Statistics of the Charge Distribution for a Localized Flaw in a Semiconductor

W. SHOCKLEY AND J. T. LAST
Shockley Semiconductor Laboratory of Beckman Instruments, Inc., Mountain View, California
(Received April 5, 1957)

A localized flaw such as a vacancy-interstitial pair, gold atom, or small void is considered. It is found that the charge situation can be described by a set of energy levels which are independent of the Fermi level but which are temperature-dependent. If the flaw is electrically neutral for a certain position of the Fermi level, then as the Fermi level is raised from the neutral level successively above each higher-lying level of the set, the flaw acquires an additional electron unit of negative charge. These levels are called the first-acceptor level, second-acceptor level, etc. The energy levels are essentially equal to $kT\ln(Z_1/Z_2)$, where Z_1 and Z_2 are the contributions of the flaw to the state sum for the system for the two conditions of charge. Similarly, as the Fermi level is lowered below the neutral point, it passes the first-donor level, second-donor level, etc., and the flaw acquires charges of plus one, plus two, etc. The statistics are derived for the distribution for the various conditions of charge, referred to as first-donor condition, neutral condition, first-acceptor condition, etc.

1. INTRODUCTION

IT has now been well established that gold and copper impurities in germanium introduce a series of levels in which the impurity center has charges which may vary from $+1$ to -3 electronic charges. The most elegant evidence has been published by Woodbury and Tyler[1] and makes use of the fact that an impurity of charge -2 contributes approximately twice as much impurity scattering as two separated charges of -1 each.

The statistics of defects with several trapping levels in semiconductors has been treated briefly by Landsberg,[2] and another treatment has been given by Champness.[3] However, a simplified and consistent scheme which lends itself to incorporation in the customary energy-level diagrams for semiconductors has not apparently been published.

It is possible to develop such a scheme which is applicable in much the same way as the simpler diagrams for group V donors and group III acceptors and this leads to a relatively simple conceptual structure for describing and interpreting the changing charge on the defect as the Fermi level and temperature vary.

For the purpose of distinguishing between imperfections with multiple possibilities for charge condition and ordinary donors and acceptors, in this article they shall be referred to as flaws. The number of flaws per unit volume is then N_f. This eliminates difficulties with subscripts a (acceptor), d (donor), c (conduction), i (intrinsic) which might arise from using center, defect, or imperfection. The theory for flaws discussed here is probably applicable to any localized imperfection, such as a vacancy-interstitial pair, a small dislocation loop, or an inclusion, provided that the dimensions are small compared to a Debye length for the mobile carriers present. No attempt will be made to ascertain the limits of validity and the flaws concerned can be best thought of as substitutional or interstitial atoms.

The theory is also restricted to such low densities of flaws that interactions between them can be neglected. No consideration is given to special effects due to degeneracy in the hole or electron gases.

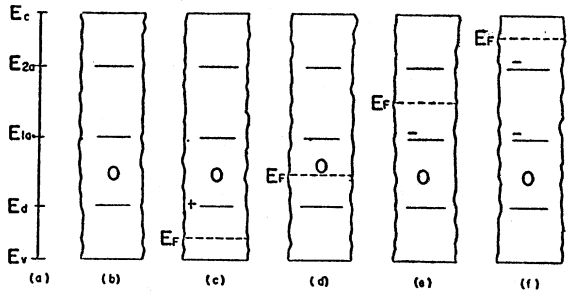

FIG. 1. Energy-level scheme for a flaw with possible charge conditions 1, 0, -1, and -2. (a) The energy levels and their notation; (b) the energy band diagram; (c)–(f) the dependence of charge condition upon the Fermi level. [To be consistent with the text, E_d should be replaced by E_{1d} in (a).]

2. PROPOSED ENERGY-LEVEL DIAGRAM

Figures 1 and 2 show the proposed energy-level diagram and the relationship between Fermi level and charge. For the particular example considered, it is supposed that the flaw may exist in four charge conditions: $+1, 0, -1$, and -2. These conditions are referred to as the first-donor, neutral, first-acceptor, and second-acceptor states.

To a good, but not perfect, approximation a flaw with the energy-level scheme of Fig. 1(a) acts like a group of three imperfections consisting of a donor and two acceptors with ionization energies of E_c-E_{1d} for electrons and $E_{1a}-E_v$ and $E_{2a}-E_v$ for holes, respectively. The functional dependence of the charge upon the flaw is closely represented by the combined behavior of the three separate imperfections.

As a preface to the exposition, it should be noted that by definition a donor with energy level E_d has a net charge of $+1$ when the Fermi level E_F is well below E_d,

[1] H. H. Woodbury and W. W. Tyler, Phys. Rev. **105**, 84 (1957).
[2] P. T. Landsberg, Proc. Phys. Soc. (London) **A65**, 604 (1956).
[3] C. H. Champness, Proc. Phys. Soc. (London) **B69**, 1335 (1956).

and is *neutral* when E_F is well *above* E_d. To a first approximation, E_d is the ionization energy; this energy is defined for the system consisting of perfect crystal except for the donor and one electron and is the energy difference between the lowest energy state with the electron on the donor and the lowest energy state with the electron far from the donor. Actually, entropy plays a role in the E_{1d} of Fig. 1 and E_{1d} is actually a difference in free energies as described in Appendix 1.

Similarly, an acceptor is *neutral* for E_F *below* E_a and negatively charged for E_F above E_a.

The figure 0 on the diagram lies below all the acceptor levels and above the donor level so that when the Fermi level lies in the interval corresponding to zero, the flaw is neutral. This is the significance of the 0 on the diagram.

On the basis of these definitions, it is readily seen that the state of charge of the flaw varies with height of the Fermi level as shown in Fig. 1(c) to (f). The net charge on the flaw will then vary as shown in Fig. 2.

Now the model of Fig. 1 actually contains some meaningless possibilities. For example, the state of charge zero can be obtained in addition to the manner shown by putting one electron on E_{1d} and one hole on either E_{1a} or E_{2a}. Such hypothetical excited states will not in general represent at all correctly the excited states of the flaw when in the neutral state.

It is thus evident that several features of the combination donor and acceptor model have no counterpart in a general model of a flaw. Nevertheless, the diagrams of Figs. 1 and 2, when properly interpreted, do usefully describe the most important aspects of the flaw.

3. INTERPRETATION AND USE OF THE ENERGY LEVEL DIAGRAM

In this section we discuss the unique way of ascribing a set of energy levels to a flaw and using them to determine the average charge and the probability of various charges on the flaw. For this purpose a set of definitions is introduced to describe the state of charge of the flaw and certain energy levels. Thus, if the flaw has charge $+2$ it is said to be in the *second-donor condition*. In general, such a condition of charge will be degenerate and may have excited states. Thus, if N_{2d} flaws per unit volume are in the second-donor condition, they will be statistically distributed over all the various possible energy states.

Similarly, the number of flaws with zero net charge is N_0 and these are said to be in the neutral condition. The conditions of negative net charge are defined in a similar manner as *acceptor conditions*.

As discussed below, a set of energy levels is introduced. These have to do with transitions of one charge unit in condition of charge according to the following scheme:

$$\cdots 2d,\ E_{2d},\ 1d,\ E_{1d},\ 0,\ E_{1a},\ 1a,\ E_{2a},\ 2a,\ \cdots.$$

Thus, each energy has a subscript corresponding to state of higher magnitude of charge of the pair to which it is related.

The values of the energy levels are defined as follows: Consider a flaw in its most positive condition, say, with charge $+n$; that is, the nth-donor condition. Let an electron be removed from the conduction band and be put into the flaw, reducing the charge to $+(n-1)$. Then the increase in free energy $E-TS$ for the semiconductor will be (as shown in Appendix 1)

$$\Delta F_{nf}=E_{nd}-E_F, \tag{3.1}$$

where E_{nd} is the change in free energy associated with adding one electron to the flaw. It includes all effects due to spin degeneracy and excited states for both states of charge. Obviously, if the flaw is initially positively charged, it can bind an electron from the conduction band and E_{nd} will lie below E_c.

Similarly, $E_{(n-1)d}$ is defined in terms of adding an electron to the $(n-1)$th donor condition to produce the $(n-2)$th donor condition. The zeroth-donor condition is the neutral condition and adding an electron to it produces the first-acceptor condition and increases the free energy by $E_{1a}-E_F$.

Now since each additional electron is put upon a less positive or even more negative flaw, it is probable that each E value will continuously increase in the sequence, going from nd through neutral and into the acceptor conditions. Finally, an E_{ma} value will be reached, which is larger than E_c. Such a negatively charged flaw will be unstable (except possibly in a degenerate n-type semiconductor) and will lose its mth negative charge to the conduction band. Similarly, a donor condition nd with energy E_{nd} lying below E_v will be unstable and will lose a hole to the valence band.

On the basis of this set of definitions, it is evident that, for any given flaw, a uniquely defined energy-level diagram like that of Fig. 1 will be obtained when the set of E_{nd} and E_{ma} values is plotted on same scale as E_c and E_v.

From the energy-level diagram, the distribution of charge values for the flaw is found as follows: For a

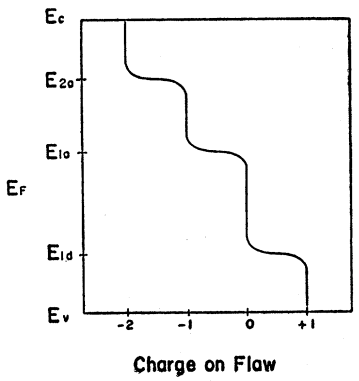

FIG. 2. The dependence of the charge on the flaw upon the Fermi level.

TABLE I. Relative probabilities for various charge conditions.

Charge condition	Increased probability for one more electron	Relative probability of being in charge condition
$3a$ (-3)	$\exp[\beta(E_F-E_{3a})]$	$W_{3a}=\exp[\beta(3E_F-E_{1a}-E_{2a}-E_{3a})]$
$2a$ (-2)	$\exp[\beta(E_F-E_{2a})]$	$W_{2a}=\exp[\beta(2E_F-E_{1a}-E_{2a})]$
$1a$ (-1)	$\exp[\beta(E_F-E_{1a})]$	$W_{1a}=\exp\ \beta(E_F-E_{1a})$
0 (0)	$\exp[\beta(E_F-E_{1d})]$	$W_0=1$
1d $(+1)$	$\exp[\beta(E_F-E_{2d})]$	$W_{1d}=\exp[\beta(E_{1d}-E_F)]$
$2d$ $(+2)$		$W_{2d}=\exp[\beta(E_{1d}+E_{2d}-2E_F)]$

given temperature with

$$\beta=1/kT \tag{3.2}$$

and a given Fermi level E_F, the ratio of the number $N_{(n-1)d}$ of flaws in charge condition $+(n-1)$ to the number N_{nd} in charge condition n is shown in the Appendix to be

$$N_{(n-1)d}/N_{nd}=\exp[\beta(E_F-E_{nd})]. \tag{3.3}$$

Thus, if E_F lies above a level, the charge state having one more electron is favored. The favoring factor is in fact exactly that associated for the simple Fermi-Dirac distribution. For example, for a simple donor level (neglecting all degeneracy), the probabilities f_n and f_p of the level being occupied by an electron or a hole are

$$f_n=\{1+\exp[\beta(E_d-E_F)]\}^{-1}, \tag{3.4}$$

$$f_p=1-f_n=f_n\exp[\beta(E_d-E_F)], \tag{3.5}$$

from which it follows that

$$N_0/N_{1d}=f_n/f_p=\exp[\beta(E_F-E_d)]. \tag{3.6}$$

According to the analysis in Appendix 1, this same relationship applies to pairs of charge states differing by an energy E_{nd} or E_{ma} as the case may be.*

The relative probability of finding any one of the various charge conditions may readily be determined by a straightforward calculation from the factors (3.3). For this purpose, consider Table I.

The first column lists the charge conditions. The second column is the factor giving the ratio of probability of the upper condition of the table compared to the lower condition; these factors are simply those of Eq. (3.3). Each factor is greater than unity if the Fermi level E_F is greater than the free-energy increase due to adding one more electron to the center. From these factors the third column is readily obtained; the relative probability of unity for the neutral condition being picked for symmetry.

The absolute probability is obtained by normalizing which consists simply of dividing each entry in column three by the sum of column three: This leads to the following form of expression for the distribution among charge conditions:

$$N_{nd}=N_0W_{nd}/\sum W\text{'s}. \tag{3.7}$$

The dependence of the distribution on E_F may be visualized as follows: If

$$E_{1d}<E_F<E_{1a}, \tag{3.8}$$

then W_0 is larger than any other W. In fact, if $E_{1a}-E_{1d}$ is several tenths of an electron volt and E_F lies near the middle, the other W's add very little to the sum of column 3 and the probability of the neutral condition is nearly unity. The probability of differing from the most favored case by two units of charge is very small and varies with E_F as $\exp(\pm 2\beta E_F)$.

It is evident that whenever E_F lies between two energy levels, then the charge condition common to the two energy differences tends to dominate the probability.

When E_F equals one of the energy levels, then the two adjacent conditions are equally probable, each having approximately probability $\frac{1}{2}$.

As E_F rises from between E_{1d} and E_{1a}, the probability of the neutral condition decreases from nearly unity to about $\frac{1}{2}$ for $E_F=E_{1a}$ and then falls as $\exp(-\beta E_F)$ until $E_F=E_{2d}$; thereafter the probability of the neutral condition falls as $\exp(-2\beta E_F)$. It falls as $\exp(-3\beta E_F)$ after E_F crosses E_{3a}.

From these examples it is seen that, except when E_F equals an energy level, one charge condition dominates the probability distribution of charge conditions. Furthermore, it is evident that these charge conditions correspond to the integral charge cases of Fig. 2. Thus, Fig. 2 does represent adequately well the average charge on the flaw when the energy levels are as interpreted in this section.

As shown in connection with Eq. (3.6), the transition from one charge condition to another in Fig. 2, is well represented by a Fermi-Dirac distribution function.

As an example of this system of treatment, the case of two separate donors treated as a composite flaw is given in Appendix 2.

APPENDIX 1. DERIVATION OF THE DISTRIBUTION LAW

The relative probability of different charge conditions on the flaw can be derived by finding the distribution that maximizes the probability. For this purpose we describe the possible states of the center in terms of its net charge, which is s electronic charges, and its state of excitation i, the various states of a degenerate energy level having different i values. The number of flaws in excited state i of charge condition s is N_{si}. Evidently the total number of flaws N_f is given by

$$N_f=\sum_{si}N_{si}. \tag{A1.1}$$

* The equilibrium and nonequilibrium behavior of flaws has been investigated from the point of view of the flaw as a recombination center, using the principle of detailed balancing. This also leads to equations like (3.3). It is planned to submit to this journal a manuscript on this subject by C.-T. Sah and W. Shockley.

The total number of electrons in the flaws

$$n_f=\sum_{si} sN_{si}, \tag{A1.2}$$

where, for reference, a flaw of zero charge is said to have zero electrons; it can be shown that the zero of electron number is not important. If the energy of state (s,i) is E_{si} (for example, the state of zero energy might be chosen as most positive flaw in lowest state and all electrons at the bottom of the conduction band), then the energy of the distribution is

$$E_f=\sum_{si} E_{si}N_{si}. \tag{A1.3}$$

In order to maximize the probability of a given distribution, the number of ways W_f of arranging the distribution N_{si} on the flaws must be known. This is evidently the number of ways of selecting groups of size N_{si} from a total number $\sum N_{si}$ and can be expressed in terms of the N_{si} values as

$$W_f=(\sum N_{si})!/\prod_{si}(N_{si}!). \tag{A1.4}$$

In order to allow the flaws to interact with electrons we also introduce a distribution in the conduction band in the customary manner by dividing the states into groups of Q_j of energy E_j occupied by n_j electrons. The values for the number of ways W_n of achieving the distribution, the total number of conduction-band electrons n and their energy E_n are

$$W_n=\prod_j \frac{Q_j!}{n_j!(Q_j-n_j)!}, \tag{A1.5}$$

$$n=\sum_j n_j, \tag{A1.6}$$

$$E_n=\sum E_j n_j. \tag{A1.7}$$

The over-all most probable distribution results from maximizing

$$\ln W=\ln W_f+\ln W_n \tag{A1.8}$$

subject to constraints on total energy, total number of electrons, and number of flaws. Using the Lagrange multiplier method, we require that

$$\ln W_f+\ln W_n-\beta(E_f+E_n)+\mu(n_f+n)+\gamma N_f \tag{A1.9}$$

be an extremum. This leads, in the usual way, to the conclusion that

$$\beta=1/kT \tag{A1.10}$$

and

$$\mu=\beta E_F, \tag{A1.11}$$

where E_F is the Fermi level and is also the chemical potential for electrons, being the change in free energy

$$F=E-kT\ln W, \tag{A1.12}$$

when one electron is added. The usual Fermi-Dirac distribution arises from setting the coefficient of δn_j equal to zero.

Setting the coefficient of δN_{si} equal to zero gives

$$\ln(N_f/N_{si})-\beta E_{si}+\beta E_F s+\gamma=0, \tag{A1.13}$$

so that

$$N_{si}=N_f\exp[\beta(sE_F-E_{si})+\gamma]. \tag{A1.14}$$

Thus, the number of centers in condition s is

$$N_s=\sum_i N_{si}=N_f\exp(\gamma+\beta sE_F)\sum_i\exp(-\beta E_{si}) =N_f\exp(\gamma+\beta sE_F)Z_s, \tag{A1.15}$$

where Z_s is the state sum over the states of the flaw for charge condition s. The quantity γ is a normalization constant. It is also the electronic part of the chemical potential for flaws.

The ratio of probability of charge condition $s+1$ compared to s is

$$N_{s+1}/N_s=(Z_{s+1}/Z_s)\exp(\beta E_F). \tag{A1.16}$$

Since the free energy per system in a set of systems with energy levels E_{si} is

$$F_s=-kT\ln Z_s, \tag{A1.17}$$

the increase in free energy resulting from moving one electron from the conduction band to convert a flaw from state s to state $s+1$ is

$$F_{s+1}-F_s-E_F=+kT\ln(Z_s/Z_{s+1})-E_F =E(s+1,s)-E_F, \tag{A1.18}$$

where

$$E(s+1,s)=kT\ln(Z_s/Z_{s+1}) \tag{A1.19}$$

gives the definition of the energy levels used in Sec. 3. In terms of this definition of $E(s+1,s)$, we have

$$N_{s+1}/N_s=\exp\{\beta[E_F-E(s+1,s)]\}. \tag{A1.20}$$

This is the equation used in Sec. 3 of the text and shows how the interpretation of $E(s+1,s)$ as a free-energy difference arises.

APPENDIX 2. TWO SEPARATE DONORS TREATED AS A COMPOSITE FLAW

The case of two separate donors treated as a composite flaw is presented here to show the relationship of the flaw energy levels E_{1d} and E_{2d} to the binding energies E_1 and E_2 of the two donors. Choosing the zero of energy as discussed in Appendix 1 [see Eq. (A1.3) where the energy E_c is taken as zero], the energy required to ionize donor "1" and put the electron into energy E_c is E_1. (That is, E_1 is absolute value of energy below E_c.) On this basis for the zero of energy and on the basis of two possible spin states for a neutral donor, the states for the two donors are as shown in Table II.

Table II leads to three state sums Z_{2d}, Z_{1d}, and Z_0 as defined in Eq. (A1.15), and using these in Eq. (A1.19) one obtains two energy levels:

$$E_{2d}=kT\ln(Z_{2d}/Z_{1d}) =-kT\ln(e^{\beta E_1}+e^{\beta E_2})-kT\ln 2 \tag{A2.1}$$

$$E_{1d}=kT\ln(e^{-\beta E_1}+e^{-\beta E_2})-kT\ln 2. \tag{A2.2}$$

In order to visualize how these levels are related to the

TABLE II. The states for the two donors described as states for a composite flaw.

Charge condition	Energy	No. of ways of forming state	No. of flaws
+2	0	1	N_{2d}
+1	$-E_1$, $-E_2$	2, 2	N_{1d}
0	$-(E_1+E_2)$	4	N_0

ionization energies E_1 and E_2, consider the two following cases:

Case (1) $E_1=E_2$

$$E_{2d}=-E_1-kT\ln 4 \tag{A2.3}$$

$$E_{1d}=-E_1. \tag{A2.4}$$

For Case (1), the energy for binding the second electron to the flaw is the same as the first, but the entropy factor is less so that the first donor level lies higher than the second donor level by $kT\ln 4$.

Case (2) $E_1-E_2>kT$

$$E_{2d}\doteq -E_1-kT\ln 2 \tag{A2.5}$$

$$E_{1d}\doteq -E_2-kT\ln 2. \tag{A2.6}$$

For Case (2), the two donors have sufficiently different energies so that they act independently and the flaw energy levels coincide with those of the donors as corrected for spin degeneracy.†

It may also be shown that the probabilities of the various charge conditions as given by Eq. (3.7) reduce to those calculated by the usual Fermi-Dirac treatment. For example, from Eq. (3.7) we obtain

$$\begin{aligned} N_0/N_f&=[1+e^{\beta(E_{1d}-E_F)}(1+e^{\beta(E_{2d}-E_F)})]^{-1}\\ &=[1+\tfrac{1}{2}(e^{-\beta E_1}+e^{-\beta E_2})\\ &\quad\times e^{-\beta E_F}(1+\tfrac{1}{2}(e^{\beta E_1}+e^{\beta E_2})^{-1}e^{-\beta E_F})]^{-1}\\ &=[1+\tfrac{1}{2}e^{-\beta(E_1+E_F)}]^{-1}[1+\tfrac{1}{2}e^{-\beta(E_2+E_F)}]^{-1}. \end{aligned} \tag{A2.7}$$

This last expression is simply the product of the probabilities given by the usual Fermi-Dirac treatment that each donor, with effective energy levels of the form $-E_1-kT\ln 2$, have an electron. The reader may easily find that similar equivalences can be established between the other charge conditions of the flaw and the corresponding states of the two donors.

† See, for example, W. Shockley and W. T. Read, Jr., Phys. Rev. 87, 835 (1952), Appendix B.

PHYSICAL REVIEW VOLUME 107, NUMBER 2 JULY 15, 1957

Anomalous Electrical Resistivity and Magnetoresistance Due to an *s-d* Interaction in Cu-Mn Alloys

KEI YOSIDA*
Department of Physics, University of California, Berkeley, California
(Received April 8, 1957)

The effect of the *s-d* exchange interaction between the conduction electrons and the Mn ions on the electrical behavior of Cu-Mn alloys is investigated from the molecular-field point of view. The magnitude of the anomalous resistivity calculated with the value of the exchange integral for a free Mn^+ ion agrees with the experimental value within a factor of three. Also the temperature dependence of the resistivity obtained by the molecular-field approximation for the antiferromagnetic spin ordering is shown to be quite similar to the behavior exhibited by the alloys with more than one atomic percent Mn. For samples with lower concentration of Mn ions, however, the theoretical result shows only a monotonic decrease of the resistance below the Néel temperature. It shows neither the resistance minimum nor maximum which has been found experimentally for the very dilute alloys. The anomalous magnetoresistance calculated on the same basis is approximately proportional to the square of the magnetization and its magnitude is in good agreement with the experimental results, especially above the Néel temperature. The magnetoresistance of ferromagnetic metals is also discussed.

I. INTRODUCTION

GERRITSEN and Linde[1] have measured the electrical resistivity of noble-metal alloys such as silver and copper containing traces of the transition elements represented by manganese and iron, and have found anomalous behavior in the resistivity at low temperature. According to their experimental results, such alloys with several atomic percent Mn show an abrupt decrease of the resistivity at low temperature, and as the concentration of Mn ions becomes lower the resistivity initially rises and then decreases, exhibiting a maximum value with lowering temperature. Gerritsen[2] has also found that these alloys show an anomalous magnetoresistance accompanying the anomaly of the resistivity. Korringa and Gerritsen[3] explained this

* On leave from the Department of Physics, Osaka University, Osaka, Japan.

[1] A. N. Gerritsen and J. O. Linde, Physica 17, 573, 584 (1951).

[2] A. N. Gerritsen, Physica 19, 6 (1953).

[3] J. Korringa and A. N. Gerritsen, Physica 19, 457 (1953).