JOURNAL OF SOLID STATE CHEMISTRY 1, 88–97 (1969)

Statistical Thermodynamics of Nonstoichiometry in Nonmetallic Binary Compounds†

R. F. BREBRICK

Lincoln Laboratory, Massachusetts Institute of Technology,
Lexington, Massachusetts 02173

Received February 5, 1969

A statistical thermodynamic analysis is given for an ordered, nonmetallic compound MX in which the predominant atomic point defects are vacancies. These are associated with singly-ionizable donor and acceptor levels and are assumed to be distributed at random over the available sites. Conduction band electrons and valence band holes are considered to be present. The partition function for the Gibbs free energy is constructed as the sum of two parts: A configurational entropy times the temperature and an excess Gibbs free energy. Utilizing the fact that both these parts must be homogeneous functions of the first degree in the numbers of defects of each kind, fixes the composition dependence of the excess Gibbs free energy when expressed accurate to the 1st power of the defect numbers and forces the densities of states in the conduction and valence bands to be expressed as functions of the numbers of defects. The resulting equations reduce to those previously obtained only in the case of nondegenerate semiconductor with small defect concentrations.

Introduction

By nonstoichiometry we mean the composition of a crystalline compound is different from that unique, stoichiometric composition required by the ideal crystal structure characterizing the phase. At a given temperature the composition can in principle vary over a range, more or less close to the stoichiometric composition, but perhaps not including it. On an atomic scale we picture the presence of so-called native, atomic point defects of varying kinds and concentrations as the simplest way in which the crystal structure is preserved while the composition is changed. In nonmetallic or semiconducting compounds, the atomic point defects give rise to donor and acceptor levels in the electronic energy band structure which can affect the electrical and optical properties as well as the thermodynamic properties.

Our object in this paper is to present a statistical thermodynamic analysis of a nonstoichiometric semiconducting compound, MX, under the assumption that the atomic point defects are distributed at random over the available sites. We believe this analysis to be more complete in a thermodynamic sense than any given hitherto under the same assumption. A few remarks may be useful in placing our treatment in perspective, since many statistical mechanical analyses of point defects in solids have been published. An essential feature of the specific model adopted here is that it takes account of conduction band electrons and valence holes and allows for the possibility that changes in composition can significantly shift the Fermi level. This in itself serves to separate our treatment from those which effectively deal with metals in explicitly or implicitly assuming the Fermi level is independent of composition [Fowler and Guggenheim (*1*), Wagner (*2*), and more recently Libowitz and Lightstone (*3*)]. In addition the statistical analysis itself is different from that often employed for solids in two straightforward but important respects. In common with Wagner (*2*) and with Libowitz and Lightstone (*3*), (1) the Gibbs free energy is constructed with the explicit recognition that aside from a grouping of terms which is the product of the temperature and the configurational entropy, the remaining terms must collectively be regarded as an excess Gibbs free energy and (2) the number of lattice sites is recognized as a variable with respect to which the Gibbs free energy must be minimized. As a consequence of (1), equations are obtained which are thermodynamically consistent even when certain composition-independent, partial excess free energies must be considered to depend upon the temperature. In contrast, the common

---

† This work was sponsored by the U.S. Air Force.

method of constructing the quasi-grand partition function results in energy terms in place of the above partial excess free energies. When it is necessary to consider these energies to vary with temperature to fit experiment, an inconsistent set of equations results (*4*) and considerable confusion has arisen in the interpretation of experimental data. Thus we find for a nondegenerate semiconductor the temperature dependence of the product of the electron and hole concentrations, np, is, aside from that of the densities of states, $\exp(g/kT)$, where g is the excess Gibbs free energy to create an electron and hole. The derivative of $\ln np$ with respect to $1/T$ is the enthalpy required to create this pair. As a consequence of (2), an additional equation characterizing the equilibrium state of the compound is obtained. For small departures from the stoichiometric composition, the content of this equation is that the Gibbs free energy/g-atom of the compound depends upon composition only through terms of the order of D/N, where D is the defect concentration and N is that of lattice sites. The Gibbs free energy/g-atom is then essentially independent of the composition of the compound—an assumption commonly made in thermochemical studies and commonly used to obtain a relation between the partial pressures of the two elementary species making up the compound.

Finally, for the first time explicit use is made of the fact that the configurational entropy and the excess Gibbs free energy must both be first degree homogeneous functions of the number of defects of each kind. As a result the densities of states of the conduction and valence bands enter some of the equilibrium equations through their partial derivatives with respect to the numbers of defects. Consequently these equations are different than those obtained before, and reduce to them only for the special case that the semiconductor is nondegenerate. Moreover, because of this homogeneity property, the composition dependence of the excess Gibbs free energy is determined when expressed accurate to any order in D/N.

For simplicity of presentation we shall consider a model in which vacancies are the only atomic point defects. There is no difficulty in considering the simultaneous presence of interstitials and place exchange defects.

Description of the Model

Consider a semiconducting compound, MX, which contains M-vacancy acceptors and X-vacancy donors. Our basic assumptions are:

I. The structure of the compound consists of two sublattices each containing N-sites. The M-atoms occupy some or all of the sites of one-sublattice, the X-atoms occupy some or all of the sites of the second. Unoccupied sites in the M sublattice are called M-vacancies. Unoccupied X-sites are called X-vacancies.

II. The M- and X-vacancies are distributed at random among the sites of the sublattices.

III. The compound MX is a semiconductor. In the perfect stoichiometric compound at 0°K the most energetic electrons just fill a set of closely spaced energy levels (the valence band) which in turn is separated from the nearest empty levels. These empty levels are also closely grouped into a band called the conduction band. We emphasize the model includes both conventional semiconductors and those in which the hopping mechanism of electronic transport predominates.

IV. Each M-vacancy is associated with an acceptor level at a fixed energy E_A. Each X-vacancy is associated with a donor level at a fixed energy E_D. The donor can be occupied by at most one electron but has a spin degeneracy of two. The acceptor level has a maximum occupancy of one hole and is also two-fold spin degenerate.

V. The density of states per unit energy in the electronic energy bands is a function of the number of sites and the total numbers of M-vacancies and of X-vacancies. This dependence has heretofore been neglected as far as the author is aware. It is discussed in more detail later.

VI. The electrons are distributed among the available energy levels according to Fermi–Dirac statistics.

VII. The excess Gibbs free energy of the compound is a linear function of the number of defects of each kind and satisfies the more generally valid restriction of being a homogeneous function of the first degree in the number of defects of each kind.

These assumptions are more likely to be valid for small deviations from stoichiometry. As the concentration of donors or acceptors increases, the originally fixed donor and acceptor levels spread out in energy, forming so-called impurity bands which eventually merge with one of the main bands. This problem is currently of interest (*5–7*). The effect can be viewed as a screened coulombic interaction among the various charged species and at low temperatures can be important at relatively low concentrations of donors or acceptors (*7*). At temperatures high enough to attain equilibrium between the compound and the surrounding vapor phase in reasonable times—the situation in which

TABLE I

List of Symbols

$V_M{}^0$	M-vacancy with a hole occupying the acceptor level at energy E_A. Also called a neutral M-vacancy in current nomenclature.
$V_M{}^-$	M-vacancy with an electron, or no hole, occupying the acceptor level i.e. an ionized M-vacancy acceptor. $V_M = V_M{}^0 + V_M{}^-$ total concentration of M-vacancies.
$V_X{}^0$	X-vacancy with an electron occupying the donor level at energy E_D. Also called a neutral X-vacancy donor.
$V_X{}^+$	X-vacancy with no electron occupying the donor level i.e. an ionized X-vacancy donor. $V_X = V_X{}^0 + V_X{}^+$ total concentration of X-vacancies.
n_s	Conduction band electron with an energy between E_s and $E_s + \Delta E_s$.
N_s	Number of conduction band levels in the energy interval E_s and $E_s + \Delta E_s$.
p_r	Valence band hole with an energy between E_r and $E_r + \Delta E_r$.
N_r	Number of valence band levels in the energy interval E_r and $E_r + \Delta E_r$.
μ_M	Chemical potential of an M-atom measured relative to an isolated M-atom in its ground state.
μ_X	Chemical potential of an X-atom measured relative to an isolated X-atom in its ground state.
E_f	Fermi level characterizing the electronic energy distribution.

we are interested—the concentration of electronic carriers is often large due to excitation across the forbidden energy gap of the semiconductor. As a result the various charged species are more effectively screened and significant shifts in the energy levels occur at higher donor and acceptor concentrations (*8*). As a result of increasing defect concentrations assumptions II and VII are eventually invalid. However, developing a model in which electronic effects (as exemplified in III, IV, and V) were assumed unimportant, Libowitz and Lightstone (*3*) found assumptions II and VII to be apparently valid for a few hydrides (presumably metallic) at defect concentrations in the 10^{21} cm^{-3} range.

Table I gives a list of symbols used. Some play a double role of identifying a defect and indicating its number. Following current usage an X-vacancy is indicated by $V_X{}^0$ when an electron occupies the associated donor level and is called a neutral X-vacancy donor. When the level is not occupied by an electron the X-vacancy donor is said to be ionized and indicated by $V_X{}^+$. The neutral M-vacancy acceptor is indicated by $V_M{}^0$ and has a hole occupying the acceptor level. The hole is removed by placing an electron on the acceptor level to form an ionized M-vacancy acceptor, $V_M{}^-$. The energies E_s and E_r are arbitrary energies inside the conduction and valence bands, respectively. The number of defects given in Table I is considerably increased over the two, V_M and V_X, required if electronic defects are neglected.

General Expression for the Gibbs Free Energy

We now desire to obtain the Gibbs free energy of the compound as an explicit function of the numbers of defects of all kinds. The general form is indicated by Eq. (1).

$$G^* = G(N, V_M{}^0, V_M{}^-, V_X{}^0, V_X{}^+, \ldots n_s, \ldots, p_r, \ldots). \tag{1}$$

The superscript asterisk used in Eq. (1) and below indicates the value of the thermodynamic potential for general values of the numbers of defects. When the asterisk is omitted, the value of the thermodynamic potential is that special one, the equilibrium value, obtained by inserting the equilibrium values for the numbers of defects.

Equation (2) serves to define a quantity, $\hat{G}$, called the excess Gibbs free energy here, in terms of the ordinary Gibbs free energy and the configurational entropy, S_{con}.

$$G \equiv \hat{G} - TS_{\text{con}} \tag{2}$$

$$= -kT \ln \sum T_n \tag{3}$$

$$= -kT \ln \sum \exp[-(\hat{G}^* - TS^*_{\text{con}})/kT]. \tag{4}$$

An explicit expression for S^*_{con} is given later. Equation (3) states formally that the Gibbs free energy can also be expressed in terms of the logarithm of a sum whose general term is T_n and which is a type of partition function. The structure of the general term, T_n, of the partition function is indicated in Eq. (4) in which the excess Gibbs free energy and the configurational entropy are to be given as explicit functions of the number of defects of each kind and the sum is over all sets of such numbers consistent with constraints of Eqs. (5)–(7).

$$N - V_M{}^0 - V_M{}^- = N_M \,(\text{a constant}) \tag{5}$$

$$N - V_X{}^0 - V_X{}^+ = N_X \,(\text{a second constant}) \tag{6}$$

$$\sum_{\text{c.b.}} n_s + V_M{}^- = \sum_{\text{v.b.}} p_r + V_X{}^+. \tag{7}$$

Equations (5) and (6) are an expression of I and indicate the numbers of vacancies and atoms of each kind add up to the same number, N, the total number of sites in each sublattice. The stipulations that the number of M-atoms, N_{M}, and the number of X-atoms, N_{X} are each constant is required later when the equilibrium numbers of defects are determined. As is well known this determination depends upon the fact that the Gibbs free energy is a minimum with respect to possible changes in the numbers of defects at constant temperature, T, pressure, P, and numbers of atoms of each kind (*9*). Equation (7) is the requirement of electrical neutrality. The first summation is over the number of electrons in each narrow energy interval of the conduction band while the second is an analogous sum for holes in the valence band.

Configurational Entropy and Densities of Electronic States

Applying the usual approximation we can replace the sum in Eq. (4) by its largest term, with negligible error (*4*). The configurational entropy part of this term can be written consistent with II, IV, and, VI as shown in Eq. (8).

$$\begin{aligned} S^*_{\mathrm{con}} = k \ln \Bigg\{ & 2^{(V_{\mathrm{X}}{}^0 + V_{\mathrm{M}}{}^0)} \binom{N}{V_{\mathrm{M}}{}^0 + V_{\mathrm{M}}{}^-} \binom{V_{\mathrm{M}}{}^0 + V_{\mathrm{M}}{}^-}{V_{\mathrm{M}}{}^0} \\ & \times \binom{N}{V_{\mathrm{X}}{}^0 + V_{\mathrm{X}}{}^+} \binom{V_{\mathrm{X}}{}^0 + V_{\mathrm{M}}{}^+}{V_{\mathrm{X}}{}^0} \Bigg\} \\ & + k \ln \sum_{\mathrm{c.b.}} \binom{N_{\mathrm{s}}}{n_{\mathrm{s}}} + k \ln \sum_{\mathrm{v.b.}} \binom{N_{\mathrm{r}}}{p_{\mathrm{r}}} \end{aligned} \qquad (8)$$

where the binomial coefficient $\binom{x}{y}$ is given by

$$\binom{x}{y} \equiv \frac{x!}{x!(x-y)!}. \qquad (9)$$

The binomial coefficient $\binom{x}{y}$ gives the number of ways in which x objects can be divided into two unordered groups, one group containing y objects and the other $x - y$ objects. Thus

$$\binom{N}{V_{\mathrm{M}}{}^0 + V_{\mathrm{M}}{}^-}$$

gives the number of ways a total of $V_{\mathrm{M}}{}^0 + V_{\mathrm{M}}{}^-$ M-vacancies can be distributed among N sites. For each of these distributions there are

$$\binom{V_{\mathrm{M}}{}^0 + V_{\mathrm{M}}{}^-}{V_{\mathrm{M}}{}^0}$$

ways that $V_{\mathrm{M}}{}^0$ holes can be distributed among a total of $V_{\mathrm{M}}{}^0 + V_{\mathrm{M}}{}^-$ acceptor levels so each contains no more than one hole. The other binomial coefficients arise similarly. The binomial coefficient

$$\binom{N_{\mathrm{s}}}{n_{\mathrm{s}}}$$

gives the number of ways that n_{s} conduction band electrons with energies in the interval E_{s} and $E_{\mathrm{s}} + \Delta E_{\mathrm{s}}$ can be distributed among the N_{s} levels in that energy range so that each level is either empty or occupied by one electron (*10*). This coefficient is then summed over all energies in the conduction band to obtain the contribution from all conduction band electrons to the configurational entropy. An analogous interpretation holds for the last sum of Eq. (8) which is over the energies of the valence band. The factor 2 raised to the $V_{\mathrm{X}}{}^0 + V_{\mathrm{M}}{}^0$ power arises from the assumed two-fold spin degeneracy of the acceptor and donor levels. If in a more general case the donor level is g_{D}-fold degenerate, while the acceptor level is g_{A}-fold degenerate, but both with a maximum occupancy of one, this factor is replaced by $g_{\mathrm{D}}^{V^0{}_{\mathrm{X}}} \cdot g_{\mathrm{A}}^{V^0{}_{\mathrm{M}}}$. If the electronic defects were ignored Eq. (8) would be replaced by

$$S^*_{\mathrm{con}} = k \ln \sum \binom{N}{V_{\mathrm{M}}} \binom{N}{V_{\mathrm{X}}}. \qquad (10)$$

Now the configurational entropy given by Eq. (8) must be an extensive function of the numbers of defects if we confine our considerations to macroscopic amounts of the compound large enough that surface effects are negligible. Mathematically speaking, S^*_{con} must be a homogeneous function of the first degree in the numbers of defects. In order for this to be so, the density of states in each band, N_{s} and N_{r}, must be considered to be doubled when two identical replicas of the compound are considered to be joined. In the past the situation has implicitly been left at that. However, one can argue that N_{s} and N_{r} ought to be a function of at least N, $V_{\mathrm{M}} \equiv V_{\mathrm{M}}{}^0 + V_{\mathrm{M}}{}^-$, and $V_{\mathrm{X}} = V_{\mathrm{X}}{}^0 + V_{\mathrm{X}}{}^+$. Suppose one considers the compound with no vacancies i.e. $N = N_{\mathrm{M}} = N_{\mathrm{X}}$. The energy levels originate from atomic levels and the total number of energy levels and in particular N_{s} and N_{r} ought to be homogeneous functions of N_{M} and N_{X}. If N_{M} and N_{X} are doubled, so must N_{s} and N_{r} be doubled for every energy interval in the bands. Now consider the stoichiometric compound with vacancies present in equal numbers i.e. $V_{\mathrm{M}} = V_{\mathrm{X}}$. The introduction of vacancies is accompanied by the introduction of donor or acceptor levels that are generally imagined outside the bands but in any case are considered different from them. Since these localized levels must basically originate from atomic levels and the total number of levels must depend on N_M and N_X, the number of levels in the bands must decrease. We are led to a general assumption that the density of states (per unit energy but *not* per unit volume) in both bands ought to be homogeneous functions of N_M, V_M, N_X and V_X, or in terms of the variables of S_{con}; N, V_M, and V_X. Thus we can write:

$$N_s = NJ_1\left(\frac{V_M}{N}, \frac{V_X}{N}, E\right) + V_M J_2 + V_X J_3 \quad (11)$$

$$N_r = NH_1 + V_M H_2 + V_X H_3. \quad (12)$$

The functions J_i and H_i are all homogeneous functions of the zeroth degree (*11*) in V_M, V_N, and N and are functions of energy E, as indicated explicitly for the argument of J_1. Equations (11) and (12) indicate a dependence of the total number of levels of all kinds on the total number of atoms in a formal way. The principle has been recognized but is usually dismissed as unimportant unless the donor and acceptor concentrations exceed about $10^{20}\,\text{cm}^{-3}$. As far as the author is aware it has not been included in the statistical analysis previously. In general the functions J_i and H_i in Eqs. (11) and (12) are unknown. A possible specific form might be

$$N_s = (N - V_M - V_X)\,n_c \qquad \text{where } n_c = n_c(E) \quad (11b)$$

$$N_r = (N - V_M - V_X)\,n_v \qquad \text{where } n_v = n_v(E). \quad (12b)$$

For example, Eq. (11b) would follow with n_c unity if the conduction band was confined to a single energy, E_c, was derived entirely from a single M-atom level, and suffered a loss of one level for each N-vacancy donor level formed. Then

$$N_s = N_M - V_N = N - V_M - V_N.$$

Inserting Eqs. (11) and (12) for the densities of states into Eq. (8) for the configurational entropy and applying Stirling's approximation for the factorials, the configurational entropy is a homogeneous first degree function of the defect numbers.

Excess Gibbs Free Energy

The excess Gibbs free energy defined by Eq. (1) must be a homogeneous function of the first degree in the defect variables since the Gibbs free energy and configurational entropy themselves are. Thus one can formally write:

$$\hat{G}^* = Ng\left(\frac{V_M{}^0}{N}, \frac{V_M{}^-}{N}, \frac{V_X{}^0}{N}, \frac{V_X{}^+}{N}, \ldots, \frac{n_s}{N}, \ldots, \frac{p_r}{N}, \ldots\right) \quad (13)$$

where g is a homogeneous zeroth degree function (*11*) of the numbers of defects as implied in the form of its argument. Expanding g in a MacLaurin's series about zero values for the numbers of each kind of defect, and discarding terms higher than the zeroth degree, (VII), we can express the excess Gibbs free energy as:

$$\begin{aligned}\hat{G}^* = Ng_0 + V_M{}^0\epsilon_M + V_X{}^0\epsilon_X + V_M{}^-(\epsilon_M + E_A) \\ + V_X{}^+(\epsilon_X - E_D) + \sum_{\text{c.b.}} n_s E_s - \sum_{\text{v.b.}} p_r E_r\end{aligned} \quad (14)$$

where the first factor of each term is the number of one kind of defect and the second factor is an excess Gibbs free energy for the creation of that defect. Some of these such as E_A, E_D, E_s, and E_r may be referred to as energy levels *but are strictly speaking, excess Gibbs free energies*. For the perfect stoichiometric compound $\hat{G}^* = Ng_0 = G^0$, i.e. g_0 is the free energy per MX pair in the perfect, stoichiometric compound. The physical significance of the other parameters in Eq. (14) can be determined by calculating the change in $\hat{G}^*$ for changes in the defect numbers consistent with the constraints of Eq. (5)–(7). Thus if one electron is promoted from an energy E_r in the valence band to an energy, E_s in the conduction band, or equivalently, if an electron at energy, E_s, and a hole at energy, E_r, are created, the change in $\hat{G}^*$ is $E_s - E_r$.

Gibbs Free Energy in Explicit Form and Equilibrium Conditions

Equation (14) for the excess Gibbs free energy and Eq. (8) for the configurational entropy with N_s and N_r given by Eqs. (11) and (12) can now be substituted into Eq. (4). Following the usual procedure we approximate the sum in Eq. (4) by its largest term and apply Stirling's approximation to the factorials in the binomial coefficients (*4*). The resultant equation for the Gibbs free energy is given by:

$$\begin{aligned}G^* = N\Bigg\{ & g_0 - kT\ln\frac{N^2}{(N - V_M)(N - V_X)} \\ & - \sum_{\text{c.b.}} J_1 kT\ln\frac{N_s}{N_s - n_s} - \sum_{\text{v.b.}} H_1 kT\ln\frac{N_r}{N_r - p_r}\Bigg\} \\ & + V_M{}^0\Bigg\{\epsilon_M - kT\ln\frac{N - V_M}{V_M{}^0} - kT\ln 2 \\ & - \sum J_2 kT\ln\frac{N_s}{N_s - n_s} - \sum H_2 kT\ln\frac{N_r}{N_r - p_r}\Bigg\}\end{aligned}$$

$$+ V_M^- \left\{ \epsilon_M + E_A - kT \ln \frac{N - V_M}{V_M^-} \right.$$

$$\left. - \sum J_2 kT \ln \frac{N_s}{N_s - n_s} - \sum H_2 kT \ln \frac{N_r}{N_r - p_r} \right\} \tag{15}$$

$$+ V_X^0 \left\{ \epsilon_X - kT \ln \frac{N - V_X}{V_X^0} - kT \ln 2 \right.$$

$$\left. - \sum J_3 kT \ln \frac{N_s}{N_s - n_s} - \sum H_3 kT \ln \frac{N_r}{N_r - p_r} \right\}$$

$$+ V_X^+ \left\{ \epsilon_X - E_D - kT \ln \frac{N - V_X}{V_X^+} \right.$$

$$\left. - \sum J_3 kT \ln \frac{N_s}{N_s - n_s} - \sum H_3 kT \ln \frac{N_r}{N_r - p_r} \right\}$$

$$+ \sum n_s \left\{ E_s - kT \ln \frac{N_s - n_s}{n_s} \right\}$$

$$- \sum p_r \left\{ E_r + kT \ln \frac{N_r - p_r}{p_r} \right\}.$$

We see that the defect variables are the factors multiplying each quantity in braces. The equilibrium distributions for these defect variables are obtained by minimizing G^* with respect to each variable subject to the constraints of Eqs. (5)–(7) and constant T and P. We note that since G^* is a homogeneous function of the first degree in the defect variables, its partial derivatives with respect to these variables can be read directly from Eq. (15) as the quantities in braces (*11*). It is seen that the summation terms arising from considering the densities of states, N_s and N_r, as a function of N, V_M, and V_N are (1) the same for the partial derivative of G^* with respect to V_M^0 as they are for the derivative with respect to V_M^-, (2) are identical in the same sense for V_X^0 and V_X^+, and (3) appear only through N_s and N_r in the partial derivatives with respect to n_s or p_r. As a consequence we shall see that the equilibrium values of the ratios V_M^-/V_M^0 and V_X^+/V_X^0 are not affected by the assumed dependencies of N_s and N_r on N, V_M and V_N made in Eqs. (10) and (11). In addition the equilibrium distribution functions for n_s and p_r have the same formal appearance as they would had these dependencies been neglected.

At constant T, P, N_M, and N_X the equilibrium conditions are given in terms of $G^*(N, V_M^0, V_M^-, V_X^0, V_X^+, \ldots, n_s, \ldots, p_r \ldots)$ of Eq. (15) by the seven equations:

$$\frac{\partial G^*}{\partial N} - \mu_M - \mu_X = 0 \tag{16}$$

$$\frac{\partial G^*}{\partial V_M^0} + \mu_M = 0 \tag{17}$$

$$\frac{\partial G^*}{\partial V_M^-} + \mu_M - E_f = 0 \tag{18}$$

$$\frac{\partial G^*}{\partial V_X^0} + \mu_X = 0 \tag{19}$$

$$\frac{\partial G^*}{\partial V_X^+} + \mu_X + E_f = 0 \tag{20}$$

$$\frac{\partial G^*}{\partial n_s} - E_f = 0 \qquad \mathrm{s} = 1, 2, \ldots \tag{21}$$

$$\frac{\partial G^*}{\partial p_r} + E_f = 0 \qquad \mathrm{r} = 1, 2, \ldots. \tag{22}$$

The chemical potentials μ_M and μ_X and the Fermi level, E_f, enter the equations as Langrangian multipliers arising from the constraints imposed by Eqs. (5)–(7). We wish to emphasize that the total number of sites in each sublattice, N, is a variable and the Gibbs free energy has been minimized with respects to variations in N in Eq. (16). That this should be done was indicated by Wagner (*2*) but as far as the author is aware, Libowitz and Lightstone (*3*) were the first to give a detailed statistical mechanical analysis in which this fact was recognized. Previous developments neglecting this fact suffer in losing one equation that serves to characterize the equilibrium situation more completely.

Distribution of Electrons and Holes

Equations (21) and (22) can be solved to give the distribution functions for electrons and holes as:

$$n_s = N_s[1 + \exp(E_s - E_f)/kT]^{-1} \tag{23}$$

$$p_r = N_r[1 + \exp(E_f - E_r)/kT]^{-1} \tag{24}$$

These are the general Fermi–Dirac distribution functions (*10*). The total number of conduction band electrons or valence band holes is obtained by summing Eqs. (23) or (24) over all the energies in the appropriate band to give

$$n \equiv \sum_{\text{c.b.}} n_s = \sum_{\text{c.b.}} N_s[1 + \exp(E_s - E_f)/kT]^{-1} \tag{25}$$

$$= \int_{\text{c.b.}} \frac{N_c dE}{1 + \exp(E - E_f)/kT}$$

$$p \equiv \sum_{\text{v.b.}} p_r = \sum_{\text{v.b.}} N_r[1 + \exp(E_f - E_r)/kT]^{-1}$$

$$= \int_{\text{v.b.}} \frac{N_v dE}{1 + \exp(E_f - E)/kT} \tag{26}$$

In going from the summation to the integrals it must be remembered that N_c and N_v are both to be considered functions of energy as well as of N, V_M, and V_X and the usual assumption has been made that the band levels are closely spaced enough in energy that the summations can be replaced by integrals. The commonly used Fermi integrals of order 1/2 are obtained from the integral forms of Eqs. (25) and (26) if the densities of states in each band vary as $(E - E_0)^{1/2}$, where E_0 is the energy of the band edge (12) and if V_M/N and V_X/N are small.

For nondegeneracy the Fermi level must lie within the energy band gap and at least a few kT away from either band edge so that:

$$1 + \exp(E_{s0} - E_f)/kT \approx \exp(E_{s0} - E_f)/kT \tag{27a}$$

and

$$1 + \exp(E_f - E_{r0})/kT \approx \exp(E_g - E_{r0})/kT \tag{27b}$$

where E_{s0} is the lower-most energy level of the conduction band and E_{r0} is the upper-most energy level of the valence band. Equations (25) and (26) for the numbers of electrons and holes then reduce to

$$n = [\exp(E_f/kT)][\sum_{\text{c.b.}} N_s \exp(-E_s/kT)] \tag{28}$$

$$p = [\exp(-E_f/kT)][\sum_{\text{v.b.}} N_r \exp(E_r/kT)] \tag{29}$$

so that

$$np = [\sum_{\text{c.b.}} N_s \exp(-E_s/kT)][\sum_{\text{v.b.}} N_r \exp(E_r/kT)] \tag{30}$$

Broad Band Semiconductors

For some semiconducting compounds such as the III–V, II–VI, and IV–VI compounds, the conduction and valence bands are of the order of a few electron volts wide. These so-called broad band semiconductors characteristically have densities of states *per unit volume* and unit energy that are small compared to the atomic density. It is therefore generally true that when a broad band semiconductor is nondegenerate, the fractions of sites vacant are also small; V_M/N, $V_X/N \ll 1$. If the densities of states per unit energy N_s and N_r, are converted to densities of states per unit volume and unit energy by dividing Eqs. (11) and (12) by the total volume, we expect that to a good approximation the new densities of states will be functions of energy and number of lattice sites per unit volume only. Moreover, the latter concentration will vary negligibly over the composition range in which the semiconductor is nondegenerate. Equation 30 can then be rewritten in terms of concentrations (number per cm^3) as

$$[n][p] = [n_i]^2 \tag{31}$$

where the intrinsic carrier concentration, $[n_i]$, is a function of temperature only for a given semiconducting compound.

Narrow Band Semiconductors

Characteristically, narrow band semiconductors have relatively large densities of states so that at relatively large vacancy fractions the electron-hole distribution may still be nondegenerate (*13*). For semiconducting, oxygen-rich, transition metal oxides, MO, a commonly used description is that for every M-vacancy or for every (M-vacancy, M-vacancy, M interstitial) complex there are two M^{+3} ions randomly distributed among the sites occupied by M-ions. In the terminology used in this paper one would say the acceptors are all doubly ionized and the valence band is confined to a single energy, E_v, with a density of states, $N_r = N - V_M$, at E_v. Thus taking the total number of sites N, to be equivalent to a concentration of about 10^{22} cm^{-3}, the hole concentration could be as large about 10^{21} cm^{-3} without the occurrence of full degeneracy. This can be seen using Eq. (26) in which the sum reduces to a single term with $N_r = N - V_M$ and $E_r = E_v$. In terms of the band picture one ascribes large densities of states to the presence of narrow *d*-bands.

Remaining Distribution Functions for the Special Case of Nondegeneracy

The distribution functions for the other defect variables obtained from Eqs. (16)–(20) all contain terms similar to

$$T = -\sum_{\text{c.b.}} J_1 kT \ln \frac{N_s}{N_s - n_s} - \sum_{\text{v.b.}} H_1 kT \ln \frac{N_r}{N_r - p_r}. \tag{32}$$

The specific terms shown arise in Eq. (16) for the minimization of G with respect to N. For the other defect variables the only change is in the subscript of the J_l, H_l functions. Using Eq. (23) and (24) for the distributions of electrons and holes, the terms above in Eq. (32) can be written as

$$T = -\sum_{\text{c.b.}} J_1 kT \ln [1 + \exp(E_f - E_s)/kT] - \sum_{\text{v.b.}} H_1 kT \ln [1 + \exp(E_r - E_f) kT]. \tag{33}$$

When the semiconductor is nondegenerate as defined by Eq. (27a,b) the exponentials in each term of Eq. (33) go to zero and so then does each term. This is also true of similar terms arising in the distribution functions for $V_M{}^0$, $V_M{}^-$, $V_X{}^0$, and $V_X{}^+$. In this special case the distribution functions obtained here by substitution of Eq. (15) into Eqs. (16)–(20) are the same as those obtained on the assumption that the densities of states N_s and N_r depended only on energy. These distribution functions are:

$$\mu_M + \mu_N = g_0 - kT\ln\frac{N^2}{(N - V_M)(N - V_X)} \qquad (34)$$

$$\mu_M = kT\ln\frac{N - V_M}{V_M{}^0} - \epsilon_M + kT\ln 2 \qquad (35)$$

$$= kT\ln\frac{N - V_M}{V_M{}^-} - \epsilon_M + (E_f - E_A) \qquad (36)$$

$$\mu_X = kT\ln\frac{N - V_X}{V_X{}^0} - \epsilon_X + kT\ln 2 \qquad (37)$$

$$= kT\ln\frac{N - V_X}{V_X{}^+} - \epsilon_X + (E_D - E_f). \qquad (38)$$

Equating Eqs. 35 and 36 and Eqs. 37 and 38 gives respectively:

$$\ln 2V_M{}^-/V_M{}^0 = (E_f - E_A)/kT \qquad (39)$$

$$\ln 2V_X{}^+/V_X{}^0 = (E_D - E_f)/kT \qquad (40)$$

Along with Eqs. (23) and (24), Eqs. (39) and (40) complete the description of the equilibrium distribution of electrons and holes over the electronic energy level structure. As mentioned earlier, Eqs. (39) and (40) as well as Eqs. (23) and (24) for n_s and p_r are still valid even if the semiconductor is degenerate, whereas Eqs. (34)–(38) for the chemical potentials must be corrected by addition of terms of the general form given by Eq. (32).

Physically, Eq. (34) can be viewed as a justification in terms of our model of an approximation commonly used, but not always recognized, in thermochemistry. Provided the vacancy concentrations are small enough so that V_M/N, $V_X/N \ll 1$ and the vapor is ideal, then Eq. 34 leads to

$$p_{M_a}^{1/a}\,p_{X_b}^{1/b} = K(T) \qquad (41)$$

where K is a function of temperature only, p_M is the partial pressure of M_a molecules (a vapor phase molecule composed of a atoms all of the element M) and p_{X_b} is the partial pressure of X_b molecules. It can be seen directly from Eq. (34) with $N - V_M = N - V_X = N$ and Eq. (41) that $\ln K(T)$ is equal to $\Delta G_f{}^0/RT$, where $\Delta G_f{}^0$ is the standard free energy of formation of MX(c) from $M_a(g)$ and $X_b(g)$ each at 1 atm.

Products other than np are functions of temperature only for a nondegenerate semiconductor. Adding Eq. (35), giving μ_M in terms of $V_M{}^0$, and Eq. 37 giving μ_X in terms of $V_X{}^0$, equating the result of Eq. (34), and rearranging we obtain

$$V_M{}^0\,V_X{}^0/N^2 = (\tfrac{1}{4})\exp\left[-(g_0 + \epsilon_M + \epsilon_X)/kT\right] = k_0(T,P). \qquad (42a)$$

Similarly adding Eqs. (36) and (37) and equating to Eq. (34), the Fermi level cancels out and we obtain

$$V_M{}^-\,V_X{}^+/N^2 = \exp\left[-(g_0 + \epsilon_M + \epsilon_X + E_A - E_D)/kT\right] = k_1(T,P). \qquad (42b)$$

Thus so-called Schottky constants are defined for each product. Strictly speaking it is the product of the vacancy fractions, e.g. $(V_M{}^0/N)(V_X{}^0/N)$, that is a function of temperature only. However, if the vacancy concentrations are small enough the number of sites per cm^3 does not vary significantly with defect concentrations, one can replace the vacancy fractions by concentrations. This substitution of course can be made throughout the equations used here.

Equations (36) and (38) for the chemical potentials, simplified as they have been by the assumption of nondegeneracy, are still more complicated than their analogs for a model in which electronic defects are neglected. Equations (35) and (37) in terms of the neutral or un-ionized vacancies have almost the same forms as for the simpler model, but only reduce to them when the fraction of neutral M- or X-vacancies is unity, i.e. when there are no ionized vacancies. This is the case when (1) either the M- or X-vacancy concentration is much larger than the other and (2) the ionization energies for the donor and acceptor levels are large compared to kT. That Condition 2 is not enough, can be seen by considering the stoichiometric compound in which the total concentration of M-vacancies must equal the total concentration of X-vacancies. If as is usually the case, the M-vacancy acceptor level is much below the X-vacancy donor level in energy, the electrons will fall from the donor levels into the acceptor levels and all the vacancies will be ionized. Thus when a single type of vacancy is in excess and is essentially neutral because of a large ionization energy of the associated donor or acceptor level, the equation for the corresponding chemical potential reduces to that obtained when electronic defects are neglected.

The chemical potentials can be written in a highly compact form in terms of the net electron concentration, $n-p$, provided (1) the semiconductor is nondegenerate and (2) that V_M/N, $V_X/N \ll 1$ (so that $N - V_M \approx N$; $N - V_X \approx N$). As mentioned earlier, the latter restriction is less confining in general than the former for broad band semiconductors. The variable, $n-p$, is particularly appropriate for broad band semiconductors since Hall measurements are generally easily made and, once enough is known concerning the band structure, n and p can be calculated. Using Eqs. (7), (31), and (42b) we can rewrite Eqs. (36) and (38) as

$$\begin{aligned}\mu_M = {} & kT\sinh^{-1}[(n-p)/2k_1^{1/2}] \\ & + kT\sinh^{-1}[(n-p)/2n_i] \\ & + \mu_M(\text{int})\end{aligned} \tag{43}$$

$$\mu_X = -\mu_M + \mu_M(\text{int}) + \mu_X(\text{int}) \tag{44}$$

where $\mu_M(\text{int})$ and $\mu_X(\text{int})$ are respectively the chemical potential of M and of X when $n-p=0$ i.e. when the semiconductor is intrinsic. It should be emphasized that the validity of Eqs. (43) and (44) is independent of the fraction of M-vacancies or X-vacancies ionized. The only restrictions are of small vacancy concentrations so that $N - V_M \approx N - V_X \approx N$ and nondegeneracy. Within these restrictions Eqs. (43) and (44) are equally valid near stoichiometry when $V_M = V_X$ and "far" from stoichiometry when one vacancy or the other is in excess. The important points in connection with Eqs. (43) and (44) are that the energy parameters of the model appear in the temperature dependent quantities, $\mu_M(\text{int})$ and $\mu_X(\text{int})$, and at a given temperature serve only to shift μ_M and μ_X equally for all values of $n-p$. The dependence of the chemical potentials upon composition arises in the inverse hyperbolic sine terms. One arises from the configurational entropy of distributing the vacancies over sites, the other [with the argument $(n-p)/2n_i$] arises from the configurational entropy of distributing electrons and holes over the available energy levels. Which predominates depends on the relative magnitudes of k_1 and n_i^2. Generally both must be considered.

Degenerate Semiconductors and Semimetals

The expressions for the chemical potentials are different than those obtained heretofore if the semiconductor is degenerate. We have not written them down explicitly, but they are readily obtainable upon substitution of Eq. (15) into Eqs. (16)–(20). Since we know very little at present concerning the functions defining the densities of states in the bands (Eqs. (11) and (12)) not much can be said. However, it is obvious that, in contrast to the situation for nondegeneracy, the results of the statistical-mechanical analysis cannot be cast in the form of mass action laws (with activity coefficients all equal to unity). For instance the product, np, formed from Eqs. (25) and (26) can be seen to be a function of composition both through the Fermi level and the densities of states.

In some cases the model used here is too simple at the donor or acceptor concentrations necessary for the onset of degeneracy for reasons summarized in Description of the Model. However, the model may very well be appropriate for semimetals. In semimetals the valence and conduction bands just touch or overlap slightly, so the electron distribution is always degenerate. However, the density of states in one or both of the bands is small so that, like semiconductors and unlike normal metals, relatively small donor or acceptor concentrations (in the 10^{16}–10^{20} cm^{-3} range) can cause significant shifts in the Fermi level. For instance HgTe and HgSe are semimetals for which the relationship between stoichiometry and carrier concentration has been investigated experimentally (*14*, *15*, *16*). On the basis of the results presented here this relationship ought to be interpretable provided the densities of states can be determined, whereas a customary mass action law analysis is invalid because of degeneracy.

Discussion

Comparisons between the results obtained here and those for models neglecting electronic defects have been made in Eq. (10) and in the paragraph following Eq. (42b). In particular the results obtained by Libowitz and Lightstone (*3*) and those obtained here are consistent in the following sense. If their results are specialized to the case that vacancies are the predominant atomic point defects in a compound MX (they considered vacancies, interstitials, and place exchange defects in M_aX_b) and if the consideration of electronic defects is eliminated from our results by taking all vacancies to be neutral and the electron and hole concentrations to be zero, then both sets of equations are identical.

The author has previously given a statistical mechanical analysis for a semiconducting compound with vacancy defects (*17*) and has presented the results for more general defect models (*18*). These previous treatments differ from the one given here in that (1) the total number of sites was not considered as a variable, (2) the conduction and valence bands were treated as being at fixed energies with effective densities of states, (3) the densities of states in the bands were taken as independent of the defect variables, and (4) the sum of the chemical potentials, $\mu_M + \mu_X$, was assumed to be independent of composition. For $N - V_M \approx N - V_X \approx N$ we have shown here that the ad hoc assumption in (4) is identical to the additional equilibrium relation obtained by considering N a variable and for a nondegenerate semiconductor (2) and (3) introduce no error. In sum the previous results are correct for a nondegenerate semiconductor in which V_M/N, $V_X/N \ll 1$. The applications by the author of these previous results to the problems of interdiffusion (*19*) and the extent and position of homogeneity ranges (*20*), and to general thermodynamic behavior (*21*), therefore require no essential correction. One need only recognize (1) the assumption made in these previous publications that $\mu_M + \mu_X$ is independent of composition is not necessary, being a consequence of the theory when V_M/N, $V_X/N \ll 1$, and (2) the energy parameters appearing in the previous work are to be considered as excess free energies.

Since the results obtained here reduce to those obtained previously for the case of a nondegenerate semiconductor and V_M/N, $V_X/N \ll 1$, we refer to previous work (*21*, *22*) for applications of the theory to experiment.

Note added in proof. Because of partial ionic character in the crystal binding, the M-atoms will bear some positive charge and the X-atoms some equal, but negative charge in the crystal. We have implicitly assumed that an un-ionized or "neutral" vacancy bears the same charge as the atom normally occupying the vacant site. Then the electroneutrality condition given by Eq. (7) follows and is independent of what this charge is.

References

1. R. H. Fowler and E. A. Guggenheim, "Statistical Thermodynamics." The University Press, Cambridge, England (1949).

2. C. Wagner, "Thermodynamics of Alloys," p. 58. Addison-Wesley, Cambridge, Mass. (1952).

3. G. G. Libowitz and J. B. Lightstone, *J. Phys. Chem. Solids* **28**, 1145 (1967).

4. G. S. Rushbrooke, "Introduction to Statistical Mechanics." Oxford University Press (1949).

5. B. I. Halperin and M. Lax, *Phys. Rev.* **148**, 722 (1966).

6. T. N. Morgan, *Phys. Rev.* **139**, A343 (1965).

7. W. W. Harvey, *J. Phys. Chem.* **71**, 2031 (1967).

8. M. B. Panish and H. C. Casey, Jr., *J. Phys. Chem. Solids* **28**, 1673 (1967).

9. E. A. Guggenheim, "Thermodynamics." North Holland, Amsterdam (1949).

10. J. Mayer and M. G. Mayer, "Statistical Mechanics," p. 111. Wiley, New York (1940).

11. D. V. Widder, "Advanced Calculus," p. 15. Prentice-Hall, New York (1947).

12. J. Blakemore, "Semiconductor Statistics," Ch. 2, Pergamon Press, New York (1962).

13. S. Van Houten, *J. Phys. Chem. Solids* **17**, 7 (1960).

14. H. Rodot, *J. Phys. Chem. Solids* **25**, 85 (1964).

15. R. F. Brebrick and A. J. Strauss (unpublished).

16. R. F. Brebrick and A. J. Strauss, "II-VI Semiconducting Compounds," p. 425. (D. G. Thomas, Ed.). W. A. Benjamin, New York (1967).

17. R. F. Brebrick, *J. Phys. Chem. Solids* **4**, 190 (1958).

18. R. F. Brebrick, *J. Appl. Phys.* **33**, 422 (1962).

19. R. F. Brebrick, *J. Appl. Phys.* **30**, 811 (1959).

20. R. F. Brebrick, *J. Phys. Chem. Solids* **18**, 116 (1961).

21. R. F. Brebrick *in* "Progress in Solid State Chemistry," Vol. 3 (H. Reiss, Ed.). Pergamon Press, New York (1966).

22. F. A. Kroger, "The Chemistry of Imperfect Crystals." North Holland, Amsterdam (1964).