) band-edge variation, (*c*) refractive index variation.

limited to a depth of the order of the minority carrier diffusion length, modified by the drift of minority carriers in the field built in by the non-uniform doping, which therefore determines the depth of the active region. The current required to achieve inversion is proportional to the pumped volume. The introduction of a layer of p-$Ga_{1-x}Al_xAs$, with a band-gap wider than in GaAs,

on the p side of the junction confines injected electrons at the potential barrier equal in height to the band-gap difference (see fig. 18). If the confinement depth is less than a diffusion length, the volume of the active region is reduced with a consequent reduction in threshold current. The wider gap material also has a smaller dielectric constant and these single heterostructure (SH) lasers provide the added advantages of reduced spreading of the optical field on the p side and better optical coupling to the active region.

To improve optical confinement on the n side, the n-GaAs layer was replaced by a layer of n-$Ga_{1-x}Al_xAs$ to form a double heterostructure (DH) (see fig. 19). The increased guiding allowed the width of the active region to be reduced still further, while maintaining strong coupling to the optical field, with a further reduction in the threshold current density.

Fig. 19



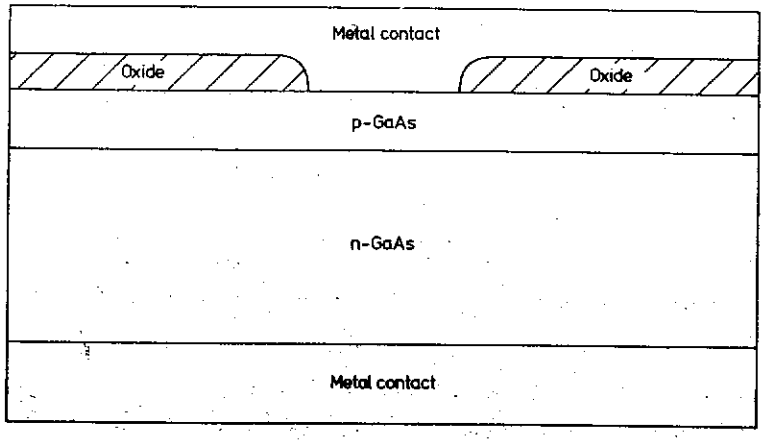
Double heterostructure. (*a*) Layer arrangement, (*b*) band-edge variation, (*c*) refractive index variation.

The degree of optical confinement of a transverse optical mode determines its ability to couple to the active region, and therefore determines the gain in the mode ; it also controls the beam width at its source and its far-field spreading. Optical confinement perpendicular to the junction plane was not well understood in homojunction lasers, although it was appreciated that both real and imaginary parts of the dielectric constant could be important and that the effects of heavy doping, doping type and levels of inversion could be critical in determining dielectric constant changes. The intentional introduction of real guiding in the DH laser reduced the importance of these considerations.

In a typical early device, 500 μm long and 50 μm wide, with an active region a fraction of a micron deep, the near-field beam dimensions have an aspect ratio of over 50 : 1, which is inconveniently large for injecting light into cylindrically symmetrical optical fibres. Moreover, it was often observed that light was emitted non-uniformly from the end face, either because high-order transverse modes were excited or because lasing action was taking place in filaments only a few microns wide. It is now generally accepted that the formation and

stabilization of filaments is related to the connection between transverse variations in the junction plane of carrier concentration and the dielectric constant, both real and imaginary, although the detailed description is still in dispute.

Fig. 20



Facet view of oxide isolated stripe geometry homojunction laser.

In order to produce emission from the lowest-order transverse mode and to eliminate filamentary lasing, devices have been designed to confine the lasing action to a narrow stripe about 10 μm wide in the transverse direction, parallel to the junction plane. A common and conveniently produced form of stripe geometry laser is one in which only a narrow stripe of active region is pumped, either because the metal contact is limited in lateral extent (see fig. 20) or because all but a narrow stripe of material is made semi-insulating by proton bombardment. Other techniques for confining the active region to a narrow stripe are reviewed by Panish (1976). In this case no intentional optical guiding in the transverse plane is introduced, and lateral variations in the dielectric constant produced by variations in the injected carrier concentration resume the importance they had for perpendicular confinement in homostructure lasers.

The relationship between doping and carrier concentration, complex dielectric constant, optical guiding and its feedback effect on carrier concentration via stimulated emission is reviewed in § 7.3.

When lasers are switched on or modulated, transient or resonance phenomena are observed. These are governed by the time-dependent behaviour of carrier and photon populations, coupled by the gain function. The effects of heavy doping on time-dependent behaviour are discussed in § 7.4.

7.2. *Gain*

In a medium where the pn product exceeds its equilibrium value, it is usually possible to describe the electron and hole populations by Fermi–Dirac functions at the same temperature with distinct quasi-Fermi levels (QFLs).

This simplification is allowed because inter-band recombination times are typically several orders of magnitude longer than the time for equilibration within a band, a process mediated by collisions with phonons and other electrons. Under these conditions it may be shown (Lasher and Stern 1964, eqn. 10) that the gain $g(\omega)$ at angular frequency ω and the spontaneous recombination rate $r_{\text{spon}}(\omega)$ are related simply by

$$g(\omega) = -\frac{\pi^2 c^2}{n^2 \omega^2} r_{\text{spon}}(\omega)[\exp((\hbar\omega - \Delta F)/kT) - 1], \tag{7.2}$$

where the electron and hole QFLs are separated by $\Delta F = F_e - F_h$, $r_{\text{spon}}(\omega)\, d\omega\, d\Omega/4\pi$ represents the rate at which photons are spontaneously emitted per unit volume into solid angle $d\Omega$ in the frequency range $d\omega$ at ω, and n represents the refractive index. This is the same definition of $r_{\text{spon}}(\omega)$ as given by eqn. (5.13), and differs from that of Lasher and Stern since ω is used as the variable. Since the spontaneous emission rate is never negative, it follows from eqn. (7.2) that for positive gain $g(\omega)$ we must have

$$\Delta F > \hbar\omega. \tag{7.3}$$

For lasing to take place a stronger condition must be met : the gain due to stimulated emission must equal the net round-trip loss due to free carrier absorption, inter-band absorption of radiation spreading into the surrounding passive medium and imperfect reflection at the facets :

$$g(\omega) = \alpha_{\text{loss}} + L^{-1} \ln(R^{-1}), \tag{7.4}$$

where α_{loss} is the effective distributed loss coefficient, L is the cavity length and R is the facet reflectivity.

To understand some of the general behaviour of lasers we may argue as follows, considering for definiteness electron injection into p-type material. At levels of injection current density J below the threshold, recombination with holes of density p proceeds by spontaneous emission, and the equilibrium minority carrier population n is given by

$$n = J/Bpd, \tag{7.5}$$

where d is the active layer thickness and B is the bimolecular recombination coefficient (see § 5.6). As the current increases at constant temperature, so does the electron population, and the electron QFL moves up the conduction band. The condition given by relation (7.3) is met for a range of frequencies above the optical edge. As the current increases further, this energy range extends and the maximum gain increases until at some energy it reaches the threshold and lasing takes place. The precise lasing frequency is determined by the cavity resonance condition—that an integral number of wavelengths fits into the Fabry–Perot structure. If the current is maintained at threshold level and the temperature is increased, two effects are likely to quench lasing. Firstly, if we suppose that the carrier population remains fixed, the QFLs will move together so that the gain at the lasing energy is reduced, as may be seen from eqn. (7.2). Secondly, the free carrier absorption is likely to increase (Turner and Reese 1964). Consequently, to maintain the threshold condition the injection level must increase and the lasing line shift to a new position determined by eqn. (7.4).

The increase of threshold current with temperature, which is characteristic of all injection lasers, is often strongly non-linear and is exacerbated by the increased Joule heating in the series resistance under continuous operation. Increasing the doping concentration in the active region moves the majority carrier QFL into the majority carrier band, making lasing easier according to eqn. (7.3), but more difficult by eqn. (7.5), because the equilibrium minority carrier concentration, and hence ΔF, is reduced. The increase in free carrier losses may also increase the threshold. Increasing the impurity concentration in the active region will affect the spectral dependence of gain via changes in the density of states of the bands due to the disordered potential. The absolute value of lasing energy is shifted by the many-body Coulomb effects of heavy doping and high-level injection.

The first attempt at predicting the temperature dependence of the threshold current to incorporate some of the effects of heavy doping was made by Lasher and Stern (1964) and was intended to model lasing from a heavily doped p-type region. Subsequent calculations have generally followed their procedure, which is now described.

By adopting a specific model for the density of states and for the optical matrix element, the spectral dependence of gain and spontaneous emission is evaluated from eqns. (5.6), (7.1) and (7.2). The spontaneous emission rate integrated over all energy is taken as a measure of the injection current density at and below the threshold. The approximation is reasonable since the injection current is seen to increase little in the super-radiance region in which the distributed gain in the active region is positive, but round-trip gain is suppressed by imperfect reflection at the facets (Cheroff *et al.* 1963).

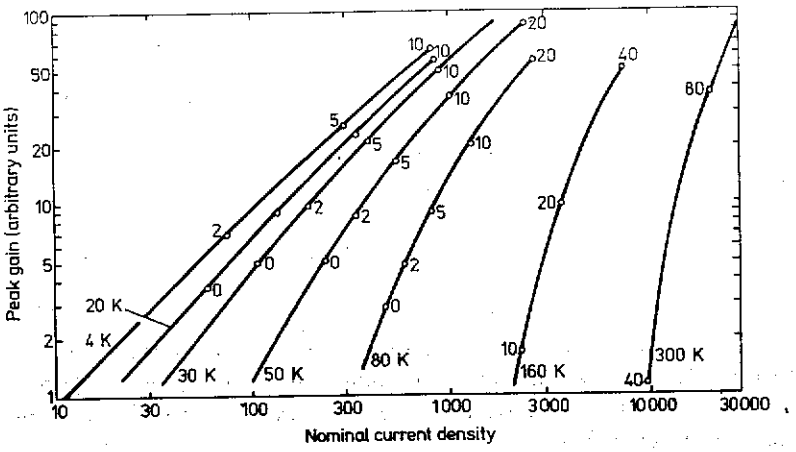
This relation between current density J and spontaneous emission rate R_{spon} can be made quantitative by estimating the internal quantum efficiency η and the active region depth d:

$$J = qdR_{\text{spon}}/\eta. \tag{7.6}$$

At a fixed temperature and for a given hole concentration, determined by the hole QFL, the spectral gain curve is recalculated for increasing minority carrier QFLs until the gain peak reaches the threshold, according to eqn. (7.4), for some values of mean absorption and coefficient and reflectance. The energy of the gain peak is taken to be the lasing energy, and the threshold current is deduced from the integrated spontaneous emission curve.

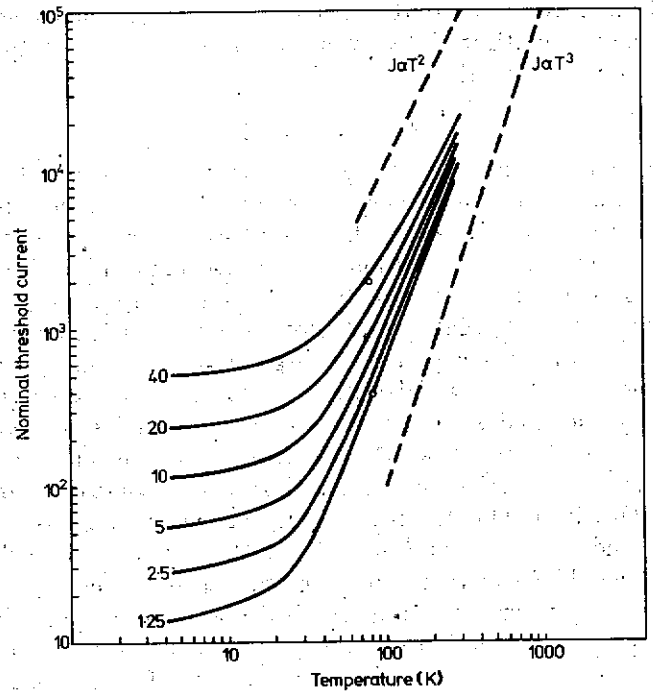
In this early work Lasher and Stern argued that at high acceptor doping levels the **k**-selection rule would break down for optical transitions. On the assumption that the final states at the valence band edge looked something like shallow acceptors, and that the electrons were initially in states near $\mathbf{k} = 0$ at the conduction band-edge, these authors used a constant matrix element evaluated by Dumke (1963) for transitions between the bottom of the conduction band and a shallow acceptor. On the grounds that the acceptor impurity band would be merged with the valence band, however, they used densities of states appropriate to unperturbed parabolic bands. Their predictions of peak gain as a function of current at a variety of temperatures are shown in fig. 21. Assuming that the major source of loss in a Fabry–Perot cavity is imperfect reflection, they chose a range of gain threshold values independent of temperature so that the temperature variation of the threshold

Fig. 21



Peak gain against nominal current density (after Lasher and Stern 1964).

Fig. 22



Nominal threshold current against temperature (after Lasher and Stern 1964).

current arises solely from the variations in QFL. For threshold gains in the range 40–100 cm^{-1}, the threshold current increases with temperature, slowly at first, and then above 20 K more rapidly as T^n where $2 < n < 3$, as shown in fig. 22.

The strong dependence on temperature is in marked contrast to the behaviour predicted by Hall (1963), who used a model with **k**-conserving optical matrix elements and densities of states appropriate to parabolic bands to find an approximately linear dependence of threshold current on temperature above 100 K.

At the time the experimental evidence favoured the results of Lasher and Stern. Burns *et al.* (1963) observed J_{th} varying as T^3 above 20 K for a diode formed by Zn diffusion into a 5×10^{17} n-type substrate, although the interpretation of Lasher and Stern was later questioned by Lamorte and Gonda (1966) who claimed that the temperature dependence of the threshold current was dominated by increased free carrier absorption losses.

Lasher and Stern also calculated the spontaneous line shape, but did not predict the low-energy tail in emission observed by Nelson *et al.* (1963) and Burns and Nathan (1964) and attributed by these authors to tail states. By replacing the $E^{1/2}$ conduction band density of states with an exponentially increasing function

$$N(E) \sim \exp(E/E_0), \tag{7.7}$$

and preserving the constant matrix element, Lasher and Stern achieved a fit to experiments at 20 K with a value of $E_0 = 15$ meV. It was later pointed out by Hwang (1970 b) that the tails on the low-temperature emission spectra obtained from low-bias diode measurements are produced by the radiative recombination of electrons and holes tunnelling into the space-charge region (Casey and Silversmith 1969).

Adams (1969) has shown that by assuming exponentially increasing densities of states at both band-edges, and by using a matrix element independent of energy and wave-vector, many results on lasing properties may be obtained analytically.

In an attempt to include the effects of doping on the density of states, Stern (1966) subsequently used Kane's semi-classical model (see § 4.2) with a self-consistently calculated screening length and the same constant optical matrix element as used by Lasher and Stern. This band tail model produced curves of maximum gain which varied less rapidly with current density than did the free electron density of states, in closer agreement with the experiments of Pilkuhn *et al.* (1964). These workers deduced a linear relation between maximum gain and injection current from measurements of threshold current versus laser length on diodes made by Zn diffusion into 5×10^{18} Sn-doped n-type layers.

Stern's calculations also illustrated the dependence of threshold current on the level of compensation in the active region for a fixed hole concentration, in this case 3×10^{18} cm^{-3}. In all cases the threshold current varied as T^2 at high temperatures, but was saturating at low temperatures. However, as the compensation level increased, this saturation set in at a higher temperature and at a higher level, whereas the coefficient of the T^2 dependence decreased, giving an overall weaker dependence on temperature, explained as follows.

The principal effect of increasing the compensation is to accentuate the band tails and to remove the rapid increase in the joint density of states near the optical edge. The spontaneous emission of radiation above the gain peak is wasteful of current intended to increase population inversion and gain. This waste is reduced if the joint density of states above the gain peak can be made to increase less rapidly. The benefit of this reduction is felt more at high temperatures, when the carrier populations are more spread out energetically.

By suspending the **k**-selection rule, Unger (1967 a, b) also calculated curves of threshold current against temperature for parabolic bands and for bands with Kane tails produced by heavy doping. The results of his analytical calculations performed by replacing the density of states in energy by one which is energy-independent, but whose magnitude depends on the degree of inversion, are in overall agreement with the corresponding results of Lasher and Stern (1964) and of Stern (1966).

Stern's band tail calculations were refined by Hwang (1970 b, c) by using the model of Halperin and Lax (see § 4.3) to describe the density of states deep in the tails, joined smoothly to the $E^{1/2}$ dependence for the unperturbed bands; the model gives significantly less pronounced tails than that of Kane. For the optical matrix element Hwang used the model of Dumke for conduction-band–acceptor transitions, but included the dependence of energy of the initial state in the conduction band. Some attempt was also made to account for the many-body band-gap narrowing by including the $n^{1/3}$ exchange term derived from a Wigner–Seitz model (Mott and Jones 1936), and also the direct screened Coulomb interaction with the impurities. This Coulomb term was shown earlier by Wolff (1962) to be cancelled by the interaction with the carriers; anyway, such a contribution should result in parallel shifts of the conduction and valence band-edges.

From this model Hwang deduced relations between maximum gain and current density intermediate between those predicted without band tails by Lasher and Stern and those with deep Kane band tails (Stern 1966). For an acceptor concentration of $N_A = 4 \times 10^{18}$ compensated with 10^{18} donors the maximum gain varied approximately as $g_{max} \sim J^b$ where b increased from 1·65 at 77 K to 3·7 at 300 K. Hwang claimed good agreement with unpublished work (Goodwin 1969). The predicted temperature dependence of the threshold current was compared with experiments on diffused diodes. For this comparison the calculations took detailed account of the effects of carrier diffusion and recombination in a material with non-uniform doping concentration, although no mention is made of the effect of the associated built-in field (Smith 1964), which is thought to be of equal importance to diffusion in diffused spontaneous diodes (Rees and Milne 1977). The calculations and experiments agree remarkably though, showing an almost exponential dependence of threshold current density on temperature over two decades. The difference between this exponential dependence and the near-power-law behaviour which Hwang predicts earlier in his calculation for a uniformly doped active region, also predicted by Stern (1966) and Lasher and Stern (1964), highlights the difference in behaviour that can be expected between diffused and epitaxially grown diodes. It must, therefore, throw some doubt on the value of other comparisons between experiments on diffused diodes and calculations on homogeneous active regions. By taking account of the effects of reabsorption,

Hwang also produced good agreement with spontaneous emission line shapes : the calculated and measured peaks agreed to within 5 meV, despite the deficiencies of his model for band-gap narrowing.

Measurements on lasers can also be processed to produce information on band tails, as shown by Moss *et al.* (1969). Using a method which combined measurements of gain versus current, with the assumption of constant matrix element and exponentially shaped band tails (see eqn. (7.7)), these workers deduced values of E_0 of about 8 meV for the conduction and valence band tails of diodes made by Zn diffusion into $1{\cdot}3\times10^{18}$ cm^{-3} n-type material. It should be noted, however, that the spectral dependence of spontaneous emission on injection levels was assumed to be determined entirely by the redistribution of carriers among fixed densities of states and the possible variable screening of band tails and band-gap shifts were ignored.

Perhaps the most realistic attempt to account for heavy doping effects in spectral gain calculations was made by Stern (1972, 1973). His model for the density of states and optical matrix element was subsequently used with considerable success by Casey and Stern (see § 5.4) to describe the absorption profile in p^+ GaAs. The calculation was intended to compare the properties of p- and n-type compensated active regions and served to show that n-type material should exhibit the lower room-temperature threshold current and the weaker temperature dependence. At low temperatures the maximum gain–current relations all approached linearity, and showed an approximate power-law dependence, $g_{max}\sim J^b$, at high temperatures where the index $b(T)$ increases with temperature most rapidly for lightly compensated p-type active regions and slowest for lightly compensated n-type regions.

The dependence of gain on current density may be inferred from measurements of differential external quantum efficiency η_e and threshold current density as a function of cavity length, as shown by Goodwin and Selway (1970) and by Goodwin and Thompson (1970). If the gain becomes pinned at the threshold at the level g_T, then the external quantum efficiency is given by

$$\eta_e=\eta_i(g_T-\alpha)/g_T,$$

where η_i represents the internal efficiency (Baird *et al.* 1964). Eliminating the loss coefficient with eqn. (7.4) at the threshold, we find that

$$g_T=\eta_i \ln (R^{-1})/L\eta_e,$$

and if we assume that $g\sim J^b$ then a plot of $\ln J$ against $(\eta_e L)^{-1}$ gives the value of b.

Goodwin and Selway (1970) found that $b=3$ for a single heterostructure laser at 300 K, and Goodwin and Thompson (1970) were able to fit the temperature variation of b to the relation

$$b=(1+(kT/E_T)^2)^{1/2},$$

which followed from Unger's (1967 b) exponential band tail arguments. Goodwin and Thompson deduced values of E_T, a measure of the extent of the band tail, of 9 meV for a SH and 7·5 meV for a diffused homostructure device.

Pinkas *et al.* (1972) used simpler arguments which assumed a loss coefficient independent of cavity length, supported by their observation of a linear dependence of η_e^{-1} on cavity length, to deduce values of b for various levels of

doping and compensation in double heterostructure lasers. They found values of b as large as 3 in heavily doped p- and n-type active layers, reducing to about 1·5 as the level of compensation increased. This behaviour is in general agreement with the prediction of Stern (1972, 1973).

Hayashi and Panish (1970) studied the temperature dependence of threshold current in SH lasers with p-type active regions. They found an exponential dependence,

$$J_{\text{th}} \sim \exp\,(T/T_0),$$

with T_0 in the region of 90 K, followed by a sharp rise at higher temperatures which they attributed to hole injection into the n-GaAs layer. Similar measurements by Hayashi *et al.* (1971) on DH lasers revealed an exponential dependence with higher values of T_0, increasing as the active region was narrowed, but without the hole-injection knee.

Stern (1976) later used the same model to make detailed predictions of gain against photon energy and current density. In an attempt to allow for many-body effects an empirical band-gap shrinkage,

$$E_g = -1{\cdot}6 \times 10^{-8}(p^{1/3} + n^{1/3}),$$

was used, as suggested by the results of Casey and Stern (1976). The calculations on a uniformly doped 4×10^{17} p-type active region were performed for direct comparison with the experiments of Hakki and Paoli (1975). These workers used measurements of the peaks and troughs introduced into the super-radiant spectrum by the Fabry–Perot resonances to obtain gain as a function of photon energy and current density. The measurements show a linear dependence of gain on current at all measured wavelengths below the threshold, and on this point the calculations are in reasonable agreement with a calculated proportionality constant intermediate between that measured by Hakki and Paoli and that measured by Thompson *et al.* (1976). However, the measured lasing peaks lie about 30 meV higher than predicted, throwing some doubt on the treatment of Coulomb effects, although the inadvertent inclusion of AlAs in the active region of the heterostructure would cause lasing peaks to shift to the blue. To avoid comparison of absolute values, Stern plots $1/J_{\text{tb}}\,(\mathrm{d}g/\mathrm{d}J)_{\text{tb}}$ against $E - E_{\text{peak}}$ and finds qualitative agreement.

It is interesting to note that, even at the low doping level of $p = 4 \times 10^{17}$ cm^{-3}, Hakki and Paoli observed a logarithmic dependence of lasing energy on current density characteristic of an exponential band tail (see eqn. (7.7)) with $E_0 = 26$ meV.

Stern (1973) also performed calculations on undoped material, assuming **k** conservation and including the effects of conduction band non-parabolicity and valence band degeneracy and warping. He found gains varying almost linearly with current density, from which we can deduce that there is an approximately linear relation between current and temperature for constant gain above 100 K.

In an attempt to test the models for optical matrix element, Bakker and Ackett (1977) compared the spectral dependence of gain, observed directly by measuring the amplification through an optically pumped region, with calculations both using and suspending the **k**-selection rule. The calculations predict that the slope of the gain–frequency curve initially decreases with **k** selection

and increases without **k** selection. At high pump levels the regions near the imperfectly reflecting facets support high optical fields and stimulated emission depresses the injected carrier concentration. Lowering of the QFL selectively reduces gain at the high photon energies, which is shown by sub-threshold measurements of spectral gain curves (Goebel *et al.* 1977). These authors compared their measurements with the results of calculations based on the four model combinations with and without Kane band tails and with and without **k**-selection rules. They concluded that the combination of no **k** selection with Kane tails best fits the results, even for lightly doped active regions, although the agreement is not so good as to preclude more refined models. They concluded that the selection rule was more nearly obeyed in samples with low hole concentration, but was broken at high levels of doping.

More recently Hwang *et al.* (1978) have observed that in double heterostructure lasers the threshold current–temperature curves are independent of the active region doping level above 100 K, suggesting that the band tails are irrelevant and that the optical transition conserves **k** between the unperturbed bands. Their conjecture is supported by the remarkably accurate agreement between their experiment and the curves of temperature dependence of threshold current calculated from Stern's (1973) undoped crystal model.

The effects of high levels of excess carriers in the absence of impurities were considered theoretically, for low temperatures, by Brinkman and Rice (1973) in connection with the energetics of electron hole droplet formation, and more especially for lasers by Brinkman and Lee (1973), where several new effects are predicted. They may be summarized as follows.

(1) An enhancement of gain by a factor of about 2 at low temperature and at an injected carrier concentration of $n = 2 \times 10^{16}$ cm^{-3}, interpreted as extension of exciton enhancement.

(2) A tail below the optical edge in the spectral gain curve due to stimulated emission associated with Auger processes involving the emission of a plasmon or single-particle excitation. In this case incoming radiation at an energy E, below the optical edge, can stimulate an electron and hole to recombine, emitting radiant energy E, and either generating a plasmon or delivering the extra energy to a free carrier.

(3) A tail on the conduction band due to the disordered potential of the randomly situated heavy holes, analogous to the band tail produced by ionized donors.

Evidence of the Auger tail in the spectral gain curve has been seen by Shacklee *et al.* (1971), and a low-energy tail on the photoemission spectrum of CdS, whose extent increases with drive intensity, has been ascribed to the increasing effects of the disordered potential with excitation level by Januzzi *et al.* (1972).

The values of gain measured by Pinkas *et al.* (1972) are consistently higher by a factor of about 2 than those predicted by Stern (1972) and Hwang (1970 c). Moreover, Shah *et al.* (1977) see a persistent enhancement of absorption at pump levels when the exciton line is screened out. Both these results are in agreement with the predictions of persistent exciton enhancement at high levels of carrier concentration.

The results of various model calculations on gain and threshold current show some general trends, which are now summarized. Calculations which assume the intrinsic semiconductor properties of **k** conservation in the optical transition between unperturbed densities of states (Hall 1963, Stern 1973) predict a roughly linear dependence of peak gain on current density and a weak dependence of threshold current on temperature, roughly independent below 100 K and increasing linearly above. Relaxing the **k**-selection rule to allow recombination between electrons and holes in parabolic bands with a probability which is independent of the initial and final states (Lasher and Stern 1964) makes for a stronger, power-law dependence of peak gain on current density where the power index increases with temperature. This strengthening of dependencies also results from the calculations of Hwang (1970 b, c) where the Halperin–Lax model for the density of states produced only a weak deviation from the unperturbed density of states and the inclusion of the energy dependence of the matrix element does not seriously inhibit recombination between the thermally occupied parts of the bands. On the other hand, if, in addition to relaxing the **k**-selection rule, the sharp threshold in the density of states at the optical edge is removed by allowing for considerable tailing on the Kane model (Stern 1966), the dependences of peak gain on current density and of threshold current density on temperature are softened again. The range of low temperature in which the current density is roughly constant is increased and a weaker dependence is predicted above this range. Both these deviations from the properties of an intrinsic semiconductor, namely loss of **k** conservation and the introduction of band tails, are likely to occur in heavily doped and especially in heavily compensated material, but since the effects they produce are opposed it is difficult to predict general trends.

In diffused diodes, where compensation is likely to be heavier, a T^3 dependence of threshold current is observed (Burns *et al.* 1963, Pilkuhn *et al.* 1964) in contrast to the weaker dependence seen in epitaxially grown diodes (Dousmanis *et al.* 1964) which, though possibly heavily doped, are not intentionally compensated.

The danger of making a direct comparison between calculations on homogeneous active layers and experiments on diffused diodes with heterogeneous active regions is clear from the work of Hwang (1970 c), who showed that inhomogeneity could change a power-law temperature dependence of threshold current to an exponential dependence.

Perhaps the simplest system to study is the now almost exclusively fabricated epitaxially grown heterostructure laser, recently results on which (Hwang *et al.* 1978) are best explained in terms of the unperturbed crystal model first used by Hall (1963).

The incorporation of band-gap narrowing effects due to the interaction of many free carriers should permit the prediction of the energy of lasing or peak in spontaneous emission as a function of doping and drive level, although the comparison will also be affected by the density of states and optical transition process. Probably the soundest calculation to incorporate many-body effects is that due to Stern (1976), where the calculated and measured value of gain peak and lasing energy differ by around 30 meV. The success obtained by Hwang (1970 c) in predicting the spontaneous peak to within 5 meV must be

regarded as fortuitous because of the inclusion of the direct interaction with ionized impurities.

Since that time our understanding of band-gap narrowing effects has progressed and improved calculations are now possible, but little progress has been made on the effect on density of states and optical matrix elements. *Ab initio* calculations of densities of states by Kane and Halperin and Lax are appropriate in limited energy ranges only and cannot deal with the intermediate region of doping when the impurity band is in the process of merging with the parent band. Consequently, these models are invariably extended beyond the doping range in which they apply, or are arbitrarily joined smoothly to the $E^{1/2}$ density of states of the unperturbed band. The models of the optical matrix element that treat breakdown due to disorder of the **k**-selection rule (Lasher and Stern 1964, Stern 1973) are at best phenomenological and, moreover, no attempt has been made to incorporate the free carrier many-body effects thought to be so important in determining energies.

7.3. *Optical guiding*

To begin this section on the effects of heavy doping on optical guiding, we review the work aimed at determining the spectral variation of refractive index in heavily doped material and material with a high concentration of injected carriers.

General expressions for the absorption coefficient and refractive index as a function of wavelength are given in eqns. (5.1)–(5.6). It is seen that the values of both $\alpha(\lambda)$ and $n(\lambda)$ near the absorption edge depend on the energy gap via the joint density of states, on the band-edge wavefunctions via the optical matrix element, and on the carrier concentration via the occupation probabilities $f(E)$. Heavy doping and population inversion will affect the spectral dependence of $\alpha(\lambda)$ and $n(\lambda)$ through these quantities. Since $n(\lambda)$ is the result of virtual transitions its dependence on these effects is weaker than that of $\alpha(\lambda)$. However, even small variations in refractive index can be crucial in optical guiding.

The spectral variation of $n(\lambda)$ can be expressed in terms of the spectral variation of $\alpha(\lambda)$ through dispersion relations (Stern 1963). These formulae are convenient for visualizing the effects of heavy doping on $n(\lambda)$ via the more easily appreciated effects on $\alpha(\lambda)$, and for inferring the dependence of $n(\lambda)$ on the observed variation of $\alpha(\lambda)$. They have also been used to deduce $n(\lambda)$ from first-principle calculations of $\alpha(\lambda)$, although the fundamental formulae (5.1) and (5.2) provide a more direct route.

The dispersion relation between the refractive index n and the extinction coefficient k in the form

$$n^2(E)-k^2(E)-1=\frac{4}{\pi}\int_0^{E_s} n(E')k(E')(E'^2-E^2)^{-1}\,\mathrm{d}E'+\sum_j G_j(E_j^2-E^2)^{-1},$$

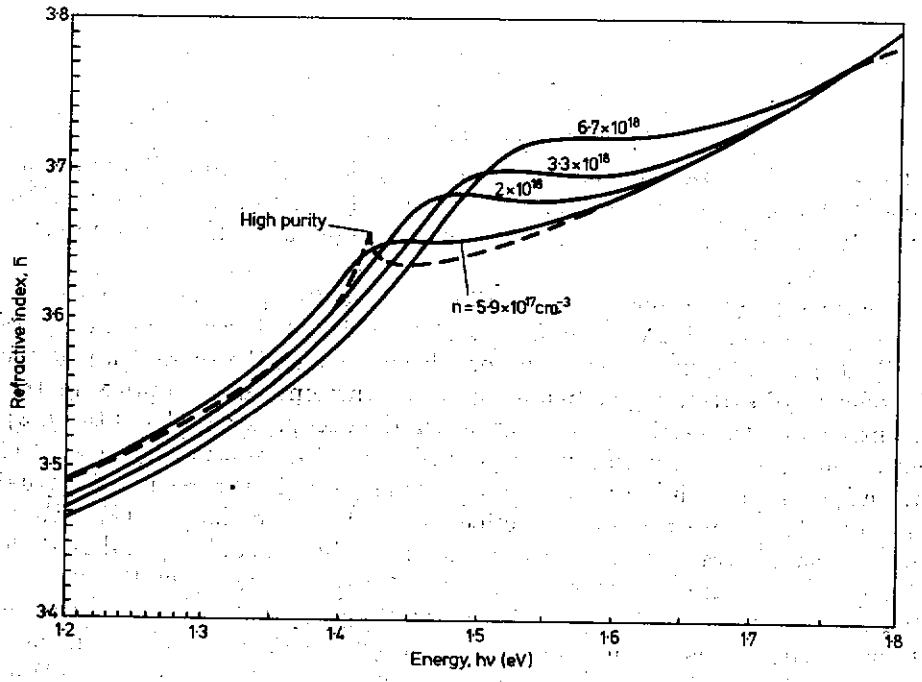
was used by Stern (1964) to show the effects of absorption on refractive index around the optical edge. The range of the integral over E' is artificially restricted by supposing that $k(E)$ is negligible above $E=E_s$, apart from some delta-function peaks of strength G_j at E_j. The relation shows that a material with a sharp absorption edge will also exhibit a peak in the refractive index near the edge. In this way data on the absorption coefficient obtained by Kudman

and Seidel (1962) on heavily doped p-type GaAs were processed to show structure in $n(\lambda)$.

Similarly, Zoroofchi and Butler (1973) processed the absorption data of Hwang (1969) to obtain the dispersion of refractive index in heavily doped n-type GaAs.

Sell *et al.* (1974) made reflectance measurements on n- and p-type GaAs from which they deduced directly the spectral dependence of refractive index (see fig. 23). Their more accurate results confirmed the trends predicted by the Kramers–Krönig analyses. The softening of the peak at the optical gap with increased doping is more noticeable in p-type than in n-type samples, the

Fig. 23



Spectral dependence of refractive index in n-GaAs (after Sell *et al.* 1974).

Moss–Burstein blue shift is evident in n-type material and the red shift, attributed to a narrowing of the energy gap, appears in p-type material. These workers also measured the refractive index of GaAs doped to about 10^{19} cm^{-3} with Si. The measured carrier concentration of $p = 2 \times 10^{18}$ is a result of heavy compensation between Si_{Ga} donors and Si_{As} acceptors, and it is interesting to note that the peak in the refractive index is considerably flatter than expected for an uncompensated sample of equivalent carrier concentration, probably as a result of increased band tailing.

These effects are more directly evident in the spectral dependence of the absorption coefficient, the most recent measurements of which are discussed in § 5.8.

For the specific purpose of investigating optical guiding in lasers, Cross and Adams (1974) devised a simple model for the variation in passive and active material by using the concept of an effective band-gap. In this model the variation of refractive index with energy, simply parametrized by fitting to measurements on intrinsic GaAs (Marples 1964) is supposed to be shifted rigidly by an amount equal to the change in the effective band-gap. The band-gap is changed

(1) by the extent to which the quasi-Fermi levels move above their respective band edges, thus allowing for the Moss–Burstein shift observed in $n(\lambda)$ in heavily doped n-type GaAs,
(2) by the average screened impurity potential seen by free carriers and
(3) by the screened exchange term evaluated with the Wigner–Seitz expression (Mott and Jones 1936).

It is not difficult to find fault with this model. For instance, no account is taken of the change in density of states in (1), and Wolff (1962) (see § 4.1) has argued that the term (2) should not appear. The optical matrix element is assumed not to vary and generally the procedure would appear to be more appropriate for modelling the absorption coefficient, whose properties are determined by real transitions. However, it remains the only attempt to represent the effects of heavy doping on $n(\lambda)$ for device modelling purposes.

Thompson (1972) has also estimated the effects of injected carriers on refractive index. With a variety of models for densities of states at the band edges, he calculated the effects of Moss–Burstein shift on optical absorption and, using the dispersion relation, he showed that the refractive index decreases approximately linearly with carrier concentration at a rate in the range 2·5 to $6{\cdot}5\times10^{-20}$ cm^3. This 'band-to-band' effect is rather larger than the reduction of refractive index due to the response of the plasma with increased free carrier concentration (Jonscher and Boyle 1966). In GaAs this plasma effect makes the refractive index decrease at the rate of $1{\cdot}2\times10^{-20}$ cm^3 (Thompson 1972).

The laser itself offers a means of measuring the spectral dependence of refractive index in inverted materials (Nathan *et al.* 1963 b). The Fabry–Perot condition determining allowed longitudinal modes of order m is

$$m\lambda = 2n(\lambda)L,$$

so that successive modes are separated in wavelength by

$$\Delta\lambda = \lambda^2/2n'L,$$

where $n' = n - \lambda\, \mathrm{d}n/\mathrm{d}\lambda$.

Measurement of the separation of peaks in the luminescence below the threshold allows n' to be determined, although it should be noted that the quantity $n(\lambda)$ determining $n'(\lambda)$ is a weighted average over the region occupied by the optical field which extends laterally into the passive material.

7.3.1. *Depth confinement*

The problem of optical confinement perpendicular to the junction plane was appreciated early in the development of diffused homojunction lasers (see, for example, McWhorter *et al.* 1963) and it was realized that any variation in both the real and imaginary parts of the refractive index could be important. Hatz

and Mohn (1967) allowed for a continuous, linear variation of optical constants in their calculations, consistent with the absence of a sharp boundary confining radiation or carriers. Zachos and Ripper (1969) observed that a parabolic decrease in dielectric constants could generate the observed field distribution. However, apart from the contribution of free carriers (Yariv and Leite 1963) no detailed models were offered for these variations.

Adams and Cross (1971) analysed the structure of transverse optical modes in SH and DH lasers by considering a three-layer waveguide model within which the optical constants are assumed to be uniform. To obtain values for $n(\lambda)$ and $\alpha(\lambda)$ they used a simplified effective gap model, allowing only for Moss–Burstein shift due to the carriers originating from impurities and the concentration of Al in the $Al_xGa_{1-x}As$ layers. In their parametrization of the absorption coefficients the blurring of the absorption edge was related to the square root of the doping concentration (Unger 1967 b). They also allowed for free carrier absorption. Qualitative agreement was obtained with observations of optimum active layer width and with the dependence of threshold current on the mode order and width of the active layer. The reduced dependence of guiding on the doping level in the active region of a DH laser is due to the relatively more important effect of alloy concentration on refractive index via the change in energy gap.

In the large optical cavity (LOC), one of the variants of the double heterostructure (Lockwood *et al.* 1970), the GaAs sandwich filling is subdivided into p-type (active) and n-type (passive) material. The optical field is still confined between the heterojunctions, whose increased separation spreads the field over a larger volume so that high-power operation does not result in damage to the facets. The active region, where gain is positive, occupies only a fraction of the waveguide with consequently reduced threshold current. By varying this fraction lasing can be stabilized in the lowest-order transverse mode, even at high power levels (Paoli *et al.* 1973). Hakki's (1974) calculations of mode stability assumed a linear relation between gain and carrier concentration (Hakki and Paoli 1973), which was later confirmed by Hakki and Paoli (1975), and correctly predicted an optimum fraction of $\frac{2}{3}$.

Measurements of near-field and far-field radiation patterns give information on transverse mode structure (Kirkby and Thompson 1972). Zoroofchi and Butler (1973) have used calculations of the spectral dependence of refractive index on doping level, rigidly shifted in energy for the wider gap $Al_xGa_{1-x}As$, to predict radiation patterns in a p^+pnn^+ LOC structure and find good agreement with experiment.

7.3.2. *Lateral confinement and filamentary lasing*

In early device fabrication no attempt was made at transverse optical confinement in the junction plane, although the active region often appeared to be spontaneously confined to longitudinal filaments. In later devices the minority carriers were injected into a narrow stripe in an attempt to reduce the aspect ratio of the beam dimensions and to control the formation of filaments. Optical guiding mechanisms for filamentary and stripe-controlled lasing are related and will be discussed in parallel. Discontinuities in the differential power efficiency, known as kinks, have been ascribed to changes in optical guiding mechanisms and will also be discussed.

In one of the first attempts to describe the structure of the optical field in a stripe geometry laser, Cross and Adams (1972) applied their dielectric model, previously used to describe depth confinement, to treat the three-dimensional dielectric slab waveguide. Their model included the effects of uniform gain in the active region. The refractive index appears to peak in the active region, although it is not clear from their paper why this is so. The predictions of higher losses and increased threshold current densities over large contact area devices are in broad agreement with observations (see, for example, Miller *et al.* 1971) and in this work appear to be connected with a reduction in depth confinement associated with the introduction of confinement in the junction plane.

The dielectric model of Thompson (1972), which predicts a reduction in the refractive index with free carrier concentration due to both plasma response and the 'band-to-band' contribution, has already been discussed. The immediate consequence of this index reduction in the high injected carrier concentration under the stripe is to inhibit guiding, since total internal reflection requires a decrease in refractive index away from the core. However, Thompson argued that stimulated emission in a region of high light intensity locally depresses the excess carrier concentration, locally increasing the refractive index and stabilizing guiding. Application of detailed arguments to this model suggested that there could exist stable filaments with a width of around 10 μm which decreases with current density as observed by Matthews *et al.* (1972) because higher refractive index steps can support a more closely confined transverse field.

Whilst it was appreciated that the full complex dielectric constant determines the optical field pattern, Schlosser (1973) showed that an increase in absorption coefficient at the core–cladding interface of a dielectric slab waveguide could result in optical guiding, even in the absence of a real refractive index step. The possibility of 'gain guiding' by the change of sign in the absorption coefficient from active to passive material was investigated by Nash (1973), who showed that typical values of gain and loss coefficient could support at least the low-order transverse modes observed in a stripe geometry laser. On the question of filaments, Nash argues that intrinsic guiding mechanisms, unrelated to spatial inhomogeneities, are inherently unstable and he favours the view that filamentation is related to material imperfections. Hence a local reduction in carrier concentration leading to real guiding via the mechanisms considered by Thompson (1972) will be accompanied by a local reduction in gain, so that the filament should wander laterally.

The mechanism of gain guiding under a stripe was investigated further by Hakki (1973) who monitored the long-wavelength spontaneous emission across a stripe in order to find the proportional variation of minority carrier concentration (see eqn. (5.15)). Using a linear relation between carrier concentration and gain, and arbitrarily assuming a trigonometrically varying field distribution precisely confined to the stripe, Hakki calculated the net gain in a mode, including the effects of loss from optical penetration into the sub-threshold regions. He was thus able to correlate the current at the lasing threshold, where net mode gain balances reflection loss at the facet, with stripe width. Agreement with experiment (Dyment *et al.* 1969) is good, and the derived linear relation between local gain and current density agrees well with the result predicted by Stern (1973).

The mechanism of gain guiding, which produces a curved wavefront in a bell-shaped gain profile, received further support from the observation by Cook and Nash (1975) of astigmatism in the emergent beam. Hakki (1975) made his earlier calculation (Hakki 1973) self-consistent by calculating the transverse optical field profile in the varying gain, assumed to be related linearly to carrier concentration, whose profile was in turn governed by recombination and diffusion away from the stripe. The self-consistent calculations of mode gain as a function of current also reproduced the observed sharp increase in threshold current density with reduced stripe width and provided design criteria relating stripe width and diffusion length to emergent beam width.

Paoli's (1977) observations of spontaneous emission across a narrow (12 μm) stripe are in very close agreement with the expected carrier profile determined by out-diffusion. Calculations which assume a linear dependence of gain on carrier concentration produce excellent agreement with experiment on optical mode width. No evidence was found for a defocusing refractive index variation deduced by Cook and Nash (1975).

By contrast, Kirkby *et al.* (1977) find evidence for both gain guiding and refractive index defocusing in narrow (10 μm) stripe lasers. From measurements on the emergent beam and the spontaneous emission profile these workers deduced the parameters describing the parabolic transverse variation of gain and dielectric constant.

On the other hand, at high power levels, or in wide (20 μm) stripe lasers, a parabolic increase in gain and decrease in refractive index away from the centre is deduced and is confirmed by the observation of a depression of the central spontaneous emission peak, produced by hole burning in the high optical field region. This inverse correlation between gain and refractive index in both narrow- and wide-stripe lasers provides strong support for Thompson's (1972) dielectric model. Self-consistent calculations after the style of Hakki (1975), but allowing for the real refractive index variation, produced values of beam parameters consistent with those measured and also predicted a sharp change in the differential power efficiency as the guiding mechanism changes from gain to real refractive index variation. This effect is offered as an explanation for observed kinks in the curve of power against current density (see, for example, Dixon *et al.* 1976). The calculations of Chinone (1977) suggest that the strong dependence of guiding on gain profile can produce apparent kinks and, furthermore, that any lateral asymmetry will serve to enhance the effect.

7.4. *Transient response*

The response of a laser to a variable pump level I carriers per unit volume per second is determined by the time-dependent behaviour of the carrier concentration n and photon density φ. The coupled equations describing their motion and their solutions are discussed in a review of transient phenomena by Adams (1973).

In a laser whose active region is assumed to be spatially uniform, the equations become

$$\left.\begin{aligned} \mathrm{d}n/\mathrm{d}t &= I(t) - n/\tau_{\mathrm{e}} - g(n)\varphi, \\ \text{and} \quad \mathrm{d}\varphi/\mathrm{d}t &= g(n)\varphi - \varphi/\tau_{\varphi}, \end{aligned}\right\} \tag{7.8}$$

where τ_e and τ_q are the spontaneous lifetime and photon decay time and $g(n)$ is the peak gain of the spectral gain curve. Heavy doping can affect the form of these equations in several ways. As we have seen in § 7.2 the relationship between peak gain and carrier concentration depends on doping and compensation level. The photon decay time depends on the degree of optical guiding through diffraction loss, which may be influenced by the active region carrier concentration via the refractive index. It has also been found useful, in order to explain some long delays, to introduce a third rate equation to distinguish between carriers in the band tail and those in the bulk density of states of the band.

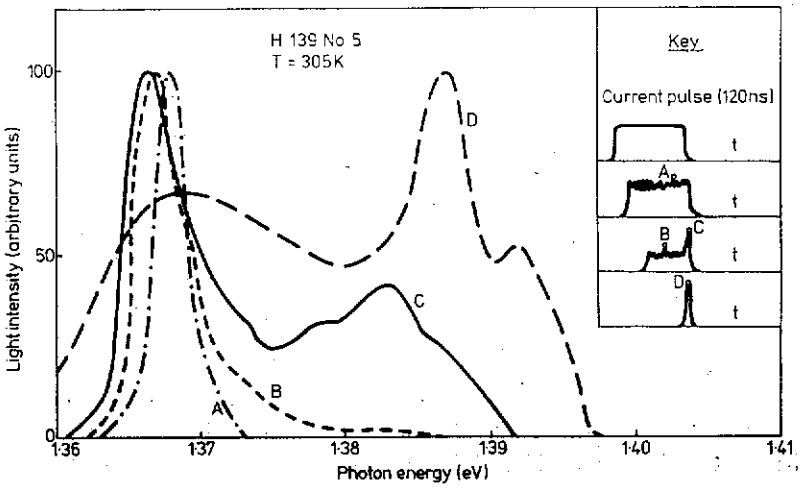
The predator–prey relations described by eqn. (7.8) are non-linear. Numerical solutions may be obtained to model the large signal transients at turn-on, when periodic decaying spikes in the photon population are predicted as the electron population falls steeply after a slow rise between the optical spikes. Alternatively, by linearizing the equations around the steady state, analytical solutions are obtained which model the small-signal optical modulation efficiency. The small-signal analysis provides an estimate of the spiking frequency ν_r and decay constant τ_0 at turn-on, and predicts a sharp peak in the modulation efficiency at ν_r. Introducing noise generators into the rate equations accounts for the discreet nature of n and φ, and the small-signal solution around the steady state provides the noise spectrum which also shows a peak at ν_r.

The influence of heavy doping on these transient properties via the gain function $g(n)$ has been investigated by several authors and the reader is referred to the review by Adams (1973) for more information.

Cross (1973) solved the rate equations (7.8) to analyse the behaviour of a filament at turn-on. The filament is supposed to be guided by the central peak in refractive index produced by the locally depressed carrier concentration. The transverse extent of the optical field, and hence its coupling to the region of gain, is allowed to vary as the carrier concentration oscillates. The width of the optical spikes, about 0·1 ns, and the period of oscillation, about 1 ns, are of the order seen experimentally (Roldan 1967).

Short delays of the order of 1 ns are invariably seen between the beginning of the current pulse and the commencement of lasing and are attributed (Konnerth and Lanza 1964) to the time required to build up the carrier concentration to the threshold value of lasing while injection at a constant rate is in competition with the spontaneous recombination rate, $n(t)/\tau_e$. These are to be distinguished from the longer delays sometimes observed (see fig. 24) and which have been explained in a variety of ways, including that of ' saturable absorption ' (Pankove 1968 b) at band tails whose population is in disequilibrium with that of the bulk band. Adams *et al.* (1973) proposed a detailed model to account for these delays, and for the existence of a critical temperature above which these delays are enhanced. They suggested that lasing in a p-type active region may be associated with electrons originating from either the conduction band tail or from the bulk of the density of states in the band. At low temperatures electrons are injected directly into the tail and lasing takes place with short delay. At high temperatures electrons are more readily injected into the more mobile states in the bulk of the band. Here they thermalize to establish population inversion so that stimulated emission occurs.

Fig. 24



Schematic of delayed emission (inset) and spectral dependence : (*a*) normal lasing, (*b*) time delay, (*c*) *Q*-switching following delayed emission, (*d*) pure *Q*-switching (after Adams *et al.* 1973).[1]

These band electrons are not thought to be in thermal equilibrium with those in the partially empty tail (Redfield *et al.* 1970) and reabsorption of light takes place, exciting electrons from the valence band to the tail states until the latter are saturated. Gain then exceeds loss and lasing commences. Gründorfer and Adams (1973) produced a quantitative analysis of this model by introducing a third rate equation to describe the population of the tail states. Thomas *et al.* (1974) measured the spectral emission and threshold current from lasers exhibiting delays and demonstrated an increase in photon energy of around 10 meV and of threshold current above the critical temperature, confirming the hypothesis of a shift from tail to bulk band of the emitted electrons and the increased absorption.

In some lasers optical emission is briefly enhanced at the end of a current pulse and in certain circumstances this is the only emission observed. Gründorfer and Adams (1974) provided an explanation for this phenomenon of internal *Q*-switching observed in homostructures and SH lasers. They argued that optical guiding at a homojunction is critically affected by the carrier concentration in the active region. During the current pulse the concentration is high, the refractive index is depressed and the optical field spreads into the passive material with the result that the gain is reduced, possibly even to below the threshold. At the end of the pulse, when the carrier concentration begins to collapse, the optical guiding and associated gain improves, resulting in enhanced emission. Thomas *et al.* (1975) later combined these ideas of delayed guiding and *Q*-switching to produce a comprehensive rate equation description. Their model was able to reproduce the rich variety in the variation of time delay with drive current and temperature and to predict correctly the regions of threshold current and temperature where spontaneous emission, lasing and *Q*-switching are likely to take place.

With their ideas of optical guiding they were also able to explain the phenomenon of ' H-pulsing ' (Gründorfer *et al.* 1975) sometimes observed in SH lasers. Using a simple extension of their Q-switching arguments that gain exceeds diffraction loss only over a narrow range of injected carrier concentration, they showed that this may occur only at the beginning and end of a current pulse, producing an initial and final burst of stimulated emission. If net gain is positive but reduced in the intervening time, then an H-sphaed emission profile will result.

Nunes *et al.* (1976) have suggested that variable optical guiding alone can account for both long delays and Q-switching. At the beginning of the current pulse the depression of the refractive index by the injected carriers results in poor guiding and sub-threshold gain. Eventually, non-radiative recombination may heat the active region sufficiently to reduce its band-gap, thereby increasing its refractive index, improving guiding and permitting delayed lasing. The guiding will be further improved at the end of the current pulse as the carrier concentration collapses, resulting in a final burst of emission. These plausible arguments are supported by more detailed calculations (Nunes *et al.* 1977), but they do not seem able to account for the existence of a threshold temperature above which delays increase, or for the blue shift of emission above this temperature.

§ 8. The silicon bipolar transistor

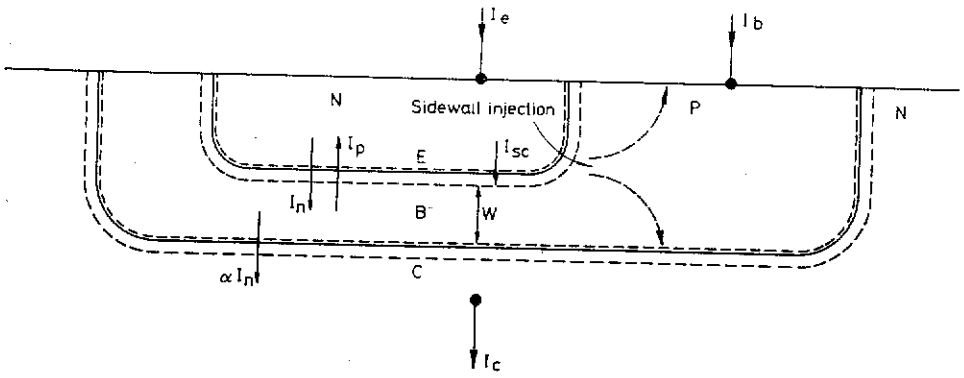
8.1. *Introduction*

The *npn* silicon bipolar transistor relies on the injection of electrons from the n-type emitter into a p-type base across a forward-biased n–p junction. Nearly all the injected electrons flow across the base to the reverse-biased collector p–n junction. Thus the collector current I_c responds to the base-emitter voltage V_{BE}, which also generates a much smaller current I_b through the base terminal. While a part of this current is due to the small minority, $1-\alpha$, of electrons that fail to cross the base but recombine with holes in the base, the greater part is usually due to hole injection from the base into the emitter. The scale of the injected current densities can be conveniently related by considering the electron and hole current densities, J_{n0} and J_{p0}, which compensate them when $V_{BE}=0$ due to spontaneous generation in the base and emitter respectively. These are in turn proportional to the minority carrier concentrations n_p and p_n. Shrinkage of the energy gap in either region, but particularly in the emitter, increases the minority carrier concentration above the value calculated classically from the majority carrier concentrations n_n and p_p and changes the values of J_{n0} and J_{p0}. Grove (1967) provides a good general reference for standard p–n junction theory.

Reduction in the effective band-gap seriously affects the performance of silicon bipolar transistors which normally have a heavily doped emitter and a moderately heavily doped base. The current gain β is defined as the ratio of collector to base current, I_c/I_b, and in practical *npn* transistors is determined largely by the ratio β_γ of the electron current I_n to the hole current I_p at the emitter–base junction. Referring to fig. 25, if α is the fraction of electrons in an *npn* transistor which reach the collector before recombination,

$$I_c=\alpha I_n, \tag{8.1}$$

Fig. 25



Planar bipolar transistor showing n-type emitter (E), p-type base (B) of width W and n-type collector (C). The dashed lines show the associated depletion regions. The currents shown are described in the text.

and

$$I_b = (1-\alpha)I_n + I_p, \tag{8.2}$$

whence

$$I_b/I_c = \beta^{-1} = \beta_\alpha^{-1} + \beta_\gamma^{-1} + \beta_\alpha^{-1}\beta_\gamma^{-1}, \tag{8.3}$$

where

$$\beta_\alpha = \alpha/(1-\alpha). \tag{8.4}$$

We have neglected generation in the emitter–base space-charge region I_{sc}, which must be added to I_b, and electron injection from the collector, which becomes important only in saturation, when the collector–base junction is forward-biased. Usually $1-\alpha$ is small and $\beta = \beta_\gamma$.

If the reduction in the effective band-gap is neglected, much higher values of β are predicted than are found in practice. In order to isolate as far as possible band-gap shrinkage from other effects, it is convenient to begin by describing the careful work of Slotboom and de Graaf (1976) on the dependence of collector current on emitter–base voltage V_{BE}, base doping and temperature.

8.2. *Calculation of the collector current*

The collector current is given by eqn. (8.1). Considering one-dimensional current flow in the base of an *npn* transistor, the electron current density J_n at the emitter–base junction is given by

$$J_n = J_{n0}\{\exp(eV_{BE}/kT) - 1\}, \tag{8.5}$$

where for uniform base doping across a base of width W and negligible recombination in the base,

$$J_{n0} = \frac{D_n n_p e}{W} = \frac{\mu_n kT n_p}{W}. \tag{8.6}$$

The expression for J_{n0} describes the diffusion of the minority electrons of equilibrium density n_p across the base to the collector edge with a diffusion coefficient D_n, equal to $\mu_n kT/e$ by the Einstein relation. It can be shown

(Wilson 1977) that eqn. (8.5) is true, even when there are fields built in to the neutral region by impurity gradients in the base, if eqn. (8.6) is generalized to

$$J_{n0} = \frac{kT}{\int\limits_{\substack{\text{neutral}\\ \text{base}}} \frac{\mathrm{d}x}{\mu_n n_p}}. \tag{8.7}$$

Here, as in eqn. (8.6), n_p is the local minority carrier density that would be present in the absence of diffusion. The exponential factor in eqn. (8.5) arises from the Boltzmann distribution factor across a well-defined emitter–base depletion region. J_{n0} is independent of voltage only if the neutral base width involved in eqn. (8.7) is not appreciably modulated by variations in V_{BE}. On the other hand, the interpretation of experiments is simplified if all the injected electrons appear as collector current, or at least if $1-\alpha \ll \beta_\gamma^{-1} \ll 1$, so that this sets an upper limit to the base width.

The expression (8.7) can be rewritten in terms of majority carriers. For any semiconductor with parabolic bands where Boltzmann statistics apply, we can write classically (Putley and Mitchell 1958)

$$p_p n_p = {n_\mathrm{i}}^2 = N_\mathrm{v} N_\mathrm{c} \exp\{-E_\mathrm{g}(0)/kT\}, \tag{8.8}$$

where p_p is the majority hole concentration, $E_\mathrm{g}(0)$ is the energy gap for zero impurity concentration and N_v and N_c are effective densities of states at the valence and conduction band edges.

The temperature dependence may be removed from eqn. (8.8) as it is found, near room-temperature, that

$$E_\mathrm{g}(0) = E_{\mathrm{g}0}(0) + \chi T, \tag{8.9}$$

where $E_{\mathrm{g}0}$ is the energy gap extrapolated to 0 K and χ is a constant. Extracting the $T^{3/2}$ dependence of N_v and N_c, we have that

$$p_p n_p = {n_\mathrm{i}}^2 = CT^3 \exp\{-E_{\mathrm{g}0}(0)/kT\}. \tag{8.10}$$

In formal analogy we define an effective energy gap $E_\mathrm{g}(N_\mathrm{A})$ for high impurity concentration, and an effective intrinsic carrier concentration n_{ie} by

$$p_p n_p = {n_{\mathrm{ie}}}^2 = N_\mathrm{v} N_\mathrm{c} \exp\{-E_\mathrm{g}(N_\mathrm{A})/kT\} \tag{8.11}$$

$$= CT^3 \exp\{-E_{\mathrm{g}0}(N_\mathrm{A})/kT\}. \tag{8.12}$$

Then, using an average mobility across the base region, eqn. (8.7) becomes

$$J_{n0} = kT\mu_n {n_\mathrm{i}}^2 / Q_{\mathrm{Be}}, \tag{8.13}$$

where Q_B is the effective base charge, given by

$$Q_{\mathrm{Be}} = \int\limits_{\text{base}} p_p({n_\mathrm{i}}^2/{n_{\mathrm{ie}}}^2)\,\mathrm{d}x \equiv \int\limits_{\text{base}} p_\mathrm{e}\,\mathrm{d}x, \tag{8.14}$$

defining p_e as an effective hole concentration.

Shrinkage of the band-gap between rigid bands, band tails, impurity bands and the effects of degeneracy can all be described formally by eqns. (8.11)–(8.14), although there is no *a priori* reason to expect that the gap shrinkage,

$$-\Delta E_g(N_A) = E_{g0}(0) - E_{g0}(N_A) = kT \ln (n_{ie}^2/n_i^2), \tag{8.15}$$

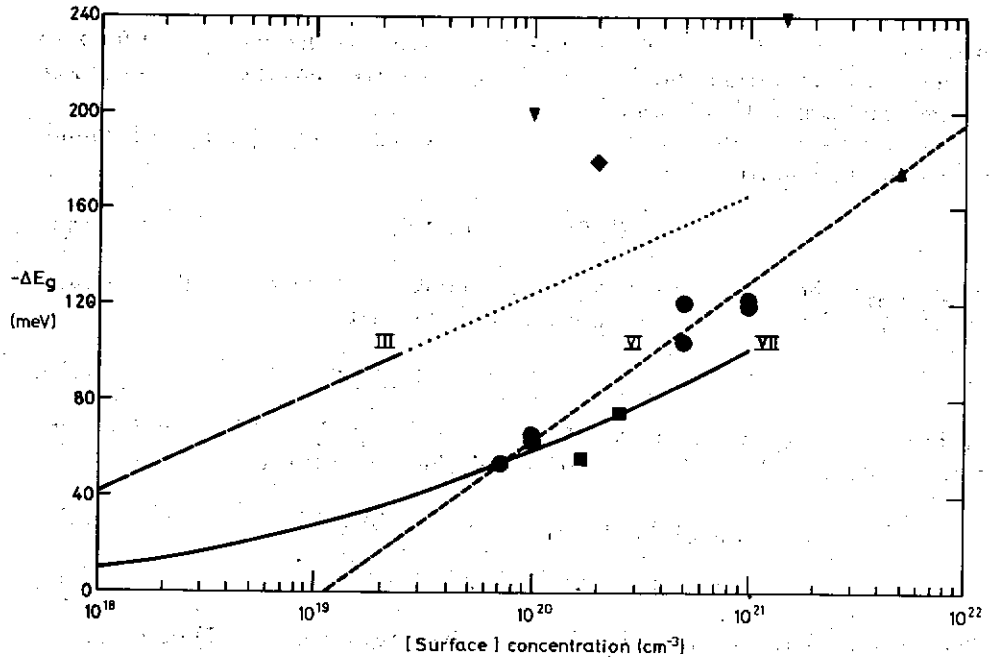
will always be independent of temperature, as it is for a shift in rigid bands with classical statistics. Although the value of ΔE_g varies across the base-doping profile, experiments determine an average value. Note also how experiments which determine an apparent activation energy of J_{n0} in fact determine

$$E_g(N_A) + (4 - \nu_n)kT,$$

when $\mu_n \sim T^{-\nu_n}$, as can be shown from eqns. (8.12) and (8.13). In practical transistors, usually $\nu_n \sim 1{\cdot}5$.

Slotboom and de Graaf (1976) compared eqns. (8.5)–(8.14) with experiment at room temperature for a number of *npn* transistors with different base impurity concentrations. In devices for which $(1-\alpha) \ll 1$, measurement of the collector current gave J_n directly and hence J_{n0} of eqn. (8.13). The value

Fig. 26



Band-gap shrinkage as a function of surface concentration deduced from measurements of transistor characteristics. ▼, Kauffman and Bergh (1968) from $I_B(T)$, phosphorus doping ; ▲, Kannam (1973) from $\beta(T)$, phosphorus doping ; ●, Buhanan (1969) from $\beta(T)$, phosphorus doping ; ■, Fair (1973) from $\beta(T)$, arsenic doping ; ◆, Lindholm (1977) from solar cell characteristics, phosphorus doping. Curve III, Slotboom and de Graaf (1976) from $I_C(T)$, boron doping ; also de Graaf *et al.* (1977), see text, boron and phosphorus doping. Curve VI, a ' best ' line for donors. Curve VII, theoretical curve of Mock (1973) calculated from tail and impurity states.

of the base charge $Q_B = \int p_p \, dx$ was found by measuring the sheet conductance in the base region under the emitter, and hence an average $\Delta E_g(N_A)$ was obtained, obtaining Q_{BE} from eqn. (8.13), and then using eqns. (8.14) and (8.15), assuming that the value of μ_n was the same as in n-type material of similar impurity concentration. An independent check was obtained from the temperature dependence of V_{BE} at constant J_n (i.e. collector current) in eqn. (8.5). To interpret this a knowledge of $\mu_n(T)$ was required and was obtained from the largely verified assumption that the temperature dependence of electron and hole mobilities is the same at a given impurity concentration. The second method gave results in good agreement with the first, consistent with the rigid band model in that $\Delta E_g(N_A)$ was independent of temperature and the T^3 factor in eqn. (8.8) was verified. This does not imply that the bands actually move rigidly, and the result was later related by Slotboom (1977) to a model involving tail states and impurity bands. The results are included in fig. 26, and the shrinkage is related to the impurity concentration N by the equation

$$-\Delta E_g(N) = 9 \times 10^{-3}\{\ln \overline{10^{-17}N} + \sqrt{[(\ln \overline{10^{-17}N})^2 + \tfrac{1}{2}]}\}. \qquad (8.16)$$

If the experiments are interpreted in terms of Boltzmann statistics, the effects of degeneracy will be wrongly ascribed to a contribution to the shift in the band-gap. The correct degenerate statistics expression for eqn. (8.9) should be

$$p_p n_p = N_v N_c \mathscr{F}_{1/2}(E_F/kT) \exp\{-(E_g(N_A) + E_F)/kT\}, \qquad (8.17)$$

where $\mathscr{F}_{1/2}$ is the Fermi–Dirac integral and E_F is the depth of the Fermi level below the valence band edge. A comparison with eqn. (8.11) shows that the change arising solely from degeneracy is given by an apparent shrinkage

$$-\Delta E_g' = -E_F + kT \ln \mathscr{F}_{1/2}(E_F/kT). \qquad (8.18)$$

This term is always negative: degenerate statistics oppose the effect of true gap shrinkage in that they decrease the minority carrier concentration for a given majority carrier concentration compared with classical statistics. In other words, while band-gap shrinkage decreases the effective base charge, degeneracy increases it. In comparing gap shrinkage results obtained from such a transistor analysis with calculations for optical experiments, the term given by eqn. (8.18) should be subtracted from those directly derived from transistor characterization using the convenient fiction of classical statistics, as is done in figs. 27 and 28. In defence of those who use classical statistics in a situation outside their applicability, it may be remarked that this may be no more artificial than the assumption that degeneracy is approached without a change in the band shape. If classical statistics is retained as a framework for device engineering, conventional theory may be retained by the use of an appropriate value of n_{ie}^2/n_i^2 or $\Delta E_g(N_A)$.

8.3. *Calculation of the current gain*

Calculation of the current gain in transistors is now straightforward but, as we shall see, comparison with experiment is difficult. By analogy with eqns.

(8.5), (8.13) and (8.14), we find the hole current density J_p in the emitter to be

$$J_p = J_{p0}\{\exp(eV_{BE}/kT) - 1\}, \tag{8.19}$$

where

$$J_{p0} = kT\mu_p n_i^2/Q_{Ee}, \tag{8.20}$$

and the effective emitter charge to be

$$Q_{Ee} = \int_{\text{emitter}} n_n(n_i^2/n_{ie}^2)\,dx = \int_{\text{emitter}} n_e\,dx. \tag{8.21}$$

Equation (8.21) applies only where the hole current substantially crosses the whole of the emitter to a contact of high recombination velocity, otherwise the integration must be taken over only the relevant part of the emitter. We can now recover the classical expression of Tannenbaum and Thomas (1956) for current gain β, modified by the use of effective base and emitter charge :

$$\beta \simeq \beta_\gamma = \frac{J_n}{J_p} = \frac{\mu_n Q_{Ee}}{\mu_p Q_{Be}}. \tag{8.22}$$

8.4. *Difficulties in experimental interpretation*

There have been published several comparisons between observed current gain and that computed from either a rigid band interpretation of absorption data (de Man 1971, van Overstraeten *et al.* 1973) or *a priori* calculations of the density of states in tails and impurity bands (van Overstraeten *et al.* 1973, Mertens *et al.* 1973, Mock 1973). The comparisons are complicated by the three-dimensional nature of real transistors, by recombination in the emitter–base depletion region, and by inadequate knowledge of the impurity and lifetime profiles in the thin layers which constitute most modern silicon transistors. The commonly used assumptions that the profiles are error functions or Gaussian are seldom adequate. Furthermore, above impurity concentrations of about 2×10^{20} it is known from Rutherford back-scattering that non-substitutional impurities are present, which in the case of arsenic have been identified as an As_4 complex from vapour-pressure measurements (Sandhu and Reuter 1971). In extreme cases the impurity is precipitated on dislocations induced by strain—' diffusion damage ' (Jaccodine 1968).

Recombination in the emitter–base depletion region gives rise to majority carrier currents I_{sc} with a dependence given by $\exp(eV_{BE}/2kT)$. At high enough voltages this current is small, and can be neglected, when it is found that current gain is constant for a range of collector currents. At higher current β falls off, as conductivity modulation by the injected carriers in the base becomes important, particularly as Ohmic drop in the base region enhances injection at the lateral edge of the emitter. Recombination in the base may itself reduce the current gain β_α, notably in some unconventional transistor structures. Electrons injected from the side wall of the emitter (see fig. 25) will be more likely to recombine on the longer path to the collector, while others, particularly those injected nearer the surface than the peak acceptor concentration in a conventional double diffused transistor, may be swept to the oxide or metal interface. The lifetime expected from the Auger process (Beck and Conradt 1973), which appears to be the fundamental limitation in silicon, is such that minority carrier loss in the emitter and base does not severely limit

carrier transport in practical structures. However, heavily doped regions, particularly those with phosphorus concentrations greater than 2×10^{20} cm^{-3}, are often structurally defective. These regions are found, for example in solar cells, to have an extremely short lifetime, and may be associated with strained regions and states in the gap in addition to those considered in this review.

Double diffused and most implanted transistors contain highly compensated neutral regions where the band tails may be more pronounced than would be expected from the concentration of majority impurities. Although these effects might be computed if an adequate *a priori* theory existed, they are not easily accommodated by the simple scheme suggested earlier. The existence of tail states in the minority carrier band is itself worrying since the low minority carrier density means that the minority carriers travel in the tails, in contrast to the majority carrier tails which are largely filled ; this tends to undermine the assumption that the minority carrier (drift) mobility is equal to the majority carrier (conductivity) mobility. A further consequence of modifications to the minority band density of states is that this can lead to gradations in the effective density of states in the minority band. Carriers tend to remain in the region of higher density of states and so an effective diffusion term must be added to the effective field term incorporated in eqn. (8.7). Very recently, both generalized fields and generalized diffusion have been analysed by Marshak and van Vliet (1978 a, b) who cite earlier papers.

8.5 *Further evidence of band-gap shrinkage in silicon devices*

With these reservations in mind we now consider three further methods by which measurements on transistor structures can be used to give information on band-gap changes. Conversely, in so far as information on heavily doped silicon is available, it can be used in transistor design.

For $(1-\alpha) \ll \beta_\gamma^{-1}$ (see eqn. (8.2)), the base current density is given by J_p. The temperature dependence of base current gives, through eqns. (8.19)–(8.21) an ' emitter activation energy ' $-\Delta E_g(N_D)$, defined similarly to eqn. (8.15), providing a region where a voltage dependence given by eqn. (8.19) is not obscured by a current I_{sc} due to recombination in the space-charge region, or to other effects. Again, experiments determine an average ' activation energy ' for the emitter for a particular impurity profile. Note, however, that the activation energy of J_{p0} is $E_g(N_D) + (4 - \nu_p)kT$, where $\mu_p \sim T^{-\nu_p}$ and ν_p is near zero in the heavily doped material of the emitter.

Some representative values of the shrinkage $-\Delta E_g(N_D)$ obtained by Kauffman and Bergh (1968), Fair (1973) and Kannam (1973) are shown in fig. 26 as a function of surface donor concentration. If the whole of the emitter is operative, the shrinkage is a mean value associated with a concentration about half that at the surface. For emitters of very high surface concentration, the defective surface regions will not contribute to Q_{Ee} as their lifetime is so short that holes generated there cannot reach the junction, while in deeper emitters even the Auger limited lifetime may give a diffusion length shorter than the emitter depth.

Another approach is to measure the temperature dependence of current gain. If, over a limited range of temperature, an activation energy ΔE_β for current gain is defined by

$$\beta = \beta_0 \exp(-\Delta E_\beta / kT), \tag{8.23}$$

then consideration of eqn. (8.22) shows that

$$\Delta E_\beta = E_g(N_D) - E_g(N_A) + (\nu_n - \nu_p)kT, \tag{8.24}$$

where the difference in mobility exponents, $\nu_n - \nu_p$, is about 1·5 in typical transistors. Values of $-\Delta E_g(N_D)$ derived from data given by Buhanan (1969) and Kannam (1973) are also shown in fig. 26 as a function of surface concentration. Though rather easier to obtain, and practically of greater importance, these values are subject to all the qualifications associated with the base current results, with the minor additional uncertainty over the relevant value of $\Delta E_g(N_A)$.

Finally, ΔE_β may be obtained directly by comparing the observed value of β with the value calculated without energy gap shrinkage. If it yields values consistent with the earlier methods, it provides a check on the validity of the earlier assumptions on minority carrier flow in the emitter. Many attempts have been made to predict the reduction in current gain of transistors from the classical expression (8.22). An early attempt by de Man (1971) pointed out the main features to be expected using the gap shrinkage values derived from optical absorption by Vol'fson and Subashiev (1967), but neglected the effects of degeneracy summarized by eqn. (8.18). Subsequently, attempts have been made by Kleppinger and Lindholm (1971), van Overstraeten *et al.* (1973), Mock (1973) and Slotboom (1977) to derive density-of-states curves from first principles using the theories of Bonch-Bruevich (1966 a, b) and Kane (1963) for tail states and of Morgan (1965) for impurity bands. Some inadequacies in these approaches have been discussed by Wilson (1977), e.g. the omission of broadening by impurity state overlap in Morgan's theory, the merging of the impurity band with tail states in heavily doped material, the incomplete treatment of many-body effects and the variation of lifetime across the emitter. More work with a wide variety of transistor structures of known impurity profile and carefully measured electrical properties is required before such models can be regarded as quantitatively verified.

Recently, a different approach was adopted by de Graaf *et al.* (1977). In the absence of experimental results for $\Delta E_g(N_D)$ for *pnp* transistors comparable to those of Slotboom and de Graaf for $\Delta E_g(N_A)$ for *npn* transistors, they assumed that the shrinkage was independent of impurity type. The lifetime was assumed to be dominated by an Auger term and computer simulations were made to determine the effect of finite lifetime and different impurity profiles on the effective emitter charge; these resulted in a weighting of the actual emitter impurity density, nearly independent of the profile, described by an effective diffusion length $L_{p\ eff}$. The agreement between the measured and computed values of this parameter was fairly satisfactory, except for very shallow arsenic emitters to which the weighting procedure was unsuited. Confirmation was obtained by measuring the emitter charge as the emitter was etched away, when it was shown that a single value of $L_{p\ eff}$ described the different structures, although the surface recombination velocity also becomes important as the junction depth falls. The same theory was used to fit two novel transistor structures: the LEC (low emitter concentration; Yagi *et al.* 1974) and the substrate-fed logic transistor (Blatt *et al.* 1975). Despite many reservations and simplifications pointed out by the authors, particularly the use of a universal relation between lifetime and carrier concentration, the work

of Slotboom and his associates relies heavily on experiment and goes some way to establish a relationship between effective band-gap narrowing and donor or acceptor concentration which is independent of impurity type. The experimental results fit the relation (8.16), but it is much better established for acceptors.

Figure 15 shows that the emitter gap narrowing determined by absorption is usually less than that expected from eqn. (8.20), consistent with the hypothesis that the more heavily doped regions do not contribute fully to the effective emitter charge owing to reduced lifetime, so that the average band-gap shrinkage is appropriate to a concentration much less than the surface concentration. Note further that n_e (or p_e) rises only slowly for impurity concentrations $N > 10^{18}$, as shown in the table, so that the weighting of the average band-gap shrinkage by high concentration is less marked than might be supposed, even when the lifetime is long enough for the heavily doped material to contribute.

Comparison of impurity concentration N and effective carrier concentration n_e or p_e according to Slotboom's formula (8.16) for silicon

n	10^{17}	3×10^{17}	10^{18}	3×10^{18}	10^{19}
n_e	$0{\cdot}79\times10^{17}$	$1{\cdot}31\times10^{17}$	$2{\cdot}8\times10^{17}$	$3{\cdot}6\times10^{17}$	$4{\cdot}1\times10^{17}$
n	3×10^{19}	10^{20}	3×10^{20}	10^{21}	
n_e	$5{\cdot}7\times10^{17}$	$8{\cdot}2\times10^{17}$	$1{\cdot}16\times10^{18}$	$1{\cdot}68\times10^{18}$	

Finally, mention should be made of the importance of energy gap shrinkage in solar cell design which can, for example, lead to an appreciable component of dark current from the heavily doped surface region and hence a reduction in the open-circuit voltage under illumination. Lindholm *et al.* (1977) show how measurements of static and dynamic response characteristics can be used to derive a value of the energy gap and quote a preliminary value of 170 meV for shrinkage in an n^+–p cell for a surface concentration of 2×10^{20} cm^{-3}. This is included in fig. 26.

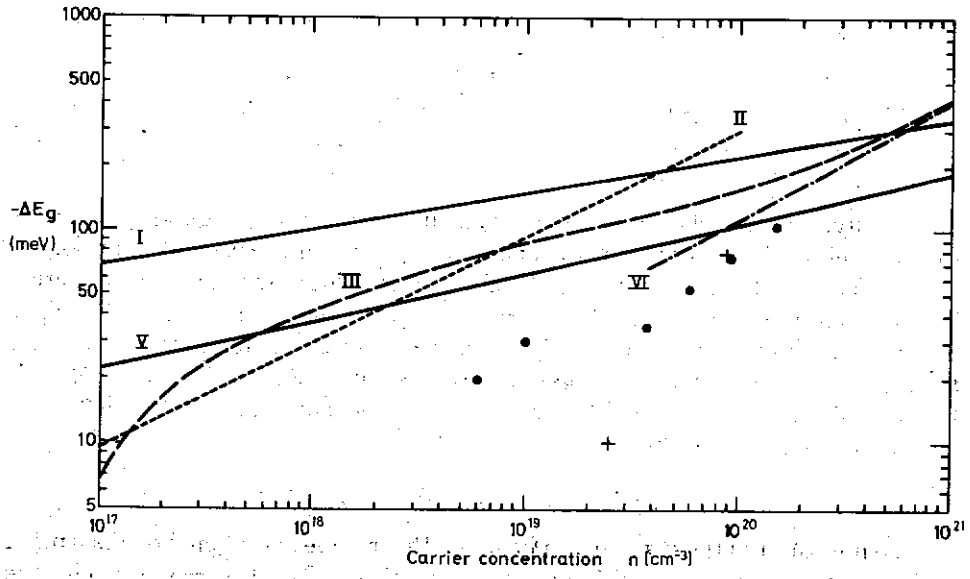
8.6. *Results and discussion*

We now discuss the comparison between theory, measurements of absorption and measurements of transistor characteristics. The theories include the various versions of the many-body theory discussed in § 3, theories of tail states (§ 4) and theories of impurity bands (§ 2). These effects are generally present simultaneously. The transistor measurements of base current and collector current in effect measure minority carrier concentration and are affected both by the rigid band shift and deviations from elementary theory associated with impurity and tail states, which can be artificially interpreted as being equivalent to a further band shift in their effect on minority carriers. For the prediction of minority carrier concentration, and hence the sense of injection, a combined theory of tail states, impurity bands and many-body effects is necessary. At the lower concentrations, the many-body theory of

itinerant carriers is less relevant, but becomes increasingly important at larger carrier concentrations. Any attempt to fit experiment to a simple theory over a wide range of doping is likely to be unsuccessful. Moreover, these measurements rely on the assumption that the minority carrier mobility is identical to that of the same sort of carrier as a majority carrier in material of the same impurity concentration. In fact, the minority carrier mobility is probably rather smaller in the minority band tails, and the above assumption has the effect of reducing the influence of the minority band tail on the gap narrowing results. In heavily doped material the majority band tail is substantially filled and will also have little influence. The precise quantitative significance of the effective band-gap measured by these experiments is not clear.

The absorption measurements involve, of course, sums over all possible transitions and with appropriate assumptions give information about the whole band structure. In this section we have extracted the information about the rigid band shift consistent with the optical measurements for comparison with the many-body theory and with transistor measurements and theory on very heavily doped material.

Fig. 27

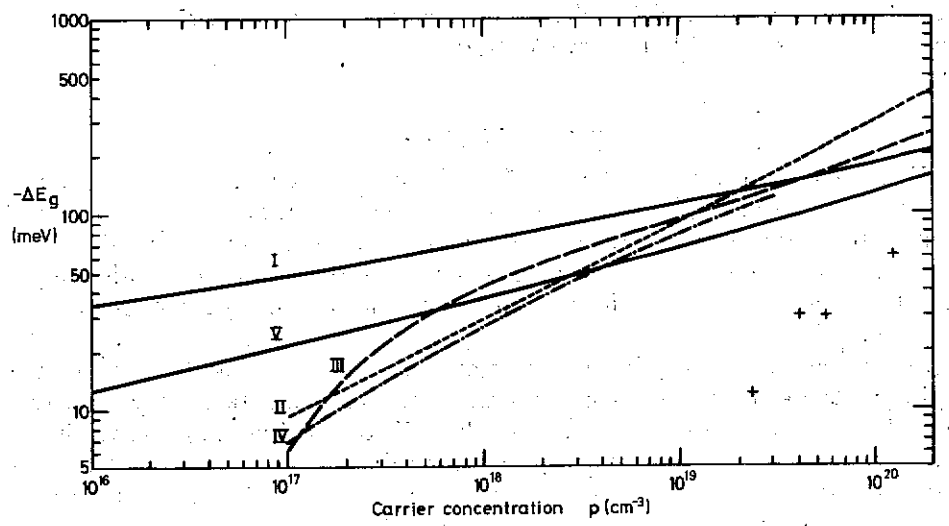


Band-gap shrinkage for n-type Si at 300 K. Curve I, theory ($-\Delta E_g^{TF}$, see § 3); curve II, theory ($-\Delta E_{g(Boltz)}^{TF}$, see § 3); curve III, assumed by de Graaf *et al.* (1977); curve V, theory ($-\Delta E_g^{L}$, see § 3); curve VI, consensus of transistor measurements in fig. 26. +, Absorption measurements (Vol'fson and Subashiev 1967); ●, absorption measurements (Balkanski *et al.* 1969).

At 300 K in n-type silicon, Balkanski's absorption measurements (Balkanski *et al.* 1969) of band-gap narrowing are shown in fig. 27 together with two results from Vol'fson and Subashiev (1967). Curve VI is equivalent to curve VI of fig. 26 and represents a consensus of transistor measurements. The mean

impurity concentration is taken to be half the surface concentration and eqn. (8.18) is used to correct for the degeneracy ignored in the transistor analysis. Curve III is our eqn. (8.16), used by de Graaf for n- and p-type silicon. In comparison, Inkson's gap narrowing result $-\Delta E_g{}^{TF}$, based on many-body effects alone (see § 3) for the fully degenerate case and shown in curve I, is too high except near $N_D = 10^{21}$, where it is now believed that only a fraction of the impurities give rise to free carriers. Inkson's result for the Boltzmann case $-\Delta E_{g(Boltz)}{}^{TF}$ is shown in curve II. Curve V shows the value of $-\Delta E_g{}^{L}$ using Lindhard's expression for the dielectric function in the many-body theory (see § 3). Agreement between this curve and both the transistor results VI and the absorption measurements is best around $N_D = 10^{20}$, when the assumption of full degeneracy is adequate and impurity bands have merged with the conduction band. The higher experimental values encountered above the solubility limit in the emitter may well be associated with structural changes. In any case,

Fig. 28



Band-gap shrinkage for p-type Si at 300 K. Curve I, theory ($-\Delta E_g{}^{TF}$, see § 3); curve II, theory ($-\Delta E_{g(Boltz)}{}^{TF}$, see § 3); curve III, experimentally fitted line from transistor measurements of Slotboom *et al.* (1976); curve IV, calculated from the tail and impurity state model of Slotboom (1977); curve V, theory ($-\Delta E_g{}^{L}$, see § 3); +, absorption measurements (Vol'fson and Subashiev 1967).

the theoretical curves refer to free carrier concentration. Below $N_D = 10^{20}$ the absorption measurements may disagree with $-\Delta E_g{}^{L}$ because of the stronger influence of tail states, or even of an impurity band. The failure of full degenerate statistics is clearly not the cause, as the same disagreement is evident in fig. 15 in the measurements at 35 K. Other results are shown in fig. 26 without correction for degeneracy where applicable. Curve III is the curve used by Slotboom in successfully modelling *npn* transistors, assuming band-gap narrowing to be the same in n- and p-type silicon. Curve VII can

be derived from the calculations of Mock (1973) for n_{ie} in the case $N_A = 0$, which uses expressions for impurity bands and tail states referred to earlier. More direct evidence for $N_D < 10^{19}$ is required, say, from measurements on *pnp* transistors.

Figure 28 shows results for *p*-type silicon at 300 K. The absorption measurements of Vol'fson and Subashiev are lower than all predicted values throughout.

Curve III shows the experimentally fitted line of Slotboom and de Graaf's (1976) *npn* transistor measurements with extrapolation above $N_A = 2{\cdot}5 \times 10^{19}$. In the region of impurity bands and tail states the theories of Morgan (1965), Kane (1963) and Bonch-Bruevich (1966 a, b) are used with different detailed assumptions by a number of authors. Slotboom's own version (Slotboom 1977) is shown in curve IV, and those of earlier authors, van Overstraeten *et al.* (1973) and Mock (1973) are compared in the experimental paper of Slotboom and de Graaf. Slotboom's version is in fair agreement as, by chance presumably, is Inkson's many-body theory expression for Boltzmann statistics, $-\Delta E_{g(\mathrm{Boltz})}{}^{\mathrm{TF}}$, shown in curve II. As degeneracy is approached, Inkson's original expression, $-\Delta E_g{}^{\mathrm{TF}}$, shown in curve I, is in agreement and the improved version $-\Delta E_g{}^{\mathrm{L}}$, in curve V, too low.

8.7. *Conclusions*

(1) For concentrations of about 10^{20} cm^{-3}, the screened exchange contribution to the many-body gap change gives a smaller shift than is observed. In *n*-type silicon, Inkson's theory gives too large an effect in the Thomas–Fermi approximation, but experiment is in agreement with the Lindhard modification suggested here. In *p*-type silicon both formulations give a shift larger than observed in absorption.

(2) For concentrations below 10^{19} cm^{-3}, an impurity band exists and, despite inadequacies in the theories for the bands and tail states, particularly the neglect of impurity wavefunction overlap and of the rigid band shift, some formulations agree well with experiment for *p*-type silicon.

(3) Results for donor impurity concentrations of about 10^{21} cm^{-3} cannot be interpreted by existing theory as precipitation is known to occur. Results are not in good agreement, but all show a substantial shift.

Acknowledgments

We are grateful to Professor Sir Nevill Mott for the initial encouragement to write this review and for many discussions during the course of the work. It is also a pleasure to acknowledge the contributions of M. J. Adams, B. T. Debney, R. C. Goodfellow, J. C. Inkson, P. T. Landsberg and M. Pepper, with whom various aspects of the article have been discussed. We thank the authors who have communicated results before publication, the authors and publishers who have allowed us to reproduce figures, and the Directors of Plessey Research (Caswell) Ltd. for permission to publish this work.

APPENDIX

Conversion of equations to SI units

The equations of this review are written in c.g.s. units, with the electrical quantities in e.s.u. unless otherwise stated. Some equations, such as the bracketed version of eqn. (2.2), are written in a reduced form and expressed explicitly in terms of a unit in conventional use. Equations of this form can be converted to SI units by inspection. The remaining numbered equations can be converted as follows.

Sections 1–6

(i) In most cases an equation remains unchanged in SI units unless the electronic charge e appears, when the change is effected by

$$e \rightarrow e/(4\pi\epsilon_0)^{1/2},$$

where ϵ_0 is the permittivity of free space.

(ii) The exceptions to (i) are eqn. (5.5), where 4π should be replaced by $1/\epsilon_0$, and eqns. (6.1), (6.2), (6.4) and (6.6), where there is no change.

Sections 7–8

There are no changes in the equations in these sections.

References

Abrahams, M. S., Buiocchi, C. J., and William, B. F., 1971, *Appl. Phys. Lett.*, **18**, 220.

Abram, R. A., and Edwards, S. F., 1972 a, *J. Phys.* C, **5**, 1183 ; 1972 b, *Ibid.*, **5**, 1196.

Ackett, G. A., Nijman, W., and 't Lam, H., 1974, *J. appl. Phys.*, **45**, 3033.

Adams, M. J., 1969, *Solid-St. Electron.*, **12**, 661 ; 1973, *Optoelectronics*, **5**, 201.

Adams, M. J., and Cross, M., 1971, *Solid-St. Electron.*, **14**, 865.

Adams, M. J., Gründorfer, S., Thomas, B., Davies, C. F. L., and Mistry, D., 1973, *I.E.E.E. Jl quant. Electron.*, **9**, 328.

Adler, S. L., 1962, *Phys. Rev.*, **126**, 413.

Alexander, M. N., and Holcomb, D. F., 1968, *Rev. mod. Phys.*, **40**, 815.

Alferov, Zh. I., Andreev, W. M., Garbuzov, D. Z., and Trikam, M. K., 1973, *Soviet Phys. Semicond.*, **6**, 1718.

Ambegaokar, V., Halperin, B. I., and Langer, J. S., 1971, *Phys. Rev.* B, **4**, 2612.

Anderson, P. W., 1958, *Phys. Rev.*, **109**, 1492 ; 1972, *Proc. natn. Acad. Sci. U.S.A.*, **69**, 1097.

Antončik, E., 1967, *Czech. J. Phys.* B, **17**, 953.

Archer, R. J., Leite, R. C. C., Yariv, A., Porto, S. P. S., and Whelan, J. M., 1963, *Phys. Rev. Lett.*, **10**, 483.

Aspnes, D. E., and Bottka, N., 1974, *Semiconds Semimetals*, **9**, 457.

Bakker, J., and Ackett, G. A., 1977, *I.E.E.E. Jl quant. Electron.*, **13**, 567.

Baldereschi, A., and Lipari, N. O., 1973, *Phys. Rev.* B, **8**, 2697 ; 1976, *Proc. 13th Int. Conf. Phys. Semicond.*, edited by F. G. Fumi (Rome : Tipografia Marves), p. 595.

Balkanski, M., Aziza, A., and Amzallag, E., 1969, *Phys. Stat. Sol.*, **31**, 323.

Beck, J. D., and Conradt, R., 1973, *Solid St. Commun.*, **13**, 93.

Berggren, K. F., 1973, *Phil. Mag.*, **27**, 1027.

Bergh, A. A., and Dean, P. J., 1976, *Light Emitting Diodes* (Oxford : Clarendon Press), Chap. 3.

Baird, J. R., Carr, W. N., and Reed, B. S., 1964, *Trans. AIME*, **230**, 286.

Black, J. F., Lanning, E., and Perkowitz, S., 1970, *Infrared Phys.*, **10**, 125.
Blatt, V., Walsh, P. S., and Kennedy, L. W., 1975, *I.E.E.E. Jl Solid-St Circuits*, **10**, 336.
Bonch Bruevich, V. L., 1963, *Soviet Phys. solid St.*, **4**, 1953; 1966 a, *Semiconds Semimetals*, **1**, 101; 1966 b, *The Electronic Theory of Heavily Doped Semiconductors* (Elsevier).
Bonch Bruevich, V. L., and Mironov, A. G., 1962, *Soviet Phys. solid St.*, **3**, 2194.
Bowers, R., 1957, *Phys. Rev.*, **108**, 683.
Brinkman, W., and Goodman, B., 1966, *Phys. Rev.*, **149**, 597.
Brinkman, W., and Lee, P. A., 1973, *Phys. Rev. Lett.*, **31**, 237.
Brinkman, W., and Rice, T. M., 1973, *Phys. Rev.* B, **7**, 1508.
Brooks, H., 1955, *Adv. Electronics Electron Phys.*, **7**, 85.
Bryant, C. A., and Keesom, P. H., 1962, *Phys. Rev.*, **124**, 698.
Buhanan, D., 1969, *I.E.E.E. Trans. electron. Devices*, **16**, 117.
Burns, G., Dill, F. H., Jr., and Nathan, M. I., 1963, *Proc. Inst. elect. electron. Engrs*, **51**, 947.
Burns, G., and Nathan, M. I., 1964, *Proc. Inst. elect. electron. Engrs*, **52**, 770.
Burstein, E., 1954, *Phys. Rev.*, **93**, 632.
Callaway, J., 1964, *Energy Band Theory* (Academic Press).
Carrol, J. E., 1974, *Physical Model for Semiconductor Devices* (Edward Arnold), Chaps. 5, 9.
Casey, H. C., and Kaiser, R. H., 1967, *J. electrochem. Soc.*, **114**, 149.
Casey, H. C., Sell, D. D., and Wecht, K. W., 1975, *J. appl. Phys.*, **46**, 250.
Casey, H. C., and Silversmith, D. J., 1969, *J. appl. Phys.*, **40**, 241.
Casey, H. C., and Stern, F., 1976, *J. appl. Phys.*, **47**, 631.
Cheroff, G., Stern, F., and Treibwasser, S., 1963, *Appl. Phys. Lett.*, **2**, 173.
Chinone, N., 1977, *J. appl. Phys.*, **48**, 3237.
Combescot, M., and Nozières, P., 1972, *J. Phys.* C, **5**, 2369.
Conwell, E. M., 1967, *Solid St. Phys.*, Suppl., **9**, Chap. 5.
Conwell, E. M., and Weisskopf, V. F., 1950, *Phys. Rev.*, **77**, 388.
Cook, D. D., and Nash, F. R., 1975, *J. appl. Phys.*, **46**, 1660.
Cross, M., 1973, *Phys. Stat. Sol.* (a), **16**, 167.
Cross, M., and Adams, M. J., 1972, *Solid-St. Electron.*, **15**, 919; 1974, *Optoelectronics*, **6**, 199.
Cuevas, M., and Fritzsche, H., 1965, *Phys. Rev.*, **139**, A1628.
Cusano, D. A., 1964, *Solid St. Commun.*, **2**, 353.
Debney, B. T., 1976, Ph.D. Thesis, University of Birmingham; 1977, *J. Phys.* C, **10**, 4719.
de Graaf, H. C., Soltboom, J. W., and Schmitz, A., 1977, *Solid-St. Electron.*, **20**, 515.
de Man, H. J., 1971, *I.E.E.E. Trans. electron. Devices*, **18**, 833.
Dingle, R. B., 1955, *Phil. Mag.*, **46**, 831.
Dixon, R. W., Nash, F. R., Hartman, R. L., and Hepplewhite, R. T., 1976, *Appl. Phys. Lett.*, **29**, 372.
Dousmanis, G. C., Nelson, H., and Staebler, D. L., 1964, *Appl. Phys. Lett.*, **5**, 174.
Dumke, W. P., 1963, *Phys. Rev.*, **132**, 1998.
Dyment, J. C., Ripper, J. E., and Zachos, T. H., 1969, *J. appl. Phys.*, **40**, 1802.
Eagles, D. M., 1960, *J. Phys. Chem. Solids*, **16**, 76.
Economou, E. N., and Antoniou, P. D., 1977, *Solid St. Commun.*, **21**, 285.
Edwards, S. F., 1958, *Phil. Mag.*, **3**, 1020; 1978, *J. non-crystalline Solids* (to be published).
Ehrenreich, H., 1963, *Phys. Rev.*, **120**, 1951.
Elliot, R. J., 1957, *Phys. Rev.*, **108**, 1384.
Fair, R. B., 1973, *I.E.E.E. Trans. electron. Devices*, **20**, 642.
Falicov, L. M., and Cuevas, M., 1967, *Phys. Rev.*, **164**, 1025.
Fan, H. Y., 1967, *Semiconds Semimetals*, **3**, 405.
Faulkner, R. A., 1968, *Phys. Rev.*, **175**, 991; 1969, *Ibid.*, **184**, 713.
Fistul, V. I., and Agaev, A. M., 1966, *Soviet Phys. solid St.*, **7**, 2461.
Frederikse, H. P. R., and Hosler, W. R., 1957, *Phys. Rev.*, **108**, 1136.
Freed, K., 1972, *Phys. Rev.* B, **5**, 4802.

FRITZSCHE, H., and CUEVAS, M., 1962, *Proc. Int. Conf. Phys. Semicond.*, Exeter, edited by A. C. Strickland (IPPS), p. 29.
GASPARD, J. P., and CYROT LACKMANN, F., 1973, *J. Phys.* C, **6**, 3077 ; 1974, *Ibid.*, **7**, 1829.
GOEBEL, E. O., HILDEBRAND, O., and LÖHNERT, K., 1977, *I.E.E.E. Jl quant. Electron.*, **13**, 848.
GOOCH, C. H., 1973, *Injection Electroluminescent Devices* (Wiley).
GOODWIN, A. R., 1969, Solid State Device Research Conference, Rochester (unpublished).
GOODWIN, A. R., and SELWAY, P. R., 1970, *I.E.E.E. Jl quant. Electron.*, **6**, 285.
GOODWIN, A. R., and THOMPSON, G. H. B., 1970, *I.E.E.E. Jl quant. Electron.*, **6**, 311.
GROVE, A. S., 1967, *Physics and Technology of Semiconductor Devices* (Wiley).
GRÜNDORFER, S., and ADAMS, M. J., 1973, *I.E.E.E. Jl quant. Electron.*, **9**, 814 ; 1974, *Electronics Lett.*, **10**, 354.
GRÜNDORFER, S., ADAMS, M. J., and THOMAS, B., 1975, *I.E.E.E. Jl quant. Electron.*, **11**, 532.
HAAS, C., 1962, *Phys. Rev.*, **125**, 1965.
HAKKI, B. W., 1973, *J. appl. Phys.*, **44**, 5021 ; 1974, *Ibid.*, **45**, 288 ; 1975, *Ibid.*, **46**, 2723.
HAKKI, B. W., and PAOLI, T. L., 1973, *J. appl. Phys.*, **44**, 4113 ; 1975, *Ibid.*, **46**, 1299.
HALL, R. N., 1963, *Solid-St. Electron.*, **6**, 405.
HALPERIN, B. I., and LAX, M., 1966, *Phys. Rev.*, **148**, 722.
HATZ, J., and MOHN, E., 1967, *I.E.E.E. Jl quant. Electron.*, **3**, 656.
HAYASHI, I., and PANISH, M. B., 1970, *J. appl. Phys.*, **41**, 150.
HAYASHI, I., PANISH, M. B., and REINHART, F. K., 1971, *J. appl. Phys.*, **42**, 1929.
HEDIN, L., and LUNDQVIST, S., 1969, *Solid St. Phys.*, **23**, 1.
HENSEL, J. C., PHILLIPS, T. G., and THOMAS, G. A., 1977, *Solid St. Phys.*, **32**, 88.
HERBERT, D. C., HURLE, D. T. J., and LOGAN, R. M., 1975, *J. Phys.* C, **8**, 3571.
HILL, D., 1976, *Proc. R. Soc.* A, **347**, 565.
HILL, D., and LANDSBERG, P. T., 1976, *Proc. R. Soc.* A, **347**, 547.
HILL, D. E., 1964, *Phys. Rev.*, **133**, A866.
HWANG, C. J., 1970 a, *J. appl. Phys.*, **41**, 2668 ; 1970 b, *Phys. Rev.* B, **2**, 4117 ; 1970 c, *Ibid.*, **2**, 4126.
HWANG, C. J., and BREWS, J. R., 1971, *J. Phys. Chem. Solids*, **32**, 837.
HWANG, C. J., and DYMENT, J. C., 1973, *J. appl. Phys.*, **44**, 3240.
HWANG, C. J., PATEL, N. B., SACILOTTI, M. A., PRINCE, F. C., and BULL, D. J., 1978, *J. appl. Phys.*, **49**, 29.
INKSON, J. C., 1973, *J. Phys.* C, **6**, 1350 ; 1976, *Ibid.*, **9**, 1177.
JACCODINE, R. J., 1968, *J. appl. Phys.*, **39**, 3105.
JANNUZZI, N., MENESES, E. A., and LEITE, R. C. C., 1972, *Solid St. Commun.*, **10**, 517.
JONSCHER, A. K., and BOYLE, M. H., 1966, *Proceedings of the IPPS Symposium on GaAs*, Reading (IPPS), p. 78.
KANE, E. O., 1963, *Phys. Rev.*, **131**, 79 ; 1966, *Semiconds Semimetals*, **1**, 75 ; 1967, *Phys. Rev.*, **159**, 624 ; 1972, *Phys. Rev.* B, **5**, 1493.
KANNAM, P. J., 1973, *I.E.E.E. Trans. electron. Devices*, **20**, 845.
KATZ, M. J., 1965, *Phys. Rev.*, **140**, 1323.
KATZ, M. J., KOENIG, S. H., and LOPEZ, A. A., 1965, *Phys. Rev. Lett.*, **15**, 828.
KAUFFMAN, W. L., and BERGH, A. A., 1968, *I.E.E.E. Trans. electron. Devices*, **15**, 732.
KAWAKAMI, T., and SUGIYAMA, K., 1973, *Jap. J. appl. Phys.*, **12**, 151.
KEESOM, P. H., and SEIDEL, G. M., 1959, *Phys. Rev.*, **113**, 33.
KELDYSH, L. V., and PROSHKO, G. P., 1964, *Soviet Phys. solid St.*, **5**, 2481.
KIRKBY, P. A., GOODWIN, A. R., THOMPSON, G. H. B., and SELWAY, P. R., 1977, *I.E.E.E. Jl quant. Electron.*, **13**, 705.
KIRKBY, P. A., and THOMPSON, G. H. B., 1972, *Optoelectronics*, **4**, 323.
KLEPPINGER, D. D., and LINDHOLM, F. A., 1971, *Solid-St. Electron.*, **14**, 407.
KOBAYASHI, N., IKEHATA, S., KOBAYASHI, S., and SASAKI, W., 1977, *Solid St. Commun.*, **24**, 67.
KOHN, W., 1957, *Solid St. Phys.*, **5**, 257.
KONNERTH, K., and LANZA, C., 1964, *Appl. Phys. Lett.*, **4**, 120.

KRIEGER, J. B., GRUENEBAUM, J., and MEEKS, T., 1974, *Phys. Rev.* B, **9**, 3627.
KRIEGER, J. B., and MEEKS, T., 1973, *Phys. Rev.* B, **8**, 2780.
KRIEGER, J. B., and STRAUSS, S., 1968, *Phys. Rev.*, **169**, 674.
KUDMAN, I., and SEIDEL, T., 1962, *J. appl. Phys.*, **33**, 771.
LAMORTE, M. F., and GONDA, T., 1966, *I.E.E.E. Jl quant. Electron.*, **2**, 9.
LANDSBERG, P. T., and ROBBINS, D. J., 1977, *J. Phys.* C, **10**, 2717.
LASHER, G., and STERN, F., 1964, *Phys. Rev.*, **133**, A553.
LAST, B. J., and THOULESS, D. J., 1974, *J. Phys.* C, **7**, 699.
LEITE, R. C. C., SARACE, J. C., OLSON, D. H., COHEN, B. G., and WHELAN, J. M., 1965, *Phys. Rev.*, **137**, A1583.
LICCIARDELLO, D. C., and THOULESS, D. J., 1975, *J. Phys.* C, **8**, 4157 ; 1976, *Ibid.*, **9**, L417.
LINDHOLM, F. A., NEUGROSCHEL, A., SAH, C. T., GODLEWSKI, M. P., and BRANDHORST, H. W., 1977, *I.E.E.E. Trans. electron. Devices*, **24**, 402.
LLOYD, P., and BEST, P. R., 1975, *J. Phys.* C, **8**, 3752.
LOCKWOOD, H. F., KRESSEL, H., SOMMERS, H. S., JR., and HAWRYLO, F. Z., 1970, *Appl. Phys. Lett.*, **17**, 499.
LOGAN, R. A., and CHYNOWETH, G., 1963, *Phys. Rev.*, **131**, 89.
LUKES, T., 1972, *J. non-crystalline Solids*, **8–10**, 470.
MAHAN, G. D., 1972, *Polarons in Ionic Crystals and Polar Semiconductors*, edited by J. F. Devreese (North Holland).
MAHAN, G. D., and CONLEY, J. W., 1967, *Appl. Phys. Lett.*, **11**, 29.
MARPLES, D. T. F., 1964, *J. appl. Phys.*, **35**, 1241.
MARSHAK, A. H., and VAN VLIET, K. M., 1978 a, *Solid-St. Electron.*, **21**, 417 ; 1978 b, *Ibid.*, **21**, 429.
MATSUBARA, T., and TOYOZAWA, Y., 1961, *Prog. Theor. Phys., Osaka*, **26**, 739.
MATTHEWS, M. R., DYOTT, R. B., and CARLING, W. P., 1972, *Electronics Lett.*, **8**, 570.
MAVROIDES, J. G., 1972, *Optical Properties of Solids*, edited by F. Abeles (North Holland).
MCWHORTER, A. L., ZEIGER, H. J., and LAX, B., 1963, *J. appl. Phys.*, **34**, 235.
MEEKS, T., and KRIEGER, J. B., 1969, *Phys. Rev.*, **185**, 1068.
MERTENS, R. P., DE MAN, H. J., VAN OVERSTRAETEN, R. J., 1973, *I.E.E.E. Trans. electron. Devices*, **20**, 772.
MILLER, A., and ABRAHAMS, E., 1960, *Phys. Rev.*, **120**, 745.
MILLER, B. I., RIPPER, J. E., DYMENT, J. C., PINKAS, E., and PANISH, M. B., 1971, *Appl. Phys. Lett.*, **18**, 403.
MOCK, M. S., 1973, *Solid-St. Electron.*, **16**, 1251.
MOORE, E. J., 1967 a, *Phys. Rev.*, **160**, 607 ; 1967 b, *Ibid.*, **160**, 618.
MORGAN, T. N., 1963, *Bull. Am. phys. Soc.*, **8**, 224 ; 1965, *Phys. Rev.*, **139**, A343.
MOSS, T. S., 1954, *Proc. phys. Soc.* B, **67**, 775.
MOSS, T. S., BURRELL, G. J., and ELLIS, B., 1973, *Semiconductor Opto-Electronics* (Butterworths).
MOSS, T. S., BURRELL, G. J., and HETHERINGTON, A., 1969, *Solid-St. Electron.*, **12**, 787.
MOTT, N. F., 1969, *Phil. Mag.*, **19**, 835 ; 1970, *Ibid.*, **22**, 7 ; 1972, *Adv. Phys.*, **21**, 785 ; 1973 a, *Proceedings of the 13th Scottish Summer School in Physics*, edited by P. G. Le Comber and J. Mort (Academic Press), p. 1 ; 1973 b, *Electron Power*, **19**, 321 ; 1974, *Metal–Insulator Transitions* (Taylor & Francis Ltd) ; 1976 a, *Commun. Phys.*, **1**, 203 ; 1976 b, *Phil. Mag.*, **34**, 643.
MOTT, N. F., and DAVIS, E. A., 1971, *Electronic Processes in Non-Crystalline Materials* (Oxford : Clarendon Press).
MOTT, N. F., and JONES, H., 1936, *The Theory of the Properties of Metals and Alloys* (Oxford : Clarendon Press), p. 137 *et seq.*
MOTT, N. F., PEPPER, M., POLLITT, S., WALLIS, R. H., and ADKINS, C. J., 1975, *Proc. R. Soc.* A, **345**, 169.
MOTT, N. F., and TWOSE, W. D., 1961, *Adv. Phys.*, **10**, 107.
NAMIZAKI, H., KAN, H., ISHII, M., and ITO, A., 1974, *Appl. Phys. Lett.*, **24**, 486.
NASH, F. R., 1973, *J. appl. Phys.*, **44**, 4696.
NATHAN, M. I., BURNS, G., BLUM, S. E., and MARINACE, J. C., 1963 a, *Phys. Rev.*, **132**, 1482.

NATHAN, M. I., FOWLER, A. B., and BURNS, G., 1963 b, *Phys. Rev. Lett.*, **11**, 152.
NELSON, D. F., GERSHENZON, M., ASHKIN, A., D'ASARO, L. A., and SARACE, J. C., 1963, *Appl. Phys. Lett.*, **2**, 182.
NISHIMURA, Y., 1974, *Jap. J. appl. Phys.*, **13**, 109.
NORTON, P., 1976 a, *J. appl. Phys.*, **47**, 308 ; 1976 b, *Proc. 13th Int. Conf. Phys. Semicond.*, edited by F. G. Fumi (Rome : Tipografia Marves), p. 603.
NUNES, F. D., PATEL, N. B., and RIPPER, J. E., 1976, *Electronics Lett.*, **12**, 574 ; 1977, *I.E.E.E. Jl quant. Electron.*, **13**, 675.
PALIK, E. D., and WRIGHT, G. B., 1967, *Semiconds Semimetals*, **3**, 421.
PANISH, M. B., 1976, *Proc. Inst. elect. electron. Engrs*, **64**, 1512.
PANKOVE, J. I., 1962, *Phys. Rev. Lett.*, **9**, 283 ; 1965, *Phys. Rev.*, **140**, A2059 ; 1968 a, *J. appl. Phys.*, **39**, 5368 ; 1968 b, *I.E.E.E. Jl quant. Electron.*, **4**, 427.
PANKOVE, J. I., and AIGRAIN, P., 1962, *Phys. Rev.*, **126**, 956.
PANTELIDES, S. T., 1974, *Solid St. Commun.*, **14**, 1255.
PANTELIDES, S. T., and SAH, C. T., 1974, *Phys. Rev.* B, **10**, 638.
PAOLI, T. L., 1977, *I.E.E.E. Jl quant. Electron.*, **13**, 662.
PAOLI, T. L., HAKKI, B. W., and MILLER, B. I., 1973, *J. appl. Phys.*, **44**, 1276.
PARMENTER, R. H., 1955, *Phys. Rev.*, **97**, 587.
PEPPER, M., 1976, *Commun. Phys.*, **1**, 147 ; 1977, *J. Phys.* C, **10**, L173.
PILKUHN, M., RUPPRECHT, H., and BLUM, S., 1964, *Solid-St. Electron.*, **7**, 905.
PILLER, H., 1972, *Semiconds Semimetals*, **8**, 103.
PINES, D., 1963, *Elementary Excitation in Solids* (Benjamin).
PINKAS, E., MILLER, B. I., HYASHI, I., and FOY, P. W., 1972, *J. appl. Phys.*, **43**, 2827.
POLLAK, M., 1972, *J. non-crystalline Solids*, **11**, 1.
PRICE, P. J., and KRIEGER, J. B., 1966, *Bull. Am. phys. Soc.*, **11**, 400.
PUTLEY, E. H., and MITCHELL, W. H., 1958, *Proc. phys. Soc.* A, **72**, 193.
RAYMOND, A., ROBERT, J. L., and PISTOULET, B., 1977, *Inst. Phys. Conf.*, Series No. 33a, p. 105.
REDFIELD, D., 1975, *Adv. Phys.*, **24**, 463.
REDFIELD, D., WITTKE, J. P., and PANKOVE, J. I., 1970, *Phys. Rev.* B, **2**, 1830.
REES, G. J., and MILNE, W. I., 1977, *Inst. Phys. Conf.*, Series No. 33a, p. 362.
RENTZSCH, R., and SHLIMAK, I. S., 1977, *Phys. Stat. Sol.* (a), **43**, 231.
RICE, T. M., 1977, *Solid St. Phys.*, **32**, 1.
RIMBEY, P. R., and MAHAN, G. D., 1974, *Phys. Rev.* B, **10**, 3419.
RODE, D. L., 1971, *Phys. Rev.* B, **3**, 3287 ; 1975, *Semiconds Semimetals*, **10**, 1.
RODE, D. L., and KNIGHT, S., 1971, *Phys. Rev.* B, **3**, 2534.
ROGACHEV, A. A., and RYVKIN, S. M., 1966, *Soviet Phys. solid St.*, **7**, 2690.
ROGACHEV, A. A., and SABLINA, N. I., 1966, *Soviet Phys. solid St.*, **8**, 691.
ROLDAN, R., 1967, *Appl. Phys. Lett.*, **11**, 346.
SANDHU, J. S., and RENTER, J. L., 1971, *IBM Jl Res. Dev.*, **15**, 464.
SASAKI, W., and KINOSHITA, J., 1968, *J. phys. Soc. Japan*, **25**, 1622.
SA-YAKANIT, V., 1978, *Phys. Rev.* B (to be published).
SCHADE, H., NELSON, H., and KRESSEL, H., 1971, *Appl. Phys. Lett.*, **18**, 121.
SCHIFF, L. I., 1968, *Quantum Mechanics* (McGraw-Hill), p. 324.
SCHLOSSER, W. O., 1973, *Bell Syst. tech. J.*, **52**, 887.
SCHUMANN, P. A., 1970, *Solid St. Technol.*, **13**, 50.
SELL, D. D., CASEY, H. C., and WECHT, K. W., 1974, *J. appl. Phys.*, **45**, 2650.
SHACKLEE, K. L., LEHENY, R. F., and NAHORY, R. E., 1971, *Appl. Phys. Lett.*, **19**, 302.
SHAH, J., LEHENY, R. F., and WEIGMANN, W., 1977, *Phys. Rev.* B, **16**, 1577.
SHEPHERD, F. D., VICKERS, V. E., and YANG, A. C., 1965, *Proc. Inst. elect. electron. Engrs*, **53**, 1759.
SHKLOVSKII, B. I., and ÉFROS, A. L., 1971, *Soviet Phys. JETP*, **33**, 468.
SLADEK, R. J., 1958, *Phys. Rev.*, **110**, 817.
SLOTBOOM, J. W., 1977, *Solid-St. Electron.*, **20**, 279.
SLOTBOOM, J. W., and DE GRAAF, H. C., 1976, *Solid-St. Electron.*, **19**, 857.
SMITH, R. A., 1964, *Semiconductors* (Cambridge University Press), Chap. 8.
SPITZER, W., and FAN, H. Y., 1957, *Phys. Rev.*, **106**, 882.

STERN, F., 1963, *Solid St. Phys.*, **15**, 299 ; 1964, *Phys. Rev.*, **133**, A1653 ; 1966, *Ibid.*, **148**, 186 ; 1971 a, *Phys. Rev.* B, **3**, 2636 ; 1971 b, *Ibid.*, **3**, 3559 ; 1972, *Semiconductor Lasers, Theory, Laser Handbook*, edited by F. T. Arecchi and E. O. Shulz-DuBois (Amsterdam : North-Holland) ; 1973, *I.E.E.E. Jl quant. Electron.*, **9**, 290 ; 1976, *J. appl. Phys.*, **47**, 5382.
SZE, S. M., 1969, *Physics of Semiconductor Devices* (Wiley–Interscience), Chaps. 6, 13.
TANNENBAUM, M., and THOMAS, D. E., 1956, *Bell Syst. tech. J.*, **35**, 1.
THOMAS, B., ADAMS, M. J., and GRÜNDORFER, S., 1975, *I.E.E.E. Jl quant. Electron.*, **11**, 528.
THOMAS, B., MISTRY, D., and DAVIES, C. F. L., 1974, *I.E.E.E. Jl quant. Electron.*, **10**, 401.
THOMPSON, G. H. B., 1972, *Optoelectronics*, **4**, 257.
THOMPSON, G. H. B., HENSHALL, G. D., WHITEAWAY, J. E. A., and KIRKBY, P. A., 1976, *J. appl. Phys.*, **47**, 1501.
TUCK, B., 1968, *J. Phys. Chem. Solids*, **29**, 615.
TURNER, W. J., and REESE, W. E., 1964, *J. appl. Phys.*, **35**, 350.
UE, H., and MAEKAWA, 1971, *Phys. Rev.* B, **3**, 4232.
UNGER, K., 1967 a, *Z. Phys.*, **207**, 322 ; 1967 b, *Ibid.*, **207**, 332.
VAN OVERSTRAETEN, R. J., DE MAN, H. J., and MERTENS, R. P., 1973, *I.E.E.E. Trans. electron. Devices*, **20**, 290.
VILMS, J., and SPICER, W. E., 1965, *J. appl. Phys.*, **36**, 2815.
VOL'FSON, A. A., and SUBASHIEV, V. K., 1967, *Soviet Phys. Semicond.*, **1**, 327.
VOLKOV, B. A., and MATVEEV, V. V., 1966, *Soviet Phys. solid St.*, **8**, 574.
VUL, B. M., ZAVARITSKAYA, E. I., KOTEL'NIKOVA, N. V., and VORONOVA, I. D., 1976, *Soviet Phys. Semicond.*, **10**, 751.
WATKINS, G. D., and CORBETT, J. W., 1964, *Phys. Rev.*, **134**, A1359.
WEAIRE, D., and SRIVASTAVA, V., 1977, *Solid St. Commun.*, **23**, 863.
WILSON, B. L. H., 1977, *Solid-St. Electron.*, **20**, 71.
WOLFF, P. A., 1962, *Phys. Rev.*, **126**, 405.
WOLFORD, D. J., STREETMAN, B. G., NELSON, R. J., and HOLONYAK, N., JR., 1976, *Appl. Phys. Lett.*, **28**, 711.
YAGI, H., SUYUKI, T. T., KOMA, K., and MIZAWA, Y., 1974, Conference abstracts, I.E.E.E. Electronic Device Meeting, Washington, p. 262.
YAMANOUCHI, C., MIZUGUCHI, K., and SASAKI, W., 1967, *J. phys. Soc. Japan*, **22**, 859.
YARIV, A., and LEITE, R. C. C., 1963, *Appl. Phys. Lett.*, **2**, 55.
YOSHINO, S., and OKAZAKI, M., 1977, *J. phys. Soc. Japan*, **43**, 415.
ZACHOS, T. H., and RIPPER, J. E., 1969, *I.E.E.E. Jl quant. Electron.*, **5**, 29.
ZIMAN, J. M., 1960, *Electrons and Phonons* (Oxford University Press), Chaps. 7, 10 ; 1961, *Phil. Mag.*, **6**, 1013 ; 1972, *Principles of the Theory of Solids* (Cambridge University Press).
ZOROOFCHI, J., and BUTLER, J. K., 1973, *J. appl. Phys.*, **44**, 3697.
ZSCHAUER, K. H., 1969, *Solid St. Commun.*, **7**, 1709.
ZUCKER, J., LAUER, R. B., and SCHLAFER, J., 1976, *J. appl. Phys.*, **47**, 2082.
ZVEREV, L. P., NEGASHEV, S. A., KRUZHAEV, V. V., and MIN'KOV, G. M., 1977, *Soviet Phys. Semicond.*, **11**, 603.
ZVYAGIN, I. P., 1963, *Soviet Phys. solid St.*, **5**, 422.