Self-compensation in semiconductors

Y. Tsur* and I. Riess
Technion, Israel Institute of Technology, 32000 Haifa, Israel
(Received 20 January 1998; revised manuscript received 4 May 1999)

The problem of self-compensation of charged dopants is analyzed. Special emphasis is given to dopants in binary oxides. It is shown that one can determine the degree of self-compensation from the properties of the host material and dopant concentration alone. It is further shown that for a native *p*-type semiconductor, donors are compensated, mostly, by native ionic defects. On the other hand, doping with acceptors allows us to increase significantly the hole concentration, i.e., self-compensation is low under high doping levels. For a native *n*-type semiconductor the opposite is true, namely, extrinsic acceptors are mainly compensated by native ionic defects. It is shown that the changes in concentration of all the charged defects are simply related by a single factor, the doping factor *f*, or its power f^k where *k* depends solely on the defect's charge. Quantitative calculations of *f* and defect concentrations are presented for Cu_2O, which was used as a model material. It is found that for *p*-type Cu_2O doping with donors results in *f* within the range of 1–10, depending on the dopant concentration and $P(O_2)$. This means that the hole concentration decreases and the electron concentration increases at most by a factor of 10. Therefore one does not expect to obtain a changeover from *p*- to *n*-type cuprous oxide by doping, under equilibrium conditions. Most of the donors are compensated by negative ionic defects. Self-compensation in the presence of amphoteric defects and Fermi level stabilization are discussed, using the former formalism. [S0163-1829(99)11335-3]

I. INTRODUCTION

Doping semiconductors is of key importance in modern technology. In many cases the theoretical calculation of the charged defect concentrations involved is simple.[1] In such cases the concentration of, e.g., electrons in an extrinsic *n*-type semiconductor equals to a good approximation the difference between donor and acceptor concentrations.

However, when dealing with wide-gap ionic semiconductors the question of self-compensation arises. Charged impurities may be compensated not only by a change in electron and hole concentrations but also by a change in ionic defect concentrations. According to Agrinskaya and Mashovets,[2] self-compensation is the deviation of the ratio n/N_D (or p/N_A) from unity, where n and p are the concentrations of electrons and holes, and N_A and N_D are the equivalent concentrations (concentration times valence) of charged, extrinsic acceptors and donors, respectively. For a more general case, we define self-compensation (η) as

$$\eta \equiv 1 - \frac{n - p - (n_0 - p_0)}{N_D - N_A}, \tag{1}$$

where the subscript 0 denotes zero concentration of dopants. This definition also takes into consideration the electron and hole concentrations (n_0 and p_0) present in the undoped solid, and that a change in both n and p is not considered as self-compensation. In compounds, n_0 and p_0 originate from deviation from stoichiometry and thermal excitation.

The parameters of interest are the concentrations of the electronic charge carriers n or p and their dependence on the dopant concentration. Obviously for a fixed temperature and chemical potential of the components (e.g., oxygen partial pressure for a binary oxide), the concentrations n, p as well as the concentration of all other native defects are functions of the dopant concentration. However, this dependence is not readily apparent.

There is no single answer to the question of self-compensation for a given doping level. As we shall see, the answer depends on the nature of the native point defects, their relative concentration, and on the nature of the dopant, whether single valued or not. We shall first present a method for calculating the concentrations of the native point defects as a function of dopant concentration. We shall then apply the method to analyze self-compensation under various conditions.

All approaches, as well as ours, assume that one can identify in the material native and extrinsic ionic defects in the sense that the defects have integer, well defined charge. Mandel's[3] approach is to guess the main compensating specie, then estimate the reaction constants and try to fit this to the experimental results. Jansen and Sankey[4] have found the trends in the native defect concentrations using a set of ten independent reactions and two constraints to evaluate 12 unknowns. This method, while formally correct, has two main drawbacks. The first one, as the authors have pointed out, is that the results are extremely sensitive to small changes in the numerical values of the reaction constants. In addition to that, it is difficult to get a better understanding of the system after applying this kind of calculations. Faschinger, Ferreira, and Sitter[5,6] used the fact that the Fermi level seems to be pinned at a rather fixed absolute value for groups of compounds. Starting from this approximate assumption they then discuss self-compensation.

The calculation of the native point defect concentrations as a function of dopant concentration can be handled by the well known method of dilute point defect chemistry.[7] Doing so, Maier[8] has argued that the effect of doping can be inferred from the neutrality condition. The trend is then that

 ©1999 The American Physical Society

effective positive dopants suppress other positive native defect concentrations and enhance negative native defect concentrations. Teramoto[9] and Neumark[10] used point defect chemistry to discuss the effect of doping with amphoteric impurities on the electron and hole concentrations.

It is the purpose of this work to analyze the problem of self-compensation of charged dopants. We shall discuss the changes in native defect concentrations with doping, following the method of point defect chemistry.[7] We shall also present a simple way of obtaining the changes in these concentrations with doping. This analysis refers to systems with dilute concentrations of point defects. Our approach is different from those of Mandel,[3] Jansen and Sankey,[4] and Faschinger, Ferreira, and Sitter.[5,6] We do not limit the discussion to one compensating specie at a time as Mandel does.[3] We do not look for explicit data of the defect energy levels and for reaction constants as done before.[2,3,9,10] Instead we show that the necessary information is available from the knowledge of the concentrations of the native defects in the undoped, host material. Since these can in many cases be determined quite accurately, e.g., by a coulometric titration method (see, for instance, Ref. 11) the experimental data needed are available. This takes the relevant energy levels and reaction constants accurately into consideration. We then show that the effect of doping on native defect concentrations can be presented by a simple power law of a single factor (called doping factor, f). Self-compensation is then expressed in terms of f. It is shown that with the aid of f one can find analytic expression for the self-compensation. Limiting values for the self-compensation can also easily be found.

II. METHOD OF CALCULATING THE CONCENTRATIONS OF NATIVE DEFECTS AS A FUNCTION OF DOPANT CONCENTRATION

A. General

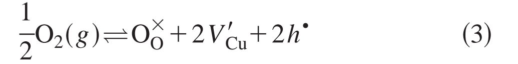
Let us consider native point defects in an ionic semiconductor. To be specific let us consider Cu_2O as an example, and let us treat it as if it were a purely ionic solid (in the sense that all defects have integer elementary charges). We start with considering native defects in the undoped host.

Cu_2O is a p-type semiconductor. There is considerable disagreement between the various works on the defect concentration and their nature in Cu_2O. We shall rely on results of recent work done in our laboratory.[11,12] The dominant point defects are neutral ones, $Cu_i^{\times}$ and $V_{Cu}^{\times}$. (We use here the Kröger-Vink notation for point defects[13] indicating the site and the relative charge of the defect.) Their concentration is approximately 0.3 cation % at 1200 K. The charged defects are in the 0.04 cation % range or lower at 1200 K. The dominant charged point defects are holes ($h^{\bullet}$), copper vacancies (V_{Cu}'), and oxygen interstitials (O_i''). Their relative concentrations depend on T and $P(O_2)$.

The neutrality equation for the undoped material is

$$p_0=[V_{Cu}']_0+2[O_i'']_0, \tag{2}$$

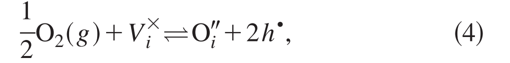
where 0 denotes, as before, zero concentration of dopants. The formation reactions of the native defects along with the corresponding mass action relations are



$$\frac{1}{2}O_2(g)\rightleftharpoons O_O^{\times}+2V_{Cu}'+2h^{\bullet} \tag{3}$$

(where $O_O^{\times}$ is an oxygen on an oxygen site),

$$p^2[V_{Cu}']^2=(K_1)^2P(O_2)^{1/2}, \tag{3$'$}$$



$$\frac{1}{2}O_2(g)+V_i^{\times}\rightleftharpoons O_i''+2h^{\bullet}, \tag{4}$$

$$p^2[O_i'']=K_2P(O_2)^{1/2}. \tag{4$'$}$$

Equations (3), (3′), (4), and (4′) hold for both the undoped material (then $p=p_0$, $[O_i'']=[O_i'']_0$, and $[V_{Cu}']=[V_{Cu}']_0$) and for the doped material.

The reaction constants K_1 and K_2 can be expressed in terms of the native defect concentrations at a given $P(O_2)$ and T, within the existence range of the undoped Cu_2O phase. We shall denote the corresponding values of $P(O_2)$, T, and native defect concentrations with the index m. The point m can be chosen arbitrarily within the existence range of the phase. From Eqs. (3′), (4′):

$$K_1(T_m)=\frac{p_{0,m}[V_{Cu}']_{0,m}}{P(O_2)_m^{1/4}}, \quad K_2(T_m)=\frac{p_{0,m}^2[O_i'']_{0,m}}{P(O_2)_m^{1/2}}. \tag{5}$$

This allows one to eliminate the reaction constants from the concentration equations (3′) and (4′) and to express the concentrations of the charged native defects in the undoped material for a given T_m in terms of $P(O_2)/P(O_2)_m$ and experimental values $p_{0,m}$, $[V_{Cu}']_{0,m}$, and $[O_i'']_{0,m}$. See Appendix A. Thus for the hole concentration Eq. (A1) yields

$$x_0^3-\frac{V_{0,m}\varphi}{V_{0,m}+1}x_0-\frac{\varphi^2}{V_{0,m}+1}=0, \tag{6}$$

with $x_0\equiv p_0/p_{0,m}$, $\varphi\equiv[P(O_2)/P(O_2)_m]^{1/4}$, and $V_{0,m}\equiv[V_{Cu}']_{0,m}/2[O_i'']_{0,m}$. This is a cubic equation for p_0 as a function of $P(O_2)$ and T, since $V_{0,m}$ is a function of T.

Introducing donors into the material must result in the reduction of the hole concentration and the enhancement of electron concentration. We therefore may write the new electron concentration as

$$n=fn_0, \tag{7}$$

with $f>1$ called the doping factor.

For Cu_2O, as well as for most semiconductors,

$$np=n_i^2(T), \tag{8}$$

which is simply the mass action relation for the creation and annihilation of an electron hole pair. As n is enhanced on doping by a factor $f>1$ the new concentration of holes, p, is reduced by the same factor f:

$$p=\frac{p_0}{f}. \tag{9}$$

In the same manner Eqs. (3′), (4′) yield for given T and $P(O_2)$ the new values of $[O_i'']$ and $[V_{Cu}']$ for the doped material:

$$[V'_{Cu}] = f[V'_{Cu}]_0, \tag{10}$$

$$[O''_i] = f^2[O''_i]_0. \tag{11}$$

The concentration of every point defect (all assumed to be dilute) changes by a simple power law, f^k, of one and the same factor f, where k is an integer. k is fixed by the charge. $k>0$ for defects having charge opposite to that of the dopant.

An immediate consequence of Eqs. (7) and (9)–(11) is that when the electron concentration is negligible with respect to other negative point defects ($[V'_{Cu}]_0, [O''_i]_0$ here) it also remains negligible under doping with donors. Of course, the electron concentration increases (by the factor f) but those of $[V'_{Cu}]$ and $[O''_i]$ increase as well. One increases by f and the other by f^2.

B. Doping a native *p*-type semiconductor with donors—change in native defect concentrations

Let us assume that D is a doubly charged donor (e.g., Co^{2+}) substituting for Cu in Cu_2O to yield $D^{\bullet}_{Cu}$. Then the neutrality equation is

$$N_D + \frac{p_0}{f} = f[V'_{Cu}]_0 + 2f^2[O''_i]_0 + fn_0, \tag{12}$$

where we have used Eqs. (7) and (9)–(11) for the native defect concentrations. (One should keep in mind that n_0 is negligible as compared to concentrations of other negative ionic defects.) The neutrality equation is a cubic equation in f. It yields the dependence of f on the dopant concentration $[D^{\bullet}_{Cu}] = N_D$, oxygen partial pressure, temperature, and the properties of the undoped material. It should be noticed that f does not depend on the nature of the dopant, only on its concentration N_D. A dependence on the nature of the dopant is introduced, in this case, only when one reaches the solubility limit.[14] Equation (12) is not affected by $D^{\times}_{Cu}$, if it exists. It would be altered if, for instance, both $D^{\bullet}_{Cu}$ and $D^{\bullet\bullet}_{Cu}$ coexist. Then f would depend on $N_D = [D^{\bullet}_{Cu}] + 2[D^{\bullet\bullet}_{Cu}]$ and thus on the nature of D, which affects the partition $[D^{\bullet}_{Cu}]/[D^{\bullet\bullet}_{Cu}]$. f would also be affected by formation of associates that contain the dopant. We limit our discussion in this section to the case that only one charged dopant defect, say $D^{\bullet}_{Cu}$, exists.

The evaluation of f in terms of the experimental values of $p_{0,m}$, $[V'_{Cu}]_{0,m}$, $[O''_i]_{0,m}$, and as a function of the degrees of freedom of the system T, $P(O_2)$, and N_D is given in Appendix B. The result is given in Eq. (B4),

$$f^3 + Gf^2 - \frac{D}{V_{0,m}^2} G^2 f - 1 - G = 0,$$

$$D \equiv \frac{N_D}{2[O''_i]_{0,m}}, \quad G \equiv \frac{V_{0,m} x_0}{\varphi}. \tag{13}$$

The equation for the (normalized) electron concentration is then Eq. (B8),

$$\tilde{n}^3 + \frac{V_{0,m}}{\varphi}\tilde{n}^2 - \frac{D}{\varphi^2}\tilde{n} - \frac{V_{0,m}+1}{\varphi^2} = 0, \quad \tilde{n} \equiv \frac{n}{n_{0,m}}. \tag{14}$$

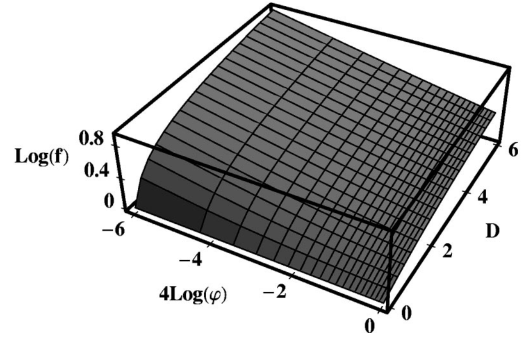


FIG. 1. f vs $P(O_2)$ and N_D for donor doping of Cu_2O at 1200 K. $\varphi = [P(O_2)/P(O_2)_{0,m}]^{1/4}$. $D = N_D/2[O''_i]_{0,m}$.

f in the polynomial, in Eq. (13), depends on parameters $[G$ and $V_{0,m}(T)]$ which do not depend on the dopant. f also depends on the parameter D which is proportional to the dopant concentration. In Fig. 1 we present the solution of Eq. (13) graphically, for the case $V_{0,m} = 0.17$. (We choose the point m at the end of the phase range, under the partial pressure for equilibrium of the undoped phases CuO/Cu_2O.) This value of $V_{0,m}$ is the relevant value for Cu_2O at 1200 K.[11] f starts from unity for $D=0$. f is a monotonic function, increasing with the donor concentration and decreasing with the oxygen partial pressure. We found experimentally[14] that the maximum value of D at ~ 1200 K is ~ 2.5, reflecting the solubility limit of Co in Cu_2O. For illustration purpose we extended the calculation up to $D=6$. However, even for these conditions and low $P(O_2)$, the doping factor f is less than one order of magnitude.

III. SELF-COMPENSATION

A. General

Self-compensation depends on the nature of the host. We shall show that this dependence can be inferred from the concentrations of the native defects in the undoped host. This will be demonstrated by discussing a few key examples of defect models. The first is a native *p*-type semiconductor doped with donors. It turns out that it is significantly self-compensated for all donor concentrations. The second is a native *p*-type semiconductor doped with acceptors. It turns out that it is only partially self-compensated, the compensation decreases with acceptor concentration. The third and fourth examples discuss amphoteric defects. Native amphoteric defects have a very different effect than extrinsic amphoteric impurities. For native amphoteric defects significant self-compensation is found. For amphoteric dopants self-compensation may vanish. In the fifth example we discuss a highly disordered compound. It may exhibit self-compensation even if the electron energy gap is small as long as the electron/hole concentration is smaller than the ionic defect concentration in the undoped material.

Another parameter of interest related to self-compensation is Fermi level pinning. Obviously, if the electron (and hole) concentration n (and p) does not vary with N_D then the Fermi level is pinned. In this case self-compensation must be high. We find Fermi level pinning, for equilibrium conditions, only in the case of extrinsic amphoteric dopants at high concentrations.

B. Self-compensation in a native *p*-type semiconductor doped with donors

Self-compensation, η, is now calculated for the above case, namely, a native *p*-type semiconductor doped with donors. η can be expressed in terms of f using Eq. (1), $N_A=0$, and $n=fn_0$,

$$\eta=1-\frac{fn_0+p_0}{N_D}\frac{f-1}{f}. \tag{15}$$

It is shown here that for a native *p*-type semiconductor, $\eta>1/2$ for all levels of donor doping within the dilute limit. Using Eq. (12) for p-Cu_2O:

$$N_D=f[V'_{\mathrm{Cu}}]_0+2f^2[\mathrm{O}''_i]_0+fn_0-\frac{p_0}{f}. \tag{16}$$

Expressing p_0 in terms of $[V'_{\mathrm{Cu}}]$ and $[\mathrm{O}''_i]$, using Eq. (2) (n_0 included) yields

$$N_D=\{(f^2-1)([V'_{\mathrm{Cu}}]_0+n_0)+(f^3-1)2[\mathrm{O}''_i]_0\}\frac{1}{f} \tag{17}$$

and

$$N_D=\frac{f-1}{f}\{(f+1)p_0+2f^2[\mathrm{O}''_i]_0\}. \tag{18}$$

Hence

$$\eta=\frac{f(1-n_0/p_0)+(2[\mathrm{O}''_i]_0/p_0)f^2}{f+1+(2[\mathrm{O}''_i]_0/p_0)f^2} \tag{19}$$

(with $f\geqslant 1$), and thus for $n_0\ll p_0$

$$\frac{1}{2}\leqslant\eta\leqslant 1 \tag{20}$$

depending on f and the ratio $2[\mathrm{O}''_i]_0/p_0$. This leads to the conclusion that a native *p*-type semiconductor (i.e., $n_0\ll p_0$) doped with donors is highly self-compensated ($\eta\geqslant 1/2$), with $\eta\to 1$ for large values of the donor concentration N_D (i.e., $f\gg 1$).

C. Doping a native *p*-type semiconductor with acceptors

1. Change in native defect concentrations

Introducing acceptors into a native *p*-type ionic semiconductor results in the enhancement of the concentration of holes. Although we did not find an acceptor for Cu_2O we shall continue to use this material as an example for the theoretical discussion. We shall assume acceptors with one negative charge as compared to the unperturbed lattice and denote them as A'. Equations (7) and (9)–(11) are being replaced now with

$$p=f_ap_0,\quad n=\frac{n_0}{f_a},\quad [V'_{\mathrm{Cu}}]=\frac{[V'_{\mathrm{Cu}}]_0}{f_a},\quad [\mathrm{O}''_i]=\frac{[\mathrm{O}''_i]_0}{f_a^2} \tag{21}$$

and the neutrality equation (12) with

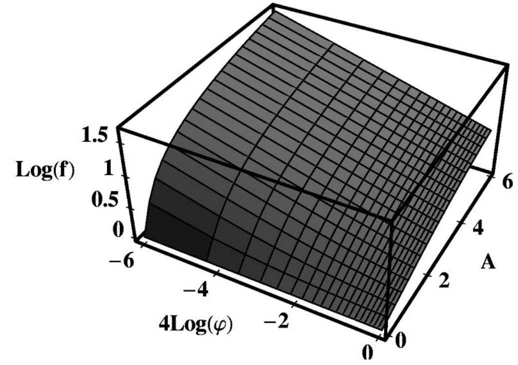


FIG. 2. f_a vs $P(O_2)$ and N_A for acceptor doping of Cu_2O at 1200 K. $\varphi=[P(O_2)/P(O_2)_{0,m}]^{1/4}$. $A=N_A/2[\mathrm{O}''_i]_{0,m}$.

$$f_ap_0=N_A+2\frac{[\mathrm{O}''_i]_0}{f_a^2}+\frac{[V'_{\mathrm{Cu}}]_0}{f_a}. \tag{22}$$

We denote f now as f_a, and we have neglected a term n_0/f_a in Eq. (22).

We further follow the analysis of Appendix B and find a cubic equation for f_a (and for $x_0f_a=p/p_{0,m}\equiv\tilde{p}$) in analogy to Eqs. (B4) and (B5):

$$(x_0f_a)^3-\frac{A}{V_{0,m}+1}(x_0f_a)^2-\frac{V_{0,m}\varphi}{V_{0,m}+1}(x_0f_a)-\frac{\varphi^2}{V_{0,m}+1}=0, \tag{23}$$

where x_0 and φ are defined as

$$x_0\equiv\frac{p_0}{p_{0,m}},\quad \varphi\equiv\left(\frac{P(O_2)}{P(O_2)_m}\right)^{1/4}. \tag{24}$$

x_0 is obtained from Eq. (6). A in Eq. (23) is defined as

$$A\equiv\frac{N_A}{2[\mathrm{O}''_i]_{0,m}}. \tag{25}$$

The solution for f_a of Eq. (23) is presented graphically in Fig. 2. Comparing f calculated for D at a given $P(O_2)$ and T with f_a calculated for the same $P(O_2)$ and T with $A=D$ one finds $f_a>f$ [except for the very dilute limit ($A,D\to 0$) where in both cases $f_a,f\to 1$]. In the heavily doped region the hole concentration is independent of the oxygen partial pressure. This takes place when N_A is greater than the native ionic defect concentrations. Note that the latter concentrations are diminished on doping with acceptors.

2. Self-compensation in a native p-type semiconductor doped with acceptors

Self-compensation on acceptor doping of a native *p*-type semiconductor is considerably different from that resulting from doping the same host with donors, discussed in Sec. III B. Starting from Eq. (1) with $N_D=0$ and $n_0\ll p_0$,

$$\eta=1-\frac{p_0(f_a-1)}{N_A}. \tag{26}$$

In analogy to Sec. III B [Eqs. (16)–(18)] the use of Eq. (22) results in

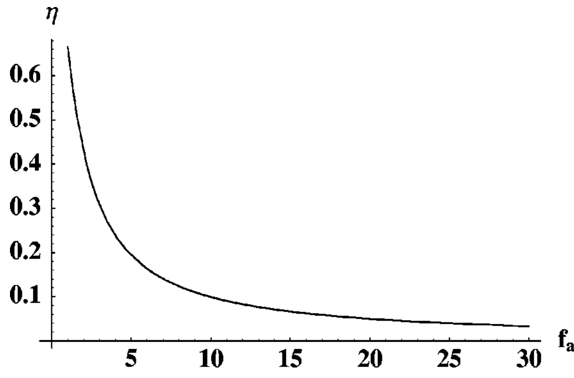


FIG. 3. η vs f_a for acceptor doping of a p-type binary oxide, with the dominant point defects O_i'' and $h^\bullet$ [i.e., the neutrality equation (2) becomes $2[O_i'']_0 \simeq p_0$].

$$N_A = \frac{f_a - 1}{f_a^2} \{ (f_a^2 + f_a) p_0 + 2[O_i'']_0 \}. \tag{27}$$

Substituting this into Eq. (26) yields

$$\eta = \frac{f_a + 2[O_i'']_0 / p_0}{f_a^2 + f_a + 2[O_i'']_0 / p_0}. \tag{28}$$

One can evaluate limits for η recalling that $0 < 2[O_i'']/p_0 < 1$ and therefore

$$\frac{1}{f_a + 1} < \eta < \frac{1}{f_a + 1/(f_a + 1)}, \tag{29}$$

with $\eta = (f_a + 1)^{-1}$ if $[V_{Cu}']_0 \simeq p_0 \gg 2[O_i'']$ and $\eta = [f_a + 1/(f_a + 1)]^{-1}$ if $2[O_i''] \simeq p_0 \gg [V_{Cu}']_0$. Figure 3 shows the latter $\eta - f_a$ relations.

This result is different from the result of Sec. III B. In the present case η is far from unity for all levels of doping. At the very dilute limit (where $f_a \sim 1$) $1/2 < \eta < 2/3$. In particular, for Cu_2O at 1200 K $\eta \sim 0.65$ at $f_a \sim 1$.

In the high doping limit, $N_A \gg p_0$, Eqs. (2), (A6), and (23) yield

$$f_a \cong \frac{A}{x_0(V_0 + 1)} = \frac{N_A}{p_0} \tag{30}$$

thus $f_a \gg 1$ and Eq. (29) yields $\eta \to 0$, i.e., negligible self-compensation. Alternatively substituting Eq. (30) into Eq. (26) yields

$$\eta(N_A \gg p_0) \cong \frac{p_0}{N_A} \ll 1. \tag{31}$$

This means that for high doping levels of acceptors η is small and the hole concentration can be enhanced significantly by acceptor doping, provided that the solubility of the dopant is high.

D. Self-compensation in the presence of native amphoteric defects as in gallium arsenide

We consider now self-compensation for the defect model discussed by Baraff and Schlüter[15] and Walukiewicz[16,17] for GaAs. For a deviation in stoichiometry with, say, excess Ga with fixed Ga chemical potential, the key ionic defects are $V_{As}^{\bullet\bullet\bullet}$ and $(Ga_{As}V_{Ga})'''$. The neutrality equation in the presence of donor $D^\bullet$ is

$$N_D + 3[V_{As}^{\bullet\bullet\bullet}] + p = 3[(Ga_{As}V_{Ga})'''] + n. \tag{32}$$

The reaction of forming the associate is

$$(Ga_{As}V_{Ga})''' \leftrightharpoons V_{As}^{\bullet\bullet\bullet} + Ga_{Ga}^{\times} + 6e' \tag{33}$$

hence

$$[(Ga_{As}V_{Ga})'''] \propto [V_{As}^{\bullet\bullet\bullet}] n^6 \tag{34}$$

Equation (32) can be written in terms of the concentrations of the native defects in the undoped material and the doping factor f as

$$N_D + 3[V_{As}^{\bullet\bullet\bullet}]_0 \frac{1}{f^3} + p_0 \frac{1}{f} = 3[(Ga_{As}V_{Ga})''']_0 f^3 + n_0 f. \tag{35}$$

Large dopant concentrations, N_D, are those of the order of the dominant native defects or higher. For large N_D values (a) there is significant self-compensation when $[(Ga_{As}V_{Ga})''']_0 f^3 > n_0 f$ which originates in part by the increase in negative native ionic defect concentration and in part by the decrease of the positive native ionic defect concentration and (b) $n - p$ increase with N_D but at a rate less than linear.

Fermi level saturation (pinning) does not occur. To see this we examine the derivative $\partial f / \partial N_D$. Fermi level saturation, i.e., no change of n (and p) on doping, requires $\partial f / \partial N_D = 0$. Differentiating Eq. (35) yields

$$\frac{\partial f}{\partial N_D} = \left(9[V_{As}^{\bullet\bullet\bullet}]_0 \frac{1}{f^4} + p_0 \frac{1}{f^2} + 9[(Ga_{As}V_{Ga})''']_0 f^2 + n_0 \right)^{-1}. \tag{36}$$

This expression is positive, showing that f increases with N_D. It vanishes only for $f \to \infty$, i.e., when the neutrality equation (35) reduces to

$$N_D \simeq 3[(Ga_{As}V_{Ga})''']_0 f^3, \quad f \gg 1 \tag{37}$$

namely, when $f \propto N_D^{1/3}$. Then $n \propto n_0 N_D^{1/3}$.

Fermi level saturation does not occur. Using the Fermi level dependence on electron concentration in the dilute case

$$\epsilon_F = \epsilon_F^0 + kT \ln n \tag{38}$$

the change in ϵ_F on doping is

$$\Delta \epsilon_F = kT \ln f. \tag{39}$$

In the limit $f \gg 1$ Eq. (37) yields

$$\Delta \epsilon_F = \frac{kT}{3} \ln \frac{N_D}{3[(Ga_{As}V_{Ga})''']_0} \tag{40}$$

and $\Delta \epsilon_F$ changes logarithmically with N_D with a coefficient $kT/3$.

The situation may be different when the amphoteric defect concentrations, $[(Ga_{As}V_{Ga})''']$ and $[V_{As}^{\bullet\bullet\bullet}]$, are not generated by thermal excitation but under nonequilibrium condi-

tions, e.g., by irradiation. This yields fixed (not necessarily equal) concentrations of $[(\mathrm{Ga_{As}}V_{\mathrm{Ga}})''']$ and $[V_{\mathrm{As}}^{\bullet\bullet\bullet}]$. The neutrality equation then yields

$$N_D = n - p - (n_0 - p_0), \tag{41}$$

i.e., increasing the dopant concentration is compensated by a change in the electron and hole concentrations.

E. Doping with amphoteric dopants, as in zinc selenide

When the dopant enters in two or more valent states all the corresponding concentrations need to be included in the neutrality equation. A particular case is when one charge state is positive and the other is negative, as in nitrogen doping of ZnSe,[18] say, with $N^{\bullet}$ and N'. For concentrations of the dopant exceeding the concentration of the native defects the neutrality equation becomes

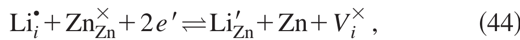
$$[N^{\bullet}] = [N']. \tag{42}$$

Large concentrations of $[N^{\bullet}]$ and $[N']$ may be generated under nonequilibrium conditions. Yet the electrons can equilibrate. As a result, n and the Fermi level are pinned, as can be seen from the mass action relation for the $N^{\bullet} + 2e' \rightleftharpoons N'$ charge transfer reaction,

$$[N^{\bullet}]n^2 = K(T)[N'], \tag{43}$$

using Eq. (42). Note that in that case the equivalent net concentration of the dopant $|[N^{\bullet}] - [N']|$ is relatively low.

Another possibility is the introduction of amphoteric impurities under thermal equilibrium conditions.[9,10] For Li in ZnSe,[10] the following reactions and corresponding mass action relations are of interest:



$$\mathrm{Li}_i^{\bullet} + \mathrm{Zn}_{\mathrm{Zn}}^{\times} + 2e' \rightleftharpoons \mathrm{Li}_{\mathrm{Zn}}' + \mathrm{Zn} + V_i^{\times}, \tag{44}$$

$$[\mathrm{Li}_i^{\bullet}]n^2 = K_3[\mathrm{Li}_{\mathrm{Zn}}'] \quad \text{for fixed Zn activity} \tag{45}$$

and the creation-annihilation reaction of electron-hole pairs with the mass action relation of Eq. (8). The neutrality equation for the undoped (p-type) ZnSe is

$$p_0 = 2[V_{\mathrm{Zn}}'']_0 + n_0 \tag{46}$$

and for the Li doped material

$$[\mathrm{Li}_i^{\bullet}] + p = [\mathrm{Li}_{\mathrm{Zn}}'] + 2[V_{\mathrm{Zn}}''] + n, \tag{47}$$

where we have also allowed for deviation from stoichiometry forming Zn vacancies.

Expressing $[\mathrm{Li}_{\mathrm{Zn}}']$ in terms of $[\mathrm{Li}_i^{\bullet}]$ using Eq. (45) and defining f by $p = fp_0$, the neutrality equation can be written as

$$N_D + fp_0 = \frac{N_D n_0^2}{K_3 f^2} + \frac{2[V_{\mathrm{Zn}}'']_0}{f^2} + \frac{n_0}{f}, \tag{48}$$

where $N_D = [\mathrm{Li}_i^{\bullet}]$. Now N_D appears in both sides of the neutrality equation and it is not obvious whether $f \geqslant 1$ or $f \leqslant 1$ (except that $f \to 1$ for $N_D \to 0$).

Differentiating Eq. (48) with respect to N_D yields

$$\frac{\partial f}{\partial N_D} = f\frac{n_0^2/K_3 f^2 - 1}{2N_D n_0^2/K_3 f^2 + fp_0 + 4[V_{\mathrm{Zn}}'']_0/f^2 + n_0/f}. \tag{49}$$

The denominator is definite positive and the sign of $\partial f/\partial N_D$ depends on the sign of the numerator. Let us examine the numerator in the limit $N_D \to 0$ $(f \to 1)$. n_0^2/K_3 can be obtained from Eq. (45),

$$\frac{n_0^2}{K_3} = \left.\frac{[\mathrm{Li}_{\mathrm{Zn}}']}{[\mathrm{Li}_i^{\bullet}]}\right|_{N_D \to 0}, \tag{50}$$

i.e., n_0^2/K_3 is the ratio of the concentrations of the amphoteric impurities in the infinite dilute limit. Depending on the characteristics of the amphoteric impurity, $n_0^2/K_3 - 1 > 0$ or $n_0^2/K_3 - 1 < 0$ and so is the corresponding sign of $\partial f/\partial N_D$. Since f changes monotonically [see Eq. (49)] it either increases monotonically above unity in the first case or decreases monotonically below unity in the second case. For the specific example of Li impurities in ZnSe,[10] $[\mathrm{Li}_i^{\bullet}] < [\mathrm{Li}_{\mathrm{Zn}}']$ for $N_D \to 0$, i.e., $n_0^2/K_3 > 1$ and $f \geqslant 1$.

Equation (49) shows that f is bound by the upper limit $f_{\max} = n_0/K_3^{1/2} > 1$. When $f \to f_{\max}$, $\partial f/\partial N_D \to 0$. Under these conditions n and the Fermi level are fixed independently of changes in N_D. Since N_D might change by orders of magnitude it is of interest to see how f becomes independent of N_D for $f \to f_{\max}$. Rearranging Eq. (48), using Eq. (46) to eliminate p_0, yields

$$N_D\left(\frac{f_{\max}^2}{f^2} - 1\right) = 2[V_{\mathrm{Zn}}'']_0\frac{f^3 - 1}{f^2} + n_0\frac{f^2 - 1}{f}. \tag{51}$$

The right hand side is definitely positive for $f > 1$. The term $(f_{\max}^2/f^2 - 1)$ vanishes as $f \to f_{\max}$, hence the other term has to diverge, $N_D \to \infty$. Under the condition $f \to f_{\max}$ the dominant terms in the neutrality equation (48) are $N_D = [\mathrm{Li}_i^{\bullet}]$ and $N_D n_0^2/K_3 f^2 = [\mathrm{Li}_{\mathrm{Zn}}']$, i.e.,

$$[\mathrm{Li}_i^{\bullet}] \simeq [\mathrm{Li}_{\mathrm{Zn}}']. \tag{52}$$

For other cases, when $n_0^2/K_3^{1/2} < 1$, f decreases towards $f_{\min} = n_0/K_3^{1/2}$. Under these conditions again $\partial f/\partial N_D \to 0$ and the amphoteric defect concentration becomes dominant.

Self-compensation can formally be considered using Eq. (1) with $N_D = [\mathrm{Li}_i^{\bullet}]$ and $N_A = [\mathrm{Li}_{\mathrm{Zn}}']$. This yields for p-ZnSe using Eqs. (46), (47),

$$\eta = \frac{(p_0 - n_0)(f + 1)}{(p_0 - n_0) + f(f + 1)p_0}. \tag{53}$$

For intrinsic ZnSe, i.e., $p_0 = n_0$, $\eta = 0$, i.e., there is no self-compensation. It should be noticed that the extrinsic defects, $\mathrm{Li}_i^{\bullet}$ and $\mathrm{Li}_{\mathrm{Zn}}'$, compensate each other as well. The net charge $N_D - N_A$ is self-compensated by V_{Zn}'' to a degree given by η of Eq. (53).

F. Self-compensation in a highly disordered compound, as in α-silver sulfide

α-Ag_2S at $T > 177\,^{\circ}C$ is highly disordered on the cation sublattice. The cation fraction of $Ag_i^{\bullet}$ and V'_{Ag} defects is larger than 0.1. The neutrality condition is

$$[Ag_i^{\bullet}]_0 = [V'_{Ag}]_0 . \tag{54}$$

Despite a low band gap[19] of 0.4 eV the electron and hole concentrations are orders of magnitude lower than $[Ag_i^{\bullet}]_0$ and $[V'_{Ag}]_0$.

On donor doping the material, the neutrality equation becomes

$$N_D + \frac{[Ag_i^{\bullet}]_0}{f} = f[V'_{Ag}]_0 . \tag{55}$$

where the concentrations fn_0 and p_0/f, the electron and hole concentrations, can still be neglected. N_D is just compensated by changes in the concentration of ionic defects. This shows self-compensation in a narrow-band-gap material. Thus the definition of large/small band gap depends on the relation of free electron/hole concentration to ionic defect concentrations. E_{gap} can be considered large as long as the electronic concentration is relatively low.

The Fermi level changes according to Eq. (39). The Fermi level in this case is not pinned as f increases monotonically with N_D.

IV. CONCLUSIONS

Self-compensation of dopants introduced into nonstoichiometric semiconductors has been analyzed. Contrary to the common practice in calculating self-compensation, we did not use explicitly the energy levels of the defects or the relevant reaction constants. Instead the method starts with the neutrality equation, and invokes the relations between defect concentrations in order to reduce the number of unknowns. This yields a polynomial equation for a factor f. The coefficients of the polynomial equation are determined by the concentrations of native defects in the undoped host and the concentration of the dopant. No explicit dependence on the chemical nature of the dopant appears, as long as the valence of the dopant defects is single. Specific detailed reference is given to p-Cu_2O, GaAs, p-ZnSe, and α-Ag_2S as examples.

The analysis assumed (a) low concentration of native point defects, i.e., they follow Boltzmann statistics; (b) detailed calculations were done assuming that the dopant is introduced in a single charge state and for amphoteric defects; formally it does not matter if the single defect is a single atomic point defect or associate of point defects; one anticipates that at elevated temperatures there will be point defects and at low temperatures associates; and (c) the system is in a state of equilibrium, except when specifically stated otherwise.

The conclusions referring to native p-type semiconductors hold for native n-type semiconductors as well, with the corresponding interchange of the role of donors and acceptors. We express therefore these conclusions for one type only, namely, native p-type semiconductors.

Our conclusions are as follows.

(1) Self-compensation and defect concentrations depend only on the nature of the host material and amount of dopant as long as the dopant is introduced in a single charge state. They can be expressed in terms of the doping factor f which is a solution of a polynomial equation.

(2) Self-compensation can also exist in narrow-band-gap semiconductors. This occurs when the concentration of native ionic defects is large compared to the concentration of electronic defects in the undoped host.

(3) The defect concentrations are enhanced or reduced on doping by a factor f^k, $k = \pm 1, \pm 2, \ldots$ (k is the charge in the Kröger-Vink notation and $k > 0$ for a charge with opposite sign as compared to that of the dopant). In the example of Cu_2O, the defect concentrations in the doped and undoped oxide are related by

$$n = fn_0, \quad [V'_{Cu}] = f[V'_{Cu}]_0, \quad [O''_i] = f^2[O''_i]_0, \quad \text{and}$$

$$p = \frac{p_0}{f}, \tag{56}$$

where 0 denotes the state of the undoped material. Equation (56) yields

$$\frac{n}{[V'_{Cu}]} = \text{const}, \quad \frac{n}{[O''_i]} \propto \frac{1}{f}, \quad \frac{[V'_{Cu}]}{p} \propto f^2,$$

$$\frac{[O''_i]}{p} \propto f^3, \quad \text{and} \quad \frac{n}{p} \propto f^2. \tag{57}$$

(4) In a native p semiconductor which is doped with donors, the dopant ions are significantly self-compensated by the native ionic defects. They are fully self-compensated for large donor concentrations. In our notation $\eta = 1$ for $f \gg 1$.

(5) Doping a native p-type semiconductor with acceptors results in enhancement of the hole concentration (and reduction of the electron concentration). For low concentrations acceptors are partially self-compensated by reducing the concentrations of negative native ionic defects. For large concentration of acceptors, self-compensation is relatively low.

(6) In a native p-type semiconductor, n is negligible with respect to the concentration of negative ionic defects. In view of conclusion (3), this relation cannot be changed by doping with donors. The electron concentration can be changed with respect to the hole concentration. The ratio n/p is increased by a factor f^2 by doping with donors: $n/p = f^2 n_0/p_0$.

The electronic conductivity depends both on the concentrations n and p and on the corresponding mobilities. Therefore the dominant conductivity may change over from p type to n type even if η is close to unity. This may occur in the case of high electron mobility as compared to that of holes. In wide-band-gap semiconductors the product np is small. Therefore in a p-type semiconductor $n \ll p$. If the material is a native p-type semiconductor then (a) the concentration of the negative native ionic defects compensating the holes is much larger than n and self-compensation of donors should occur, (b) it is not likely that the relation $n_0/p_0 \ll 1$ can be changed into $n \gg p$ by donor doping.

(7) Large changes in n or p, i.e., large values of f, require large solubility of the dopant. However, this is a problematic requirement for many of the ionic semiconductors.[14]

(8) f and η depend on the host stoichiometry. For an oxide, e.g., as for Cu_2O, one can determine the value of f as a function of the oxygen chemical potential, temperature, and dopant concentration [e.g., Eq. (13)], and equivalently find the electron or hole concentration under the same conditions [e.g., Eq. (14)]. In addition, the self-compensation η has a simple relation to f.

(9) When the dopant enters in two (or more) charge states with the same polarity (say, $D^{\bullet}_{\mathrm{Cu}}$ and $D^{\bullet\bullet}_{\mathrm{Cu}}$) an extension of the calculations is straightforward but results in more complicated equations. Self-compensation then also depends on the characteristics of the dopant via the concentration ratio (e.g., $[D^{\bullet}_{\mathrm{Cu}}]/[D^{\bullet\bullet}_{\mathrm{Cu}}]$). This also holds in the very-dilute limit. Yet, the main general trends for donors are like those encountered for a single valenced donor. One has to bear in mind that the introduction of dopants with two charge states allows for a new conduction mechanism via an impurity band. This may significantly change the electrical properties of the material.

(10) Native amphoteric defects contribute to self-compensation of the dopant. They do not lead, under equilibrium, to Fermi level pinning. Extrinsic amphoteric defects, on the other hand, yield donors and acceptors that, at high concentrations, are mutually compensating, resulting in pinning of the Fermi level. In some cases self-compensation may not occur at all, as in ZnSe with $p_0=n_0$ in the undoped material.

ACKNOWLEDGMENT

This research was supported by the Basic Research Foundation, administered by the Israel Academy of Sciences and Humanities.

APPENDIX A

Finding the concentrations of the native defects in the undoped material as a function of the oxygen partial pressure goes as follows. We first eliminate the reaction constants from the mass action relations. For the example of Cu_2O, let us express $[V'_{\mathrm{Cu}}]_0$ and $[\mathrm{O}''_i]_0$ using Eq. (5) as

$$[V'_{\mathrm{Cu}}]_0=\frac{\varphi}{x_0}[V'_{\mathrm{Cu}}]_{0,m},\quad [\mathrm{O}''_i]_0=\left(\frac{\varphi}{x_0}\right)^2[\mathrm{O}''_i]_{0,m},\quad \text{(A1)}$$

where

$$x_0\equiv\frac{p_0}{p_{0,m}},\quad \varphi\equiv\left(\frac{P(\mathrm{O}_2)}{P(\mathrm{O}_2)_m}\right)^{1/4}.\quad \text{(A2)}$$

We next write an equation for p_0 at T_m as a function $P(\mathrm{O}_2)$ in the undoped material. Substituting Eqs. (A1), (A2) into the neutrality equation (2) yields a cubic equation for x_0. This expresses x_0 (or p_0) as a function of φ [or $P(\mathrm{O}_2)$]:

$$p_{0,m}x_0=\frac{\varphi}{x_0}[V'_{\mathrm{Cu}}]_{0,m}+2\left(\frac{\varphi}{x_0}\right)^2[\mathrm{O}''_i]_{0,m}.\quad \text{(A3)}$$

$p_{0,m}$ in Eq. (A3) can be expressed in terms of $[V'_{\mathrm{Cu}}]_{0,m}$ and $[\mathrm{O}''_i]_{0,m}$ using the neutrality equation (2). This yields

$$([V'_{\mathrm{Cu}}]_{0,m}+2[\mathrm{O}''_i]_{0,m})x_0=\frac{\varphi}{x_0}[V'_{\mathrm{Cu}}]_{0,m}+2\left(\frac{\varphi}{x_0}\right)^2[\mathrm{O}''_i]_{0,m},\quad \text{(A4)}$$

which can be rewritten as

$$x_0^3-\frac{V_{0,m}\varphi}{V_{0,m}+1}x_0-\frac{\varphi^2}{V_{0,m}+1}=0,\quad \text{(A5)}$$

where

$$V_{0,m}\equiv\frac{[V'_{\mathrm{Cu}}]_{0,m}}{2[\mathrm{O}''_i]_{0,m}}\quad \text{(A6)}$$

is a concentration ratio which is determined from measurements on the undoped material[11] at T_m and $P(\mathrm{O}_2)_m$.

APPENDIX B

We demonstrate here how to evaluate f for the case of donor doped cuprous oxide, discussed in Sec. II B. Using Eq. (A1) in Eq. (12),

$$fN_D+p_{0,m}x_0=f^2\frac{\varphi}{x_0}[V'_{\mathrm{Cu}}]_{0,m}+f^3\left(\frac{\varphi}{x_0}\right)^2 2[\mathrm{O}''_i]_{0,m}.\quad \text{(B1)}$$

Substituting into Eq. (B1) $p_{0,m}$ using the neutrality equation (2) and using $V_{0,m}$ of Eq. (A6) yields

$$f^3+\frac{V_{0,m}x_0}{\varphi}f^2-\frac{Dx_0^2}{\varphi^2}f-\frac{V_{0,m}+1}{\varphi^2}x_0^3=0,\quad \text{(B2)}$$

where

$$D\equiv\frac{N_D}{2[\mathrm{O}''_i]_{0,m}}.\quad \text{(B3)}$$

We may rewrite Eq. (B2) using the relation between x_0 and f of Eq. (6) to eliminate x_0^3,

$$f^3+Gf^2-\frac{D}{V_{0,m}^2}G^2f-1-G=0,\quad G\equiv\frac{V_{0,m}x_0}{\varphi}.\quad \text{(B4)}$$

It is also straightforward to write an equation for the electron concentration under doping. Equation (B2) can be written as

$$\left(\frac{f}{x_0}\right)^3+\frac{V_{0,m}}{\varphi}\left(\frac{f}{x_0}\right)^2-\frac{D}{\varphi^2}\left(\frac{f}{x_0}\right)-\frac{V_{0,m}+1}{\varphi^2}=0.\quad \text{(B5)}$$

Equation (B5) is a cubic equation for the normalized electron concentration. This is because, using Eq. (A2),

$$\frac{f}{x_0}=\frac{f p_{0,m}}{p_0}. \tag{B6}$$

Using Eq. (8), p_0 and $p_{0,m}$ are expressed in terms of n_0 and $n_{0,m}$ which together with Eq. (7) yield

$$\frac{f}{x_0}=\frac{n}{n_{0,m}}\equiv\tilde{n}. \tag{B7}$$

Substituting Eq. (B7) into Eq. (B5) yields

$$\tilde{n}^3+\frac{V_{0,m}}{\varphi}\tilde{n}^2-\frac{D}{\varphi^2}\tilde{n}-\frac{V_{0,m}+1}{\varphi^2}=0. \tag{B8}$$

*Present address: Materials Research Laboratory, the Pennsylvania State University, University Park, PA 16802.

[1] N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Saunders College, Philadelphia, 1976).
[2] N. V. Agrinskaya and T. V. Mashovets, Fiz. Tekh. Poluprovodn. **28**, 1505 (1994) [Semiconductors **28**, 843 (1994)].
[3] G. Mandel, Phys. Rev. **134**, A1073 (1964).
[4] R. W. Jansen and O. F. Sankey, Phys. Rev. B **39**, 3192 (1989).
[5] W. Faschinger, S. Ferreira, and J. Sitter, J. Cryst. Growth **151**, 267 (1995).
[6] W. Faschinger, J. Cryst. Growth **159**, 221 (1996).
[7] F. A. Kröger, *The Chemistry of Imperfect Crystals* (North-Holland, Amsterdam, 1974).
[8] J. Maier, Angew. Chem. Int. Ed. Engl. **32**, 313 (1993).
[9] I. Teramoto, J. Phys. Chem. Solids **33**, 2089 (1972).
[10] G. F. Neumark, J. Appl. Phys. **51**, 3383 (1980).
[11] O. Porat and I. Riess, Solid State Ionics **81**, 29 (1995).
[12] O. Porat and I. Riess, Solid State Ionics **74**, 229 (1994).
[13] F. A. Kröger and H. J. Vink, Solid State Phys. **3**, 307 (1956).
[14] Y. Tsur and I. Riess, Z. Phys. Chem. (Munich) **207**, 181 (1998).
[15] G. A. Baraff and M. Schlüter, Phys. Rev. B **33**, 7346 (1986).
[16] W. Walukiewicz, J. Vac. Sci. Technol. B **5**, 1062 (1987).
[17] W. Walukiewicz, Phys. Rev. B **37**, 4760 (1988).
[18] K. Kondo, H. Okuyama, and A. Ishibashi, Appl. Phys. Lett. **64**, 3434 (1994).
[19] H. Rickert, *Electrochemistry of Solids* (Springer-Verlag, Berlin, 1982).