Solid-State Electronics Vol. 22, pp. 783–791
© Pergamon Press Ltd., 1979. Printed in Great Britain

A PROPOSAL FOR A NEW APPROACH TO HETEROJUNCTION THEORY

M. J. ADAMS
Department of Electronics, University of Southampton, Southampton SO9 5NH, England

and

ALLEN NUSSBAUM
Department of Electrical Engineering, University of Minnesota, Minneapolis, MN 55455, U.S.A.

(*Received* 24 *March* 1977; *in revised form* 15 *January* 1979)

Abstract—The Anderson theory of heterojunctions is reviewed with particular reference to the derivation of the affinity rule. We consider in detail the fundamental assumptions on which the Anderson model is based, analyzing arguments which either confirm or dispute these underlying ideas. An alternate model is then proposed, and again we discuss the arguments for and against. Our conclusion is that further work must be done both theoretically and experimentally before a definite determination can be made as to the more appropriate of these two theories.

1. INTRODUCTION

Until recent years the heterojunction occupied a position of relative obscurity in the hierarchy of semiconductor topics of current interest. However, with the development of heterostructure injection lasers[1] for applications in optical communications and elsewhere, the semiconductor heterojunction has now achieved a sphere of much wider interest (for a comprehensive review, see e.g. Ref. [2]). In the case of the injection laser, the use of a double heterojunction structure has made possible the achievement of room-temperature c.w. operation by confining the region of radiative recombination to a well-defined, narrow layer of material. The resulting structure is, however, a difficult one to analyse from the viewpoint of a detailed theoretical understanding of the carrier transport effects. A recent study[3] has revealed several areas of uncertainty in modelling such a structure, most of which are due to fundamental questions in the physics of heterojunctions.

The earliest theoretical model for heterojunction behaviour, due to Anderson[4], utilised the Schottky depletion approximation to analyse an abrupt junction between two semiconductors. An alternative model allowing for a graded transition of energy gaps rather than an abrupt interface was proposed by Oldham and Milnes[5]. Other models have included the effects of interface states[6] and of minority carriers[7] in the region of the heterojunction. More recently the fundamental question of how the energy band discontinuity is distributed between conduction and valence bands has been resurrected[8, 9].

It is this latter problem with which we are concerned here. In particular the conventional assumption regarding the division of energy discontinuities at an interface—the electron affinity rule—is examined with reference to recent criticisms[8]. An alternative proposal is made for calculating energy band discontinuities by assuming continuity of the intrinsic level rather than the vacuum level. Some results of applying this rule to common heterojunctions are given and compared with the results of the affinity rule.

We consider throughout this paper the anisotype (p–n) heterojunction, specifically with n-type energy gap (E_{GN}) larger than p-type (E_{GP}). However, the arguments presented are perfectly general and may equally well be applied to the anisotype case with $E_{GP} > E_{GN}$ and to the isotype n-n or p-p situations.

2. PERTINENT PROPERTIES OF HOMOJUNCTIONS

In order to show the basis of our proposed modification of the Anderson model, we briefly review the pertinent concepts of homojunction physics. Firstly, for a typical N-type homogeneous semiconductor (Fig. 1), the Fermi level E_F will lie a distance K above the intrinsic level E_I, where K, being a measure of the shift of the Fermi function from its symmetric position in intrinsic material, should logically be identified as the chemical potential. It then follows that the position of E_I with respect to the arbitrary reference level shown should be identified as the electrostatic energy $-eV$. This confirms to the customary definition[10] of the electrochemical potential $\bar{\mu}$, namely

$$E_F = \bar{\mu} = K - eV.$$

The figure also shows the vacuum level E_{vac} and the associated work function W and electron affinity χ.

All of these quantities can be applied to the homojunction of Fig. 2 (taken as symmetric for simplicity). To determine the actual shape of the bands in the transition region, we impose the conditions that E_F is everywhere constant and that the displacement vector $\vec{D}$ has a continuous normal component D_n at the junction. These are used in conjunction with the Poisson–Boltzmann equation, which has the dimensionless form[11]

$$\frac{d^2u}{dr^2} = \sinh u - a. \tag{1}$$

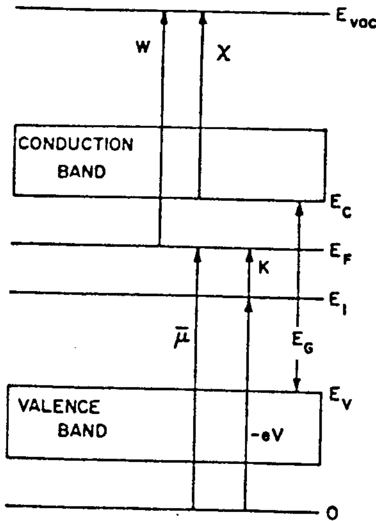

Fig. 1. Energy band structure for n-type semiconductor in equilibrium.

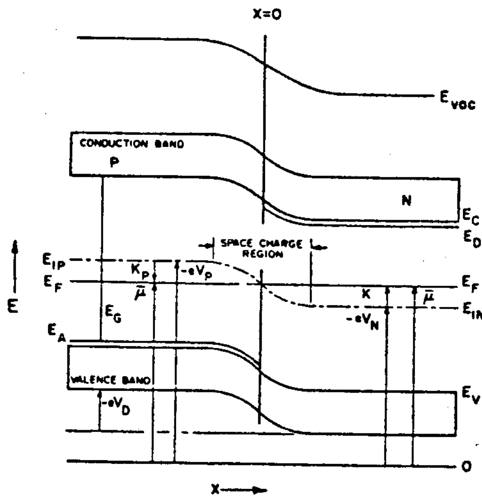

Fig. 2. Energy band structure for symmetrical homojunction.

Equation (1) comes from Poisson's equation by letting the volume charge density be expressed as $p-n+N_D-N_A$, by using Boltzmann statistics in the form $n=n_i\exp(K/kT)$, $p=n_i\exp(-K/kT)$ (in accordance with our discussion above of the significance of the chemical potential), and by using the reduced variables $u=K/kT$, $a=(N_D-N_A)2n_i$, $r=x/L_D$ where $L_D=(\epsilon kT/2n_ie^2)^{1/2}$ is the intrinsic Debye length. The first integral of (1) is

$$\frac{du}{dr}=\surd(2(\cosh u-au+C)) \tag{2}$$

where C is an integration constant. This relation is actually two equations, one for each region of the junction. Applying appropriate subscripts and using the continuity of D_n (so that du/dr is continuous at $r=0$), the value $u(0)$ of u at the junction becomes

$$u(0)=\frac{C_N-C_P}{a_N-a_P}. \tag{3}$$

Knowing $u(0)$, we can integrate eqn (2) from $r=0$ to each edge of the space-charge region, obtaining equilibrium band structures which depend on the impurity concentrations on each side of the junction. The function u is a measure of the position of E_I with respect to E_F; once the behaviour of E_I is known, all other levels are automatically determined. We shall return to this point when we consider heterojunctions, and shall also show how (3) should be modified.

3. THE ANDERSON THEORY OF HETEROJUNCTIONS

The well-known depletion approximation solutions [10] to the Poisson–Boltzmann equation, when applied to a heterojunction, are

$$\begin{aligned}V_N(x)&=-(eN_D/2\epsilon_N)(x-x_N)^2+V_{N0}\\ V_P(x)&=(eN_A/2\epsilon_P)(x+x_P)^2+V_{P0}\end{aligned} \tag{4}$$

so that at the junction $r=0$, we obtain

$$\begin{aligned}V_N(0)-V_{N0}&=V_{DN}=\frac{-e}{\epsilon_N}\frac{N_D}{2}x_N^2\\ V_P(0)-V_{P0}&=V_{DP}=\frac{e}{\epsilon_P}\frac{N_A}{2}x_P^2\end{aligned} \tag{5}$$

where V_{DN} and V_{DP} are diffusion or barrier potentials which represent the band-bending on each side of the junction. Dividing the first relation by the second gives

$$\left|\frac{V_{DN}}{V_{DP}}\right|=\frac{N_D\epsilon_Px_N^2}{N_A\epsilon_Nx_P^2}. \tag{6}$$

To eliminate $(x_N/x_P)^2$ from (6), we differentiate eqn (4) to obtain the normal components of $\vec{D}=\epsilon\vec{E}$ and use these derivatives in the boundary condition

$$\epsilon_N\left(\frac{dV}{dx}\right)_N=\epsilon_P\left(\frac{dV}{dx}\right)_P \tag{7}$$

to obtain

$$\left|\frac{V_{DN}}{V_{DP}}\right|=\frac{N_A\epsilon_P}{N_D\epsilon_N}=\kappa. \tag{8}$$

From a knowledge of κ and some other constants of the junction, it is now possible to construct an energy band diagram in accordance with the Anderson model [4] as follows: the position of E_C and E_V on each side of the junction are located with respect to the uniform Fermi level by the respective impurity concentrations. The known values of the electron affinities set the position of the two vacuum levels, as shown in Fig. 3, and the difference in energies between these two levels is identified as the diffusion potential V_D, where

$$V_D=V_{DN}+V_{DP}. \tag{9}$$

Solving (8) and (9) simultaneously gives the individual components of V_D and these magnitudes are then transferred down to all other levels: E_C, E_I and E_V. This

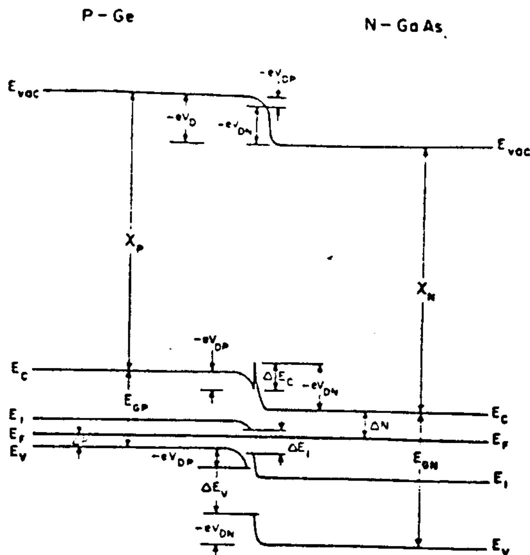

Fig. 3. "Conventional" heterostructure band diagram with continuity of vacuum level at $x = 0$.

crucial feature of the Anderson theory thus implicitly assumes that the boundary condition on D_n is applicable to the vacuum level. Figure 3, based on an example of Milnes and Feucht[2], shows the typical "spike-and-notch" structure appearing in the conduction band, a conventional discontinuity in the valence band, and a somewhat smaller one for the intrinsic level. As Fig. 3 indicates, the quantity Δ_P is the distance from E_V to E_F on the P side and Δ_N is the distance from E_F to E_C on the N side. In terms of these quantities we can write the conduction band discontinuity as

$$\Delta E_C = (E_{GN} + eV_{DN}) - (E_{GN} - \Delta_N - \Delta_P + E_{GP} - eV_{DP}) = \Delta_N + \Delta_P + eV_{DN} + eV_{DP} - E_{GP} \quad (10)$$

where this result is simply a consequence of the conservation of energy, and similarly

$$\Delta E_V = -eV_{DN} + E_{GN} - \Delta_N - \Delta_P - eV_{DP}. \quad (11)$$

But

$$E_{GN} + \chi_N + eV_{DN} + eV_{DP} = E_{GN} - \Delta_N - \Delta_P + E_{GP} + \chi_P$$

or

$$eV_{DN} + eV_{DP} + \Delta_N + \Delta_P = E_{GP} + \chi_P - \chi_N. \quad (12)$$

Using (12) in (10) gives

$$\Delta E_C = \chi_P - \chi_N \quad (13)$$

and in (11)

$$E_V = (E_{GN} - E_{GP}) - (\chi_P - \chi_N). \quad (14)$$

Equation (13) is known as the affinity rule and eqn (14) is the form applicable to the valence band.

4. AN ALTERNATE TREATMENT OF HETEROJUNCTION THEORY

We have two reservations about the Anderson theory. The first involves the use of the depletion approximation. It has been found to give reasonable results in some instances, such as in determining the equilibrium band structure of asymmetric junctions[12]. In other situations, however, its use leads to very unrealistic current-voltage characteristics[11] and each application should be carefully justified. The other difficulty is concerned with the nature of the boundary condition. If, as indicated in connection with Fig. 1, the position of E_I serves as a measure of the electrostatic potential for the carriers in a semiconductor, then the continuity of D_n requires that we apply the boundary condition to du/dr for a heterojunction in a manner like that for the homojunction. That is, the relation

$$\epsilon_P\left(\frac{dV}{dx}\right)_P = \epsilon_N\left(\frac{dV}{dx}\right)_N$$

becomes, in terms of the normalized variables previously used

$$\sqrt{(\epsilon_P n_{iP})}\left(\frac{du}{dr}\right)_P = \sqrt{(\epsilon_N n_{iN})}\left(\frac{du}{dr}\right)_N.$$

The use of (2) at $r = 0$ gives

$$\epsilon_P n_{iP}(\cosh u(0) - a_P u(0) + C_P) = \epsilon_N n_{iN}(\cosh u(0) - a_N u(0) + C_N)$$

where, as in (3), $u(0)$ is the value of u at $r = 0$. Rearranging gives

$$(R-1)\cosh u(0) + (a_N - a_P R)u(0) + (RC_P - C_N) = 0 \quad (15)$$

where

$$R = \frac{\epsilon_P n_{iP}}{\epsilon_N n_{iN}}$$

and where the boundary conditions $du/dr = d^2u/dr^2 = 0$ at the edges of the space-charge region when applied to (1) and (2) permit us to write

$$\left.\begin{aligned} a_i &= \sinh u_i \\ C_i &= u_i \sinh u_i - \cosh u_i \end{aligned}\right\} \quad i = N, P. \quad (16)$$

Turning again to the Milnes and Feucht example[2], the appropriate values turn out to be

$$u_P = -8.66, \quad u_N = 22.44, \quad R = 4.03 \times 10^6.$$

A Newton–Raphson solution of (15) shows that there are two roots, with values $u(0) = -11.075, -2.042$ and this is verified by the plot of eqn (15) over the range of R from 0 to ∞, as shown in Fig. 4. Note that while this graph indicates the presence of two roots for any value of R, only the centre branch of the three-member curve has

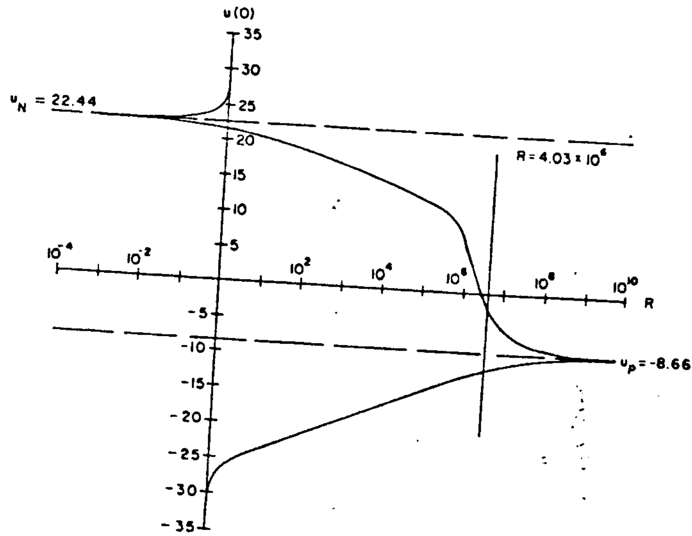

Fig. 4. Solutions to eqn (15) for Milnes–Feucht example.

physical significance. Values of $u(0)$ which lie outside the limits imposed by u_N and u_P would cause the bands to bend away from each other at the junction and violate our original continuity condition. Because of the effort involved in solving transcendental equations of the form (15), it is more efficient to combine it with (16), obtaining

$$R = \frac{\cosh u(0) - \cosh u_N + (u_N - u(0)) \sinh u_N}{\cosh u(0) - \cosh u_P + (u_P - u(0)) \sinh u_P}. \tag{17}$$

This result is plotted for various useful combinations of u_N and u_P, and for principal branches only, in Fig. 5.

The energy band diagram using the root $u(0) = -2.04$ will appear as shown in Fig. 6. Again using conservation of energy, we have that

$$\Delta E_C = -E_{GP} + eV_{DP} + \Delta E_{IP} - eV_D - \Delta E_{IN} + E_{GN} + eV_{DN} \tag{18}$$

where ΔE_{IP} and ΔE_{IN} are the respective heights of the intrinsic levels above the valence band edges. This simplifies to

$$\Delta E_C = (E_{GN} - \Delta E_{IN}) - (E_{GP} - \Delta E_{IP}).$$

From the usual definition[10] of the intrinsic level, namely

$$E_I - E_V = \Delta E_I = \frac{E_G}{2} + \frac{3kT}{4} \log\left(\frac{\mu_e}{\mu_h}\right)$$

we obtain

$$\Delta E_C = \frac{E_{GN} - E_{GP}}{2} - \frac{3kT}{4} \log(c_N/c_P) \tag{19}$$

where

$$c = \mu_e/\mu_h.$$

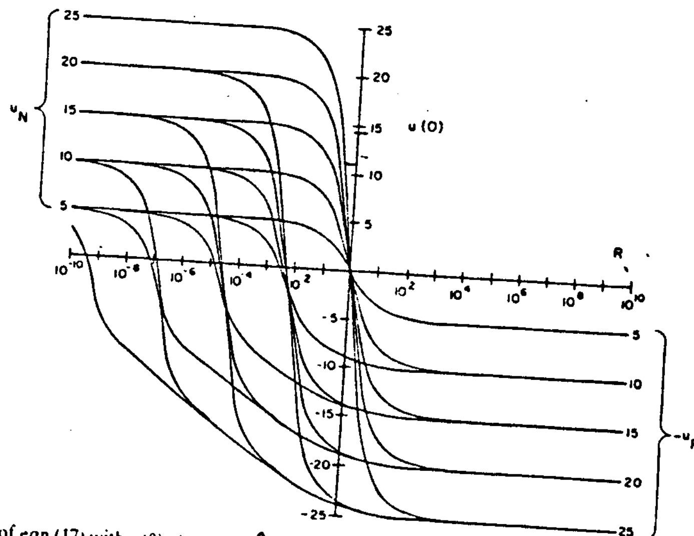

Fig. 5. Solutions of eqn (17) with $u(0)$ plotted vs R for various values of u_N, u_P. Note that interchanging u_N and u_P has the effect of replacing R by R^{-1}, thus giving a further set of solutions.

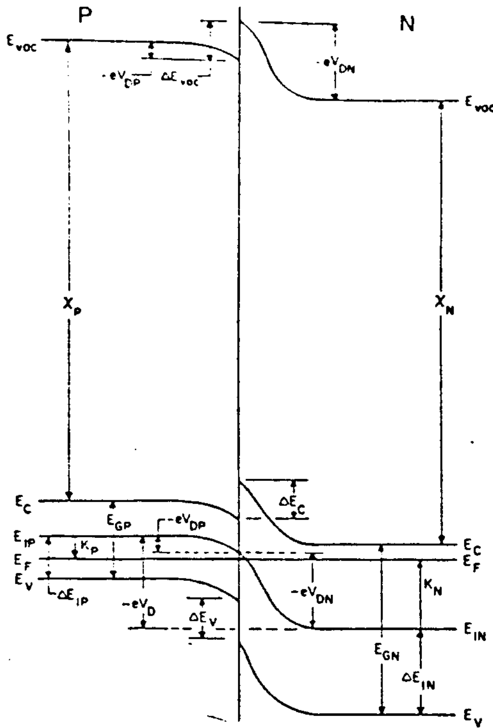

Fig. 6. New model heterostructure band diagram with continuity of intrinsic level at $x = 0$.

Similarly, the corresponding relation for the valence band discontinuity is

$$\Delta E_V = \frac{E_{GN} - E_{GP}}{2} + \frac{3kT}{4} \log (c_N/c_P) \qquad (20)$$

and it may be further shown that the vacuum level has a discontinuity given by the equation

$$\Delta E_{vac} = \frac{E_{GN} - E_{GP}}{2} - (\chi_P - \chi_N) - \frac{3kT}{4} \log (c_N/c_P). \qquad (21)$$

Since the second term on the right in both (19) and (20) is usually not more than about 20% of the first term in these equations, it follows that the discontinuities predicted by these new versions of the affinity rule are roughly the same in the two bands; this is in sharp contrast to Anderson's rules, which can lead to very large asymmetries in these quantities.

5. THE BEHAVIOR OF THE INTRINSIC LEVEL

Let us now examine in more detail the assumption about the continuity of the intrinsic level, which was introduced in an *ad hoc* manner in the previous section. It has been previously pointed out by Chang[23] that the Fletcher boundary conditions[24], which stipulate the carrier concentrations at the edges of the space-charge region in a *PN* junction under low-level injection, can be generalized to include heterojunctions. In reaching this conclusion, Chang found that eqns (13) and (14) should be replaced by eqns (19) and (20) in order to use a definition of barrier potential that would be valid for structures with discontinuities in energy levels. It is possible to eliminate any dependence on the Fletcher conditions and, in addition, any questionable assumptions about the nature of the intrinsic level, by starting with the situation illustrated in Fig. 3 for E_C, E_I, and E_V. The energy $-eV_D$ associated with the diffusion potential will have a contribution from each of the two regions and thus should be

$$\begin{aligned} -eV_D &= kT \log (n_{0N}/n_{iN}) + kT \log (n_{iP}/n_{0P}) \\ &= \frac{kT}{2} \log (n_{0N}/n_{iN})^2 + \frac{kT}{2} \log (n_{iP}/n_{0P})^2 \\ &= \frac{kT}{2} \log (n_{0N}/p_{0N}) + \frac{kT}{2} \log (p_{0P}/n_{0P}) \end{aligned}$$

using Boltzmann non-degenerate statistics and the usual symbols. Hence, for a heterojunction

$$-eV_D = (kT/2) \log (n_{0N}p_{0P}/n_{0P}p_{0N})$$

and this, of course, reduces to the customary expression if the two materials are identical. The argument of the logarithm is

$$\frac{n_{0N}p_{0P}}{n_{0P}p_{0N}} = \left(\frac{m_{eN}m_{hP}}{m_{eP}m_{hN}}\right)^{3/2} \frac{e^{(E_{CP}-E_{CN})/kT}}{e^{(E_{VN}-E_{VP})/kT}}$$

so that the diffusion potential is now

$$-eV_D = \frac{3kT}{4} \log \frac{c_P}{c_N} + \frac{1}{2}[(E_{CP} - E_{CN}) + (E_{VP} - E_{VN})].$$

But the barrier should also be a measure of what is happening in the conduction band; in fact, it should be $E_{CP} - E_{CN}$ corrected by the discontinuity ΔE_C, or

$$-eV_D = (E_{CP} - E_{CN}) + \Delta E_C.$$

If these two expressions for the diffusion potential are to be consistent, it then follows that

$$\begin{aligned} \Delta E_C &= \frac{(E_{VP} - E_{VN}) + (E_{CN} - E_{CP})}{2} - \frac{3}{4} kT \log \left(\frac{c_N}{c_P}\right) \\ &= \frac{E_{GN} - E_{GP}}{2} - \frac{3}{4} kT \log \left(\frac{c_N}{c_P}\right). \end{aligned}$$

In the same way, there should be an expression

$$-eV_D = (E_{VP} - E_{VN}) - \Delta E_V$$

involving the discontinuity in the valence band and this, when combined with the original expression, leads to

$$\Delta E_V = \frac{E_{GN} - E_{GP}}{2} + \frac{3}{4} kT \log \left(\frac{c_N}{c_P}\right).$$

We thus see that if three different possible definitions of the barrier potential are to be mutually consistent, then

the discontinuities in the valence and conduction bands must have the form of eqns (19) and (20).

Although nothing was stipulated in the argument just given about the specific nature of E_I at the junction, we can deduce an important conclusion about the behavior of this level from the fact that we have arrived at our proposed affinity rule, eqns (19) and (20), in a different manner from that of Section 4. Returning to eqn (18), it is seen that the reason why $-eV_D$ and its two components drop out is because we have implicitly required that

$$-eV_D = -eV_{DN} + -eV_{DP}.$$

Thus, if E_I is the level which represents a measure of the internal electrostatic potential and if it is taken to be continuous, then the analysis of Section 4 is valid. These assumptions, however, are not crucial, as has just been shown.

Another way of comparing the Anderson affinity rule with the one that we are advancing is based on the following physical argument: consider a homojunction with the band structure of Fig. 2 and apply tension to the N region only. This process will lower the symmetry from cubic to tetragonal, there will be an associated change in the size of the energy gap, and we will have created a heterojunction with a virtually perfect lattice match at the interface. Some internal level (not necessarily the intrinsic level) can be chosen to specify the electrostatic potential and it must be continuous; a discontinuity implies that work can be done with zero displacement, and this is possible only if the field is infinite. Furthermore, the final arrangement of the energy levels and of the inevitable discontinuities will depend only on the internal rearrangements in the distorted lattice, so that—as we are predicting—the new band structure depends only on inherent parameters of the two parts of the junction and not on work functions. On the other hand, the change in the surface dipole layer should result in a discontinuity in the vacuum level, and this effect is also a feature of our model.

6. THE DOUBLY-INTRINSIC HETEROJUNCTION

In our search for specific physical situations which might be serve as a basis for using one model in preference to the other, it was realized that it is possible to integrate eqn (1) twice and express the relative chemical potential u in terms of standard functions (i.e. to derive an analytic expression for the shape of the bands) in one special case—namely, intrinsic material. In fact, this is a situation for which the depletion approximation would be completely invalid in any case, so that the Anderson approach would require some modification. As previously noted, however, this is not regarded as a serious objection.

Let us therefore extend the treatment of Section 2 to a heterojunction composed of intrinsic germanium and intrinsic gallium arsenide. Equation (1) then reduces to

$$\frac{d^2u}{dr^2} = \sinh u \tag{22}$$

with a first integral

$$du/dr = \sqrt{[2(\cosh u + C)]}. \tag{23}$$

To evaluate the arbitrary constant C, we use the boundary condition that $u = 0$ for intrinsic material in the equilibrium regions of the junction, where du/dr vanishes, so that

$$C = -1.$$

Hence

$$du/dr = \pm\sqrt{[2(\cosh u - 1)]} = 2\sinh(u/2). \tag{24}$$

There is an ambiguity of sign in this result which we shall discuss shortly. Rearranging (24)

$$dr = du/\{2\sinh(u/2)\} = \tfrac{1}{2}\cosh(u/2)\,du$$

and integrating again

$$r = \log_e \tanh(u/4) + C'. \tag{25}$$

To evaluate the second arbitrary constant C', let

$$u = u(0) \quad \text{at} \quad r = 0. \tag{26}$$

Then

$$C' = -\log\tanh\{u(0)/4\}$$

and

$$r = \log\left[\frac{\tanh(u/4)}{\tanh\{u(0)/4\}}\right] \tag{27}$$

or

$$u = 4\,\text{arc}\tanh[e^r \tanh\{u(0)/4\}]. \tag{28}$$

This expression for u is equivalent to one obtained for the potential in the semiconductor part of a metal-intrinsic Schottky diode by McKelvey[25].

To apply these relations to a specific heterojunction, we again consider the Ge–GaAs pair. Equation (17) in the case of a doubly-intrinsic structure simplifies to

$$R = \frac{\cosh u_1(0) - 1}{\cosh u_2(0) - 1} \tag{29}$$

where the subscripts 1, 2 refer to GaAs and Ge, respectively. Using the constants cited by Milnes and Feucht[2], we obtain the Anderson-model energy band diagram of Fig. 7, which shows that the barrier is 0.27 eV. It therefore follows that

$$|u_1(0)| + |u_2(0)| = 0.27/kT \tag{30}$$

Using $R = 4.03 \times 10^6$, as indicated on Fig. 4, we must solve (29) and (30) simultaneously using a numerical

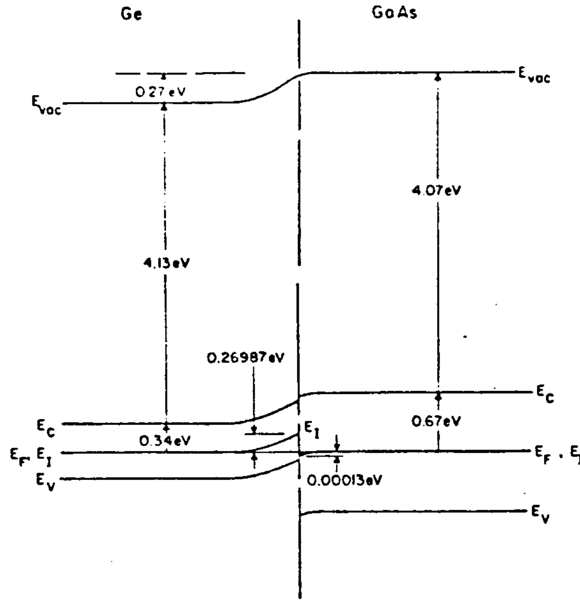

Fig. 7. Energy band diagram for a doubly-intrinsic Ge/GaAs heterojunction, assuming a continuous vacuum level.

method; the Newton–Raphson procedure (which turned out to be somewhat difficult to apply) gives magnitudes of

$$|u_1(0)| = 0.00013, \quad |u_2(0)| = 0.26987$$

and these results are also incorporated into Fig. 7.

Because of the tremendous difference in these two values of relative chemical potential at the junction, Fig. 7 cannot show the band bending with any kind of reasonable scale. Furthermore, it is also difficult to show the continuity of the normal component of $\vec{D}$ across the junction, since the band diagram indicates $\mathrm{d}u/\mathrm{d}r$, which is a measure of E_n rather than D_n. It is simple to show, however, that in the normalized units of eqn (1), we may write

$$D_n = -\sqrt{(2n_i\epsilon kT)}\,\mathrm{d}u/\mathrm{d}r.$$

The significance of this remark lies in the way in which we resolve the sign ambiguity mentioned in connection with eqn (24). The positive sign is valid on the GaAs side of the junction and the negative sign only on the Ge side. Thus $\mathrm{d}u/\mathrm{d}r$ must change sign as the junction is crossed and this prevents D_n from being continuous unless we require that $\mathrm{d}u/\mathrm{d}r$ be identically zero on both sides.

A similar difficulty arises in connection with the final solution in eqn (28). In the equilibrium part of the Ge region, u can go to zero as required because of the term $\exp(r)$, where r is now large and negative. This same term, however, will prevent the boundary condition from being satisfied in the GaAs equilibrium region, and again we reach the conclusion that the resolution of this difficulty is for $u(0)$ to be zero.

The energy band structure based on these considerations is shown in Fig. 8, and it indicates a sharp disconinuity in the vacuum level. Although this may appear to be an objection to the model we are proposing, we believe that there is precedent for such a conclusion. As an example, Ashcroft and Mermin[21] discuss the work involved in removing an electron from the interior of a crystal through a face with, say, a (100) orientation and returning it via a (111) face. It is well established, both theoretically and experimentally, that there is a difference in work functions for these two orientations. Conservation of energy then implies that there will be a redistribution of charge between the faces and a field between them. Thus, there can be a comparatively abrupt change in the vacuum level over a relatively short distance. We shall return to this important question in Sec 8.

7. NUMERICAL EXAMPLES OF THE TWO RULES

In this section we take two common semiconductor pairs which have been extensively studied in the heterojunction literature and compare the predictions of the two affinity rules discussed in Sections 3 and 4 respectively, with experimental results. These heterojunction systems are (i) Ge–GaAs, and (ii) GaAs–$\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{As}$. A third system which has received equally extensive study, viz. Ge–Si will not be discussed here; the 4% lattice constant mismatch implies less accurate measurements of ΔE_C, ΔE_V, in view of the presence of interface states.

For the p-Ge/N-GaAs heterojunction example given by Milnes and Feucht[2] the appropriate numerical values are $\chi_N = 4.07\,\mathrm{eV}$, $\chi_P = 4.13\,\mathrm{eV}$, $E_{GN} = 1.45\,\mathrm{eV}$, $E_{GP} = 0.7\,\mathrm{eV}$ at room temperature. Hence the electron affinity rule, eqns (13) and (11), yields $\Delta E_C = 0.06\,\mathrm{eV}$, $\Delta E_V = 0.69\,\mathrm{eV}$. The other parameter values for this example[2] are $(N_D - N_A)_N = 10^{16}\,\mathrm{cm}^{-3}$, $(N_A - N_D)_P = 3\times 10^{16}\,\mathrm{cm}^{-3}$, $\Delta_N = 0.1\,\mathrm{eV}$, $\Delta_P = 0.14\,\mathrm{eV}$, $\epsilon_N = 11.5$, $\epsilon_P = 16$. Hence eqns (8) and (9) may be solved for V_{DN}, V_{DP} to yield $eV_{DN} = 0.42\,\mathrm{eV}$ and $eV_{DP} = 0.10\,\mathrm{eV}$.

Applying the new model to this same example, with

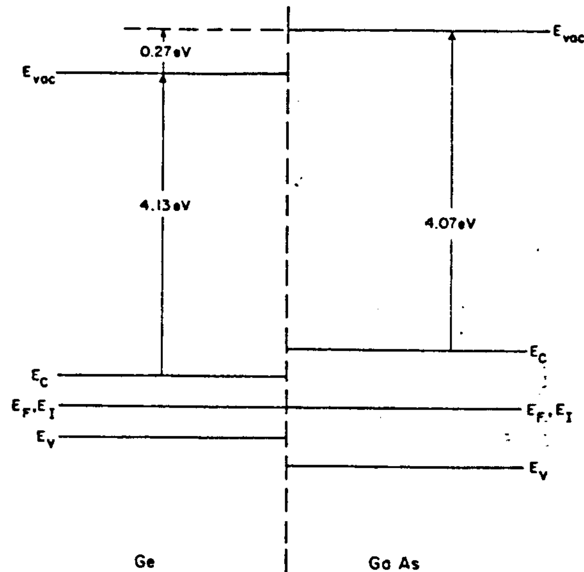

Fig. 8. Energy band diagram for a doubly-intrinsic Ge/GaAs heterojunction, assuming a continuous intrinsic level.

$C_N = 21.67$, $C_p = 2$ at room temperature, eqns (19) and (20) give $\Delta E_C = 0.33$ eV. $\Delta E_V = 0.42$ eV.

It follows that for the new model $eV_{DN} = -K(-\infty) + K(0) = 0.17$ eV and $eVD_P = K(\infty) - K(0) = 0.62$ eV. For comparison we note that the available experimental values of ΔE_C are in the range 0.15–0.19 eV[2] and it would seem to be impossible to decide the relative accuracies of the two models from these figures. ~~figures.~~

For a typical heterostructure laser containing a p-GaAs/n-$Al_{0.2}Ga_{0.8}As$ heterojunction, the appropriate parameters are $\chi_P = 4.07$ eV, $E_{GP} = 1.45$ eV, $E_{GN} = 1.70$ eV; however, the value of χ_N is in some doubt. We may estimate the value of χ_N by interpolation from the values for AlAs and GaAs. Using the values given in Ref. [2], this estimate yields $\chi_N = 3.96$ eV. Hence the conventional affinity rule, eqns (13) and (14), yields $\Delta E_C = 0.11$ eV, $\Delta E_V = 0.14$ eV. From the new model assuming $C_N = C_P$ [13] (and noting that any small deviations from this will have negligible effect), eqns (19) and (20) give $\Delta E_C = \Delta E_V = 0.125$ eV. Hence the two models seem to give fairly close agreement. An experimental value[14] for p–p heterojunctions in this material yields $\Delta E_C = 0.22$ eV, $\Delta E_V = 0.03$ eV in sharp disagreement with both models. Further experimental evidence[15, 16] on laser junctions confirms this trend, although the results are open to discussion as a result of possible band-gap grading[17] in the immediate vicinity of the heterojunction.

A recent attempt[9] to verify the conventional affinity rule used InP/CdS junctions for which eqn (13) yields a negative value of ΔE_C. Whilst this is in direct conflict with our result (19), we would question the validity of obtaining V_{DN}, V_{DP} values from a capacitance-voltage plot[9]. Since in this case the wide-gap region was highly doped, we refer the reader to the warnings on this procedure given by Cserveny[7], Hampshire *et al.*[18] and Tansley[19].

8. THE BEHAVIOUR OF THE VACUUM LEVEL

We have indicated that one of the principal changes which we are proposing in the Anderson model is the shift of the continuity condition involving D_n from the vacuum level to the intrinsic level and this change has resulted in a discontinuity ΔE_{vac} in the vacuum level, as well as a new form of the affinity rule. The validity of the original form has recently been discussed by Kroemer[8] at considerable length. Briefly his criticisms are: (i) the affinity rule, eqn (13), is in most practical cases a small difference between two large numbers χ_P, χ_N and therefore difficult to determine accurately. (ii) the values of χ_P, χ_N are different for different crystal planes and this variation alone can swamp the differences ΔE_C, ΔE_V. (iii) the affinity rule is implicitly based on conditions measurable only at a free surface, whereas the true situation at a semiconductor heterojunction interface is likely to be quite different as a result of interactions.

As a result of these criticisms Kroemer[8] proposes discarding the affinity rule and attempting to calculate ΔE_C, ΔE_V from a detailed knowledge of the band structures of the two materials. Since then, Frensley and Kroemer[20] have performed this calculation and believe that they have in fact confirmed the Anderson form of the affinity rule. However, their approach involves a number of assumptions—plus the approximations which are inherent in the pseudopotential method—and there is no really satisfactory way of accurately determining the validity of this calculation.

With regard to the location of the discontinuity, Ashcroft and Mermin[21] have pointed out that there are some inconsistencies in the common definitions of work function and the associated concepts. They point out that when an electron leaves a periodic structure, within the first few lattice spacings it is subject to a potential and a force which arise from the very complicated charge arrangement left behind and not, as usually stated, to an image force which falls off as $1/r^2$ and hence requires the

vacuum level to be placed at infinity. Instead, the force field is the well-known multipole expansion, but with one significant difference: the monopole term vanishes for an essentially neutral material, the dipole term does likewise if there is inversion symmetry, and for a simple cubic lattice (which is not unreasonable for some metals) even the quadrupole term will disappear. Hence, the lowest order non-vanishing contribution involves $1/r^5$, and this falls off so rapidly, that the potential energy is essentially zero very close to the surface. This is the proper position for the vacuum level, provided we bring the electron to rest (in principle). Furthermore, the energy involved in reaching this location should depend on the nature of the surface as well as the position of the Fermi level and this indicates why there is a difference in work functions associated with different crystalline directions. As a corollary, there must be a discontinuity between the vacuum levels associated with surfaces of different orientation on a given crystal, and this statement agrees with experimental results. We thus conclude that there is nothing inherently objectionable about a discontinuity in the vacuum level, provided one understands what is meant by the term. This point has been discussed at length by Hagstrum[22], who speaks about a near-surface vacuum level and a second one at infinity.

9. SUMMARY AND CONCLUSIONS

Our purpose here has been to present the evidence for and against each of two possible models for the ideal heterojunction. On the theoretical side, we have shown that the customary affinity rule leads to an identification of the barrier or diffusion potential in terms of external rather than internal characteristics of the junction, and objections to this can be raised. On the other hand, if our proposed conclusion about the continuity of the intrinsic level is correct, then other criticisms have to be answered. Those that have been expressed are (1) since the intrinsic level is fictitious in the sense that it normally does not correspond to occupied states and (2) electrostatic potential can only be specified to within an arbitrary constant, it is not proper to attribute special properties to E_I or to identify it as the absolute potential. Our reply to these points is that our use of E_I (following Shockley[26]) is simply to express Boltzmann statistics in a symmetrical form, which results in a very simple version of the Poisson equation, namely eqn (1). We have also discussed the original objections of Kroemer[8] and his later[20] reversal of this position. Here we should remark that pseudopotential calculations require approximations which are sometimes difficult to justify, so that some skepticism might be in order. The example of the doubly-intrinsic heterojunction worked out in Section 6 appears to support our alternate model, but it can be argued that the associated questions that are then raised about the nature of the vacuum level have not been fully answered.

The experimental aspect of choosing one model in preference to another is equally difficult. Although we have mentioned some prior work in which the fit between theory and experiment is not very satisfactory, it must be recognized that obtaining reliable data in the first place is not an elementary matter. Material parameters, such as impurity concentrations and dielectric constants, must be accurately known and the preparation of a heterojunction with good lattice match and little or no surface states is not a trivial matter. We are therefore justified in saying that the examples we have cited indicate that there is some difficulty with present theory, but the evidence is not overwhelming as yet.

We consequently summarize our survey of the Anderson model and our proposal for an alternate band structure by expressing the opinion that a strong case for the most appropriate of these two alternatives cannot be made at this time. What is needed, we believe, are more basic theoretical studies—with regard to both band structure and current-voltage diode characteristics—as well as more extensive experiments designed to confirm or contradict some of the ideas that we have raised.

Acknowledgement—One of us (M.J.A.) is indebted to the U.K. Science Research Council for the award of a Research Fellowship.

REFERENCES

1. For a recent review of heterostructure lasers, see P. R. Selway, *Proc. IEE* 123, 609 (1976).
2. A. G. Milnes and D. L. Feucht, *Heterojunctions and Metal Semiconductor Junctions*. Academic Press, New York (1972).
3. K. A. Shore and M. J. Adams, *J. Optical and Quantum Electron.* 8, 269 (1976).
4. R. L. Anderson, *Solid-St. Electron.* 5, 341 (1962).
5. W. G. Oldham and A. G. Milnes, *Solid-St. Electron.* 6, 121 (1963).
6. W. G. Oldham and A. G. Milnes, *Solid-St. Electron.* 7, 153 (1964).
7. S. I. Cserveny, *Int. J. Electronics* 25, 65 (1968).
8. H. Kroemer, *Crit. Rev. Solid State Sci.* 5, 555 (1975).
9. J. L. Shay, S. Wagner and J. C. Phillips, *Appl. Phys. Letts.* 28, 31 (1976).
10. A. Nussbaum, *Semiconductor Device Physics*. Prentice-Hall, Englewood Cliffs, New Jersey (1962).
11. A. Nussbaum, *Phys. Stat. Sol.* (*a*) 19, 441 (1973).
12. D. E. Fulkerson, *IEEE Trans. Elec. Dev.* ED-15, 404 (1968).
13. A. A. Immorlaca and G. L. Pearson, *Appl. Phys. Letts.* 25, 570 (1974).
14. R. Dingle, W. Wiegmann and C. H. Henry, *Phys. Rev. Letts.* 33, 827 (1974).
15. Zh. I. Alferov, V. M. Andreev, V. I. Korolkov, E. L. Portnoi and D. N. Tret'yokov, *Sov. Phys.-Semiconductors* 2, 843 (1969).
16. Zh. I. Alferov, V. M. Andreev, V. I. Murygin and V. I. Stremin, *Sov. Phys.-Semiconductors* 3, 1234 (1970).
17. J. F. Womac and R. H. Rediker, *J. Appl. Phys.* 43, 4129 (1972).
18. M. J. Hampshire, G. A. Wilson and M. M. Chadda, *Brit. J. Appl. Phys.* (*J. Phys. D*) 1, 853 (1968).
19. T. L. Tansley in *Semiconductors and Semimetals*, Vol. 7, (Edited by R. K. Willardson and A. C. Beer), Vol. 7, Academic Press, New York (1971) p. 293.
20. W. R. Frensley and H. Kroemer, *J. Vac. Sci. Tech.* 13, 810 (1976).
21. N. W. Ashcroft and N. D. Mermin, *Solid State Physics*. Holt, Rinehart and Winston, New York (1976).
22. H. D. Hagstrum, *Surf. Sci.* 54, 197 (1976).
23. Y. F. Chang, *J. Appl. Phys.* 36, 3350 (1965).
24. N. H. Fletcher, *J. Elec.* 2, 609 (1957).
25. J. P. McKelvey, *Solid State and Semiconductor Physics*. Harper and Row, New York (1966).
26. W. Shockley, *Electrons and Holes in Semiconductors*. D. van Nostrand, Princeton (1950).