

General Theory of Impurity Diffusion in Semiconductors via the Vacancy Mechanism

S. M. Hu

IBM Components Division, East Fishkill Laboratory, Hopewell Junction, New York 12533

(Received 16 October 1968)

A general theory is developed for impurity diffusion in semiconductors via the vacancy mechanism, which introduces and unifies a number of new and existing concepts into a self-consistent phenomenological formalism. The thermodynamics of the vacancy-impurity-semiconductor system is first analyzed based on an energy-band model, in which the activity coefficients of the vacancy and of the impurity and the concentration of the vacancy-impurity pairs are obtained. The relationship between the diagonal and the off-diagonal phenomenological coefficients is discussed. The diagonal elements are then determined from random-walk theory, using an appropriate atomistic model. A special treatment is given to a tight-binding approximation. It is shown that the validity of Seitz's analysis of the "chemical pump effect" depends primarily on the correlation effect which he neglected. A vacancy flux from the interior, however, will be induced in the initial period as a consequence of the lowering of the vacancy activity coefficient in the donor-doped region. On the assumption of a quasiequilibrium vacancy concentration and a moderately low impurity concentration, the general theory reduces to a particularly simple form

$$J_A = -D_A^* \gamma_v^{-1} (1 + \partial \ln \gamma_A / \partial \ln N_A) \partial N_A / \partial x,$$

where J_A and N_A are, respectively, the flux and the concentration of the impurity; D_A^* is its diffusivity under intrinsic condition; and γ_v and γ_A are the activity coefficients of the vacancy and the impurity, respectively.

1. INTRODUCTION

ALTHOUGH impurity diffusion in semiconductors has long been practiced in the electronic device industry, since the work of Fuller and Dunlap,¹ a general theory on this subject is still lacking. Reiss² first treated the thermodynamics of an impurity in semiconductors using an energy-band model under non-degeneracy conditions. Longini and Greene³ extended the thermodynamic treatments to defects, such as the vacancy, in semiconductors. Since it has been accepted that certain elements diffuse in semiconductors via a vacancy mechanism, it may be expected that the thermodynamics of the vacancy will strongly affect diffusion in semiconductors. There is experimental evidence^{4,5} that indeed the diffusivity of an impurity as a function of doping concentration in the semiconductor is fairly well described by such a model. Analyses of one- and two-impurity diffusion in semiconductors based on a model of the effect of enhanced vacancy concentration and of the internal electric field have been given by Hu and Schmidt.⁶ Certain assumptions in that work, however, have not been formally justified. In this context, one may ask the following questions: How is the diffusivity related to the vacancy concentrations? How is it affected by the vacancy-

impurity association? Is the vacancy concentration maintained at its equilibrium value during the diffusion process? How about the "chemical pump effect?" Are there any interferences of the various fluxes? How are the interferences and the phenomenological coefficients affected by the correlation effect?

The purpose of this paper is to develop a general theory of diffusion in the ternary system vacancy-impurity-semiconductor in the formalism of irreversible thermodynamics, in order that the above questions may be properly answered. The thermodynamics of the vacancy-impurity-semiconductor system will first be analyzed to obtain the activity coefficients of the various constituents and the concentration of the vacancy-impurity pairs under fairly general conditions. Then the fundamental relationships between the phenomenological coefficients are discussed. It will be shown that in the particular ternary system considered, the off-diagonal elements of the phenomenological coefficients can be expressed in terms of the diagonal elements by applying appropriate constraints. General expressions for the diffusivities and the diagonal phenomenological coefficients are then derived from a given atomistic model, taking into account the correlation effect. Special discussion is then given to the case where the vacancy and the impurity are strongly associated. Furthermore, the validity of Seitz's chemical-pump effect⁷ will be examined in the light of the general theory. Finally, impurity diffusion in semiconductors will be analyzed from the simplification of the general theory based on an assumption that a quasiequilibrium vacancy concentration is maintained throughout. Vacancy transients will be treated in a separate paper.

¹ C. S. Fuller, *Phys. Rev.* **86**, 137 (1952); C. S. Fuller and J. A. Ditzenberger, *J. Appl. Phys.* **25**, 1439 (1954); **27**, 544 (1956); W. C. Dunlap, Jr., *Phys. Rev.* **94**, 1531 (1954); W. C. Dunlap, Jr., H. V. Bohm, and H. P. Mahon, Jr., *ibid.* **96**, 822 (1954).

² H. Reiss, *J. Chem. Phys.* **21**, 1209 (1953).

³ R. L. Longini and R. F. Greene, *Phys. Rev.* **102**, 992 (1956).

⁴ M. W. Valenta and C. Ramasastry, *Phys. Rev.* **106**, 73 (1957); J. Fairfield and B. J. Masters, *J. Appl. Phys.* **38**, 3148 (1967).

⁵ M. F. Millea, *J. Phys. Chem. Solids*, **27**, 315 (1966).

⁶ S. M. Hu, Electrochemical Society Meeting, Dallas, 1967 (unpublished); S. M. Hu and S. Schmidt, *J. Appl. Phys.* **39**, 4272 (1968).

⁷ F. Seitz, *Phys. Rev.* **74**, 1513 (1948); *Acta Cryst.* **3**, 355 (1950); *Acta Met.* **1**, 355 (1953); *J. Phys. Soc. Japan* **10**, 679 (1955). See, also, A. D. LeClaire, *Phil. Mag.* **3**, 921 (1958).

2. THERMODYNAMICS OF A VACANCY-IMPURITY-SEMICONDUCTOR SYSTEM

A. Chemical Potentials in a Neutral System

We begin by considering a simple, electrically neutral system consisting of three constituents A , B , and v , denoting, respectively, the impurity, the semiconductor (solvent), and the vacancy. Let N_S denote the total number of lattice sites in the ensemble, and N_A , N_B , and N_v denote the numbers of lattice sites occupied respectively by A , B , and v . Possible configurations of the system involving solute clustering will be assumed to be of negligible consequence, with the exception of vacancy-impurity pair formation. The concentration of the pair is denoted by N_{Av} . We shall restrict our discussions to a relatively dilute system, such that $N_B \gg N_A \gg N_v$. The Gibbs free energy of the system is given by

$$G_1 = G_1^0 + N_A u_A + N_v u_v - N_{Av} E_\alpha - kT \ln \Omega, \quad (2.1)$$

where G^0 is the free energy of the same lattice occupied by B only; u_A is the free energy associated with exchanging a solvent atom in the interior of the crystal with an impurity atom at the surface; u_v is the free energy of vacancy formation; E_α is the binding energy of the vacancy-impurity pair, and $k \ln \Omega$ is the entropy of configuration. We have neglected long-range interactions between the vacancy and the impurity.

For given concentrations of A , v , and B , the free energy of a volume element can be minimized by two processes, i.e., the rearrangement of A and v in the volume element, such as A - v pair formation, and the exchange of lattice constituents with neighboring volume elements. However, the latter process also affects the free energy of neighboring volume elements and cannot be considered alone. It occurs at a finite rate toward eventual heterogeneous equilibrium between the source phase and the semiconductor phase for both the impurity and the vacancy. The formation of vacancy-impurity pairs is a homogeneous process within the volume element, and can be considered as instantaneous in comparison with the slow diffusion process. Hence, the free energy of any volume element is always minimized with respect to N_{Av} , i.e.,

$$\mu_{Av} = (\partial G_1 / \partial N_{Av})_{N_A, N_v} = 0. \quad (2.2)$$

Then, following the procedure of Lidiard,⁸ one has

$$N_{Av} = N_v (N_A / N_S) \alpha / [1 + [\alpha - (Z + 1)] N_A / N_S], \quad (2.3)$$

where Z is the coordination number, and α is an association factor given by

$$\alpha = Z \exp(E_\alpha / kT). \quad (2.4)$$

The chemical potentials of the impurity and the vacancy

are, respectively, given by

$$\begin{aligned} \mu_A &= (\partial G_1 / \partial N_A)_{N_v, N_{Av}} \\ &= kT \ln(N_A / N_S) - kT \ln(1 - N_A / N_S) \end{aligned} \quad (2.5)$$

and

$$\begin{aligned} \mu_v &= (\partial G_1 / \partial N_v)_{N_A, N_{Av}} = kT \ln(N_v / N_v^0) \\ &\quad - kT \ln[1 + (N_A / N_S)(\alpha - Z - 1)], \end{aligned} \quad (2.6)$$

where N_v^0 is the equilibrium vacancy concentration at infinite dilution of A in the neutral system, and is given by

$$N_v^0 = N_S \exp(-u_v / kT). \quad (2.7)$$

The equilibrium vacancy concentration N_v^e at a given impurity concentration N_A , is, therefore,

$$N_v^e = N_v^0 [1 + (N_A / N_S)(\alpha - Z - 1)]. \quad (2.8)$$

One may identify the second terms in the right-hand sides of Eqs. (2.5) and (2.6) as activity coefficients, i.e.,

$$\gamma_A = (1 - N_A / N_S)^{-1} \quad (2.9)$$

and

$$\gamma_v = \frac{1}{1 + (N_A / N_S)[\alpha - (Z + 1)]}. \quad (2.10)$$

γ_A and γ_v reduce to unity for very small N_A .

B. Chemical Potentials in an Electrically Active System

In the following, we extend Longini and Greene's thermodynamical treatment of a vacancy-impurity-semiconductor system to the case in which there is formation of vacancy-impurity pairs. The free energy of the ensemble can be written

$$G = G_1 + G_2, \quad (2.11)$$

where G_1 is the free energy of the atomic system as given in (2.1) and G_2 is the free energy of the electronic subsystem. The number of states at the energy levels E_c , E_A , E_v , and E_p are given by N_c , N_A , N_v , and N_p , respectively, pertaining to the conduction band, the impurity centers (donor or acceptor), the vacancies (acceptor) and the valence band. We must, however, take into account the change in the energy levels in the impurity and the vacancy arising from Coulombic interaction. The range of Coulombic interaction is determined by the screening distance R , which has been given by Dingle⁹ for semiconductors and is concentration-dependent. For simplicity, however, we shall consider the interaction to operate only on the nearest neighbors, specifically the vacancy-impurity pair. We shall also neglect the screening effect and regard the energy of Coulombic interaction as a constant. For this purpose, one only has to consider the pairs in which both A and v are ionized. Assume that there are p_A

⁸ A. B. Lidiard, *Phil. Mag.* **5**, 1171 (1960).

⁹ R. B. Dingle, *Phil. Mag.* **46**, 831 (1955).

positive charges amongst the N_A impurity centers, and p_{Av} of these p_A charges are amongst A atoms that are in the N_{Av} vacancy-impurity pairs; that there are n_v negative charges amongst the N_v vacancies, and n_{Av} of these charges are amongst the vacancies that are in the N_{Av} vacancy-impurity pairs. The number of vacancy-impurity pairs in each of which a positive charge stays in A and a negative charge stays in v is denoted by n_β . From the concentrations N_{Av} , p_{Av} , n_{Av} , and n_β , one can obtain the concentrations of the four differently charged vacancy-impurity pairs $N_{A^0v^0}$, $N_{A^0v^-}$, $N_{A^+v^0}$, and $N_{A^+v^-}$ through the following relationships:

$$N_{A^+v^-} = n_\beta, \quad (2.12)$$

$$N_{A^0v^-} = n_{Av} - n_\beta, \quad (2.13)$$

$$N_{A^+v^0} = p_{Av} - n_\beta, \quad (2.14)$$

$$N_{A^0v^0} = N_{Av} - n_{Av} - p_{Av} + n_\beta. \quad (2.15)$$

By counting the numbers of ways of sequentially placing n_c electrons in the N_c conduction-band states, p_p holes in the N_p valence-band states, $(p_A - p_{Av})$ positive charges in the $(N_A - N_{Av})$ free-impurity atoms, $(n_v - n_{Av})$ negative charges the $(N_v - N_{Av})$ free vacancies, selecting n_β pairs from N_{Av} for placing both a positive and a negative charge in each, and then successively placing the rest of the $(n_{Av} - n_\beta)$ negative charges and $(p_{Av} - n_\beta)$ positive charges in the successively remaining $A-v$ pairs, and permitting combinations of spin degeneracies, one obtains the entropy of electronic configuration as

$$k \ln \Omega_e = k \ln \frac{N_c!}{n_c!(N_c - n_c)!} \frac{N_p!}{p_p!(N_p - p_p)!} \frac{(N_A - N_{Av})!}{(p_A - p_{Av})!(N_A - N_{Av} - p_A + p_{Av})!} \frac{(N_v - N_{Av})!}{(n_v - n_{Av})!(N_v - N_{Av} - n_v + n_{Av})!} \\ \times \frac{N_{Av}!}{n_\beta!(n_{Av} - n_\beta)!(p_{Av} - n_\beta)!(N_{Av} - n_{Av} - p_{Av} + n_\beta)!} + k \ln g_A^{p_A} g_v^{n_v}, \quad (2.16)$$

where g_A and g_v are the spin degeneracy factors of the A and the v levels, respectively. The Gibbs free energy of the electronic ensemble is given by

$$G_2 = n_c E_c + n_v E_v - p_A E_A - p_p E_p - n_\beta E_\beta - kT \ln \Omega_e, \quad (2.17)$$

where E_β is the energy of Coulombic interaction between a vacancy and an impurity atom. The total free energy of the system is minimized with respect to n_c , p_A , n_v , and p_p under the constraint of local-charge neutrality

$$n_c - p_A + n_v - p_p = 0. \quad (2.18)$$

using the method of Lagrange multiplier, and with respect to n_β , p_{Av} , n_{Av} , and N_{Av} , for which the constraints are implicit in the configurational entropy expressions (2.16) and (2.2)–(2.5) using simple partial differentiations. After some lengthy algebraic manipulations, one obtains

$$N_{Av} = N_v \left[\frac{N_A}{N_S} \frac{\alpha}{1+\xi} \left(1 + \frac{1+\beta\xi}{1+\zeta} \xi \right) \right] / \left[1 + \frac{N_A}{N_S} \left[\frac{\alpha}{1+\xi} \left(1 + \frac{1+\beta\xi}{1+\zeta} \xi \right) - (Z+1) \right] \right], \quad (2.19)$$

$$p_{Av} = N_v \left[\frac{N_A}{N_S} \frac{\alpha\xi}{1+\zeta} \left(\frac{1+\beta\xi}{1+\zeta} \right) \right] / \left[1 + \frac{N_A}{N_S} \left[\frac{\alpha}{1+\xi} \left(1 + \frac{1+\beta\xi}{1+\zeta} \xi \right) - (Z+1) \right] \right], \quad (2.20)$$

$$n_{Av} = N_v \left[\frac{N_A}{N_S} \frac{\alpha\xi}{1+\xi} \left(\frac{1+\beta\xi}{1+\zeta} \right) \right] / \left[1 + \frac{N_A}{N_S} \left[\frac{\alpha}{1+\xi} \left(1 + \frac{1+\beta\xi}{1+\zeta} \xi \right) - (Z+1) \right] \right], \quad (2.21)$$

$$n_\beta = N_v \left[\frac{N_A}{N_S} \frac{\alpha\beta\xi}{1+\xi} \left(\frac{\zeta}{1+\zeta} \right) \right] / \left[1 + \frac{N_A}{N_S} \left[\frac{\alpha}{1+\xi} \left(1 + \frac{1+\beta\xi}{1+\zeta} \xi \right) - (Z+1) \right] \right], \quad (2.22)$$

$$n_v = N_v \frac{\xi}{1+\xi} \left[1 + \frac{N_A}{N_S} \left[\frac{1+\beta\xi}{1+\zeta} - (Z+1) \right] \right] / \left[1 + \frac{N_A}{N_S} \left[\frac{\alpha}{1+\xi} \left(1 + \frac{1+\beta\xi}{1+\zeta} \xi \right) - (Z+1) \right] \right], \quad (2.23)$$

$$p_A = N_A \frac{\zeta}{1+\zeta} \left\{ 1 - \frac{N_v}{N_S} \frac{\alpha\xi}{1+\xi} \left(\beta - \frac{1+\beta\xi}{1+\zeta} \right) / \left[1 + \frac{N_A}{N_S} \left[\frac{\alpha}{1+\xi} \left(1 + \frac{1+\beta\xi}{1+\zeta} \xi \right) - (Z+1) \right] \right] \right\}, \quad (2.24)$$

where α has been defined in Eq. (2.4), and

$$\beta = \exp(E_\beta/kT), \quad (2.25)$$

$$\zeta = g_A \exp[(E_A - E_p)/kT], \quad (2.26)$$

$$\xi = g_v \exp[(E_F - E_v)/kT]. \quad (2.27)$$

The chemical potential of vacancy is given by the sum of partial differentiations of Eq. (2.1) and (2.17) with respect to N_v , holding other concentrations constant:

$$(\partial G_1/\partial N_v) = u_v + kT \ln \left[\frac{(N_v - N_{Av})}{(N_S - (Z+1)N_{Av})} \right], \quad (2.28)$$

$$(\partial G_2/\partial N_v) = kT \ln \left[\frac{(N_v - N_{Av} - (n_v - n_{Av}))}{(N_v - N_{Av})} \right]. \quad (2.29)$$

Note that

$$u_v = -kT \ln N_{v^0} = -kT \ln [N_v^*/(1+\xi^*)], \quad (2.30)$$

where N_{v^0} is the concentration of neutral vacancy at thermal equilibrium, N_v^* is the total vacancy concentration at equilibrium under intrinsic conditions (i.e., $N_A=0$), and ξ^* is the value of ξ under intrinsic condition. Substituting u_v from (2.30), and N_{Av} , n_v , and n_{Av} from Eqs. (2.19), (2.21), and (2.23), into Eqs. (2.28) and (2.29), one obtains

$$\begin{aligned} \mu_v &= \partial G_1/\partial N_v + \partial G_2/\partial N_v \\ &= kT \ln \frac{N_v}{N_v^*} - kT \ln \frac{1+\xi}{1+\xi^*} \\ &\quad - kT \ln \left(1 + \frac{N_A}{N_S} \left[\frac{\alpha}{1+\xi} \left(1 + \frac{1+\beta\xi}{1+\xi} \right) - (Z+1) \right] \right). \end{aligned} \quad (2.31)$$

The equilibrium vacancy concentration N_{v^0} in a semiconductor with an impurity concentration of N_A can be obtained by setting $\mu_v=0$ in the above equation:

$$N_{v^0} = N_v^* \left[\frac{(1+\xi)/(1+\xi^*)}{\left[1 + \xi(1+\beta\xi)/(1+\xi) \right] - (Z+1)} \right] \left[\frac{\alpha/(1+\xi)}{\left[1 + \xi(1+\beta\xi)/(1+\xi) \right] - (Z+1)} \right]. \quad (2.32)$$

One can identify the activity coefficient from the definition $N_{v^0} = \gamma_v^{-1} N_v^*$. At sufficiently high N_A such that $\xi \gg 1$, yet not so high that the system remains nondegenerate ($\zeta \gg 1$), the above equation reduces to

$$N_{v^0} \cong N_v^* \left[\frac{\xi/(1+\xi^*)}{\left[1 + (N_A/N_S) [\alpha\beta - (Z+1)] \right]} \right], \quad \zeta \gg 1, \quad \xi \gg 1. \quad (2.33)$$

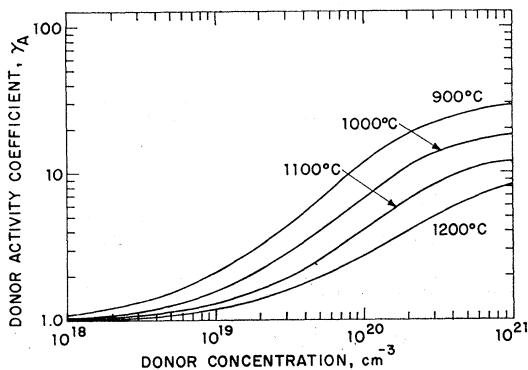


FIG. 1. Donor activity coefficient in Si calculated using Fermi statistics, on the assumption of constant donor level and constant energy-band structure.

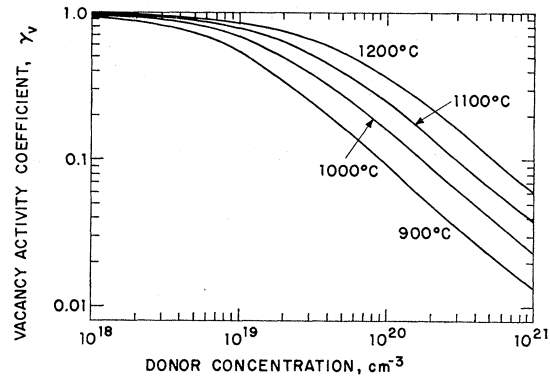


FIG. 2. Vacancy activity coefficient in Si calculated using Fermi statistics, on the assumption of constant donor and vacancy levels and constant energy-band structure.

The chemical potential of A is similarly obtained:

$$\mu_A = kT \ln \frac{N_A}{N_S} - kT \ln \left(\frac{1+\zeta}{1+\zeta^*} \right) - kT \ln \left(1 - \frac{N_A}{N_S} \right). \quad (2.34)$$

The activity coefficient can be likewise defined by the equation $\mu_A = kT \ln(\gamma_A N_A/N_S)$. The Fermi level E_F needed for the evaluation of μ_v and μ_A can be determined in the conventional manner¹⁰ by assuming N_v to be negligible.

Without specifying the impurity species, we shall assume that $g_A \exp(E_A/kT) \cong 1$. Hence, $\zeta = \exp(-E_F/kT)$. The activity coefficients of the impurity A and of the vacancy are then evaluated numerically. The results for silicon are shown in Figs. 1 and 2. In the calculation of these activity coefficients, we have, for simplicity neglected the effect of defect association, and write [c.f. Eqs. (2.32) and (2.34)]

$$\gamma_v \cong (1+\xi^*)/(1+\xi), \quad (2.35)$$

$$\gamma_A \cong (1+\zeta^*)/(1+\zeta). \quad (2.36)$$

The intrinsic carrier concentration n_i is estimated using the empirical expression of Morin and Maita.¹¹ A vacancy acceptor level of $E_c - 0.4$ eV is assumed,¹² with a spin-degeneracy factor of 2. In the calculation of N_c , we have assumed $m_e^*/m_0 = 1$, since no data appear available at diffusion temperatures. Figures 1 and 2 show the activity coefficients of donor and vacancy, respectively, under the condition that the impurity level remains discrete and constant.

It should be noted that at diffusion temperatures, taking Si as an example, the intrinsic carrier concentration is in the order of 10^{19} cm^{-3} , and the system should remain nondegenerate until the impurity concentration is much higher than 10^{19} cm^{-3} . On the other

¹⁰ For example, see J. S. Blakemore, *Semiconductor Statistics* (Pergamon Press, Inc., New York, 1962), p. 106ff.

¹¹ F. J. Morin and J. P. Maita, *Phys. Rev.* **96**, 28 (1954).

¹² G. D. Watkins and J. W. Corbett, *Discussions Faraday Soc.* **31**, 86 (1961).

hand, at impurity concentrations of $> 2 \times 10^{18} \text{ cm}^{-3}$, the average distance between impurity ions becomes smaller than the diameter of the Bohr orbit of the electron in silicon, and the electrons become completely delocalized. This leads to the lowering of the conduction band and the formation of an impurity band,¹³⁻¹⁸ and consequently the lowering of the Fermi level for a given electron concentration. One then should expect a lower rate of variation of γ_v and γ_A with N_A than would be predicted from an assumption of a constant conduction band and complete ionization using Fermi statistics. As an example, the Fermi level versus electron concentration in silicon at 1000°C is calculated by adopting Kane's model of the impurity band,¹⁵ and is shown in Fig. 3. From Fig. 3, one sees that it is not necessary to consider the formation of impurity band at diffusion temperatures, at least up to $\sim 3 \times 10^{20} \text{ cm}^{-3}$ in Si, particularly when one uses the approximation of Boltzmann statistics.

3. PHENOMENOLOGICAL EQUATIONS OF DIFFUSION

Analyses in this section will be quite general, and should be applicable to diffusions in elemental crystal-line solids including, but not limited to, semiconductors.

A. Relationships between Phenomenological Coefficients

The linear phenomenological equation of an irreversible process in matrix notation is conventionally represented by¹⁹

$$\mathbf{J} = [\mathbf{L}] \mathbf{X}, \quad (3.1)$$

where \mathbf{J} and \mathbf{X} are column vectors representing the flux and the driving force in the component space, and $[\mathbf{L}]$ is the matrix of phenomenological coefficients, with elements defined by

$$L_{ij} = (\partial J_i / \partial X_j)_{X_1, X_2, \dots, X_j = 0}. \quad (3.2)$$

In the present treatment, the thermodynamic forces X_i will be restricted to $-\partial \mu_i / \partial x$. Irreversible thermodynamics imposes certain restrictions on and prescribes certain relationships among these phenomenological coefficients. However, the determination of these coefficients must resort to atomistic modeling. Generally, one can determine the diagonal elements L_{ii} in a rather straightforward manner, for example, by

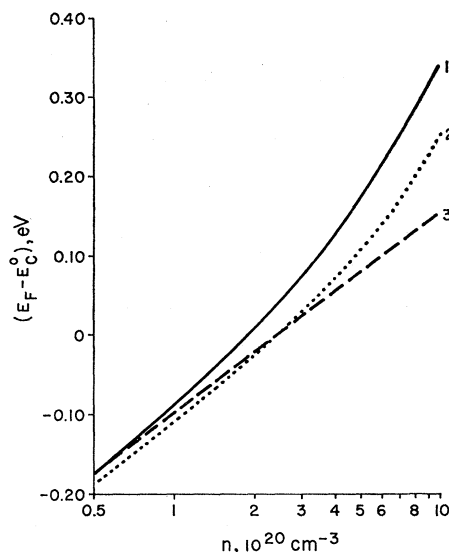


FIG. 3. Fermi level in heavily doped Si at 1000°C ($\kappa=12$, $m_{(n)}^*/m_0=1.08$). 1: Constant conduction band and complete ionization, Fermi statistics; 2: Kane's model of the impurity band; 3: Constant conduction band and Boltzmann statistics.

studying the random walk of a representative particle of the interested species, with an appropriate atomistic mechanism. It appears, however, that the evaluation of the off-diagonal elements L_{ij} ($i \neq j$) has not been discussed properly in literature. Bardeen and Herring²⁰ first pointed out that the off-diagonal elements may not be negligible even in a vacancy mechanism of diffusion, and suggested the determination of the cross-coefficients from the consideration of correlation effect. Their treatment, which rests on the assumption of local equilibrium of vacancy concentration, is appropriate only for tracer diffusion in its isotope host lattice. In this section, it is shown that the cross-coefficients in the particular system $A-B-v$ are actually determined by the diagonal elements. This is strictly a consequence of certain constraints (thermodynamic as well as atomistic) imposed on the coupled system. Thus, if one considers a representative particle of a certain species executing random walk in a fixed spatial frame, one has the flux of that species determined by its own driving force, and its associated diagonal coefficient alone. But, although the fluxes are here defined with respect to a fixed spatial frame, the trajectory of the random walk of the particle is imbedded in a nonstationary lattice as a result of the occurrence of the fluxes of various lattice constituents under certain constraints. Thus, the particle displacement in the lattice at the end of its random walk differs from its displacement in a fixed spatial frame. Hence, the off-diagonal elements of the phenomenological coefficients arise strictly as a consequence of the interaction of fluxes under these restric-

¹³ F. Stern and R. M. Talley, *Phys. Rev.* **100**, 1638 (1955).

¹⁴ P. A. Wolff, *Phys. Rev.* **126**, 405 (1962).

¹⁵ E. O. Kane, *Phys. Rev.* **131**, 79 (1963).

¹⁶ T. N. Morgan, *Phys. Rev.* **139**, A343 (1965).

¹⁷ B. I. Halperin and M. Lax, *Phys. Rev.* **148**, 722 (1966); **153**, 802 (1967).

¹⁸ V. L. Bonch-Bruyevich, *The Electronic Theory of Heavily Doped Semiconductors* (Elsevier Publishing Co., Inc., New York, 1966).

¹⁹ S. R. deGroot and P. Mazur, *Non-Equilibrium Thermodynamics* (North-Holland Publishing Co., Amsterdam, 1962), pp. 64-69.

²⁰ J. Bardeen and C. Herring, in *Atomic Movements* (American Society for Metals, Cleveland, 1951), p. 51.

tions, while the effect of the presence of various components j on the thermodynamic force of component i is properly taken care of in the function $X_i(N_i, \partial N_i/\partial x, N_j, \partial N_j/\partial x, \dots)$ or $\mu_i(N_i, N_j, \dots)$. There are two restrictions that are appropriate to the system under present consideration, i.e., the Gibbs-Duhem relation regarding the linear dependency of thermodynamic forces, and the conservation of lattice sites in any spatial element regarding the material fluxes. (There would be additional restrictions, such as the energy balance if Soret effect and Dufour effect are included in the analysis.) These are expressed as

$$\sum N_j \frac{\partial \mu_j}{\partial x} = 0, \quad (3.3)$$

$$\sum J_i = 0. \quad (3.4)$$

Because of the linear relationships (3.3) and (3.4), Eq. (3.1) in its full form is superfluous. One can eliminate one flux and one thermodynamic force (say of the n th component) from the expression, so that in the reduced form all the $(n-1)$ fluxes and the $(n-1)$ forces become independent. If one first eliminates the n th flux by using (3.4), the rate of entropy production of the diffusion process becomes¹⁹

$$\sigma = \sum_{i=1}^n J_i X_i = \sum_{i=1}^{n-1} J_i (X_i - X_n). \quad (3.5)$$

For the above entropy rate expression, one can define a new set of phenomenological coefficients l_{ij} , such that the flux expression becomes

$$J_i = \sum_{j=1}^{n-1} l_{ij} (X_j - X_n), \quad i=1, 2, \dots, n-1. \quad (3.6)$$

After replacing X_j with $-\partial \mu_j/\partial x$, and the elimination of $-\partial \mu_n/\partial x$ by the use of (3.3), one obtains

$$J_i = -\sum_{j=1}^{n-1} \left(l_{ij} + \frac{N_j}{N_n} \sum_{k=1}^{n-1} l_{ik} \right) \frac{\partial \mu_j}{\partial x}. \quad (3.7)$$

It has been shown that¹⁹ elements l_{ij} in the l_{ij} scheme obey Onsager reciprocal relation. On the other hand, in the L_{ij} scheme [i.e., Eq. (3.1) unreduced with respect to relationships (3.3) and (3.4)], there will be n -fold arbitrariness in the coefficients L_{ij} , so that L_{ij} 's do not necessarily obey the Onsager relation. However, symmetrical L_{ij} values can be obtained by disposing of the n -fold arbitrariness in an elegant way suggested by deGroot and Mazur (Ref. 19). In the l_{ij} scheme, there are only $(n-1)^2$ nonarbitrary coefficients, of which $\frac{1}{2}n(n-1)$ are distinct because of Onsager relation in the l_{ij} scheme. If one eliminates $\partial \mu_n/\partial x$ from Eq. (3.1) by means of (3.3), one obtains a set of flux expressions, each being a function of $\partial \mu_j/\partial x$, $j=1, 2, \dots, (n-1)$. These driving forces are now all independent, so that the sum of the coefficients of $\partial \mu_j/\partial x$ for all fluxes must

vanish according to (3.4). Thus, $(n-1)$ equations of the following²¹:

$$\sum_{i=1}^n \left(L_{ij} - \frac{N_j}{N_n} L_{in} \right) = 0, \quad j=1, 2, \dots, n-1, \quad (3.8)$$

must be independently fulfilled. The n -fold arbitrariness and the $(n-1)$ relations represented by Eq. (3.8) reduce the n^2 coefficients in the L_{ij} scheme also to $(n-1)^2$ sensible coefficients.

We now consider the present case of a three-component system. We assume that the diagonal elements L_{AA} , L_{BB} , and L_{vv} are obtainable from an appropriate atomistic model as to be discussed later. We can choose to eliminate J_B and $\partial \mu_B/\partial x$ in the l_{ij} scheme. A symmetrical L_{ij} scheme can then be obtained by adopting the following set of solutions (for a fuller discussion see de Groot and Mazur¹⁹):

$$L_{AA} = l_{AA}, \quad (3.9)$$

$$L_{vv} = l_{vv}, \quad (3.10)$$

$$L_{Av} = l_{Av}, \quad (3.11)$$

$$L_{vA} = l_{vA} = L_{Av}, \quad (3.12)$$

$$L_{AB} = -(l_{AA} + l_{Av}) = -(L_{AA} + L_{Av}), \quad (3.13)$$

$$L_{vB} = -(l_{vA} + l_{vv}) = -(L_{Av} + L_{vv}), \quad (3.14)$$

$$L_{BA} = L_{AB}, \quad (3.15)$$

$$L_{Bv} = L_{vB}. \quad (3.16)$$

The five equations (3.12)–(3.16) relate coefficients in the L_{ij} scheme, and imply the use of one Onsager relation, one relation from (3.8), and the disposition of threefold arbitrariness. Together with the three diagonal elements, one still needs a ninth relation. This is provided by the following, also from (3.8):

$$(L_{AA} - (N_A/N_B)L_{AB}) + (L_{vA} - (N_A/N_B)L_{vB}) + (L_{BA} - (N_A/N_B)L_{BB}) = 0. \quad (3.17)$$

Substituting (3.13) and (3.14) into (3.17), one obtains

$$L_{Av} = \frac{1}{2}(L_{BB} - L_{AA} - L_{vv}) = L_{vA}. \quad (3.18)$$

With L_{Av} given above, one can rewrite Eqs. (3.13) and (3.14) in terms of diagonal elements

$$L_{AB} = \frac{1}{2}(L_{vv} - L_{AA} - L_{BB}) = L_{BA}, \quad (3.19)$$

$$L_{vB} = \frac{1}{2}(L_{AA} - L_{BB} - L_{vv}) = L_{Bv}. \quad (3.20)$$

²¹ There is a customary argument that $\sum_i L_{ij} = 0$ is a necessary condition for the constraint $\sum_i J_i = 0$ to be fulfilled whatever the driving forces. [For example, see P. Shewmon, *Diffusion in Solids* (McGraw-Hill Book Co., New York, 1963), p. 123. Also, Ref. 20.] This is incorrect because the driving forces are linearly dependent from the Gibbs-Duhem relation. However, it is a sufficient condition, and it can be a solution by suitable disposition because of the n -fold arbitrariness in the coefficients L_{ij} . In fact, in the set of L_{ij} coefficients to be selected subsequently, in Eqs. (3.18)–(3.20), one can see that $\sum_i L_{ij} = 0$ is satisfied.

Thus, in the particular ternary system vacancy-impurity-solvent lattice, the off-diagonal elements are all fixed once the diagonal elements are determined.

It is interesting to consider the special case of a *binary system*, specifically the diffusion of vacancy in a lattice of pure element B . Since in such a system the concentration gradients of the vacancy and of the solvent B are equal and opposite in sign, while the diffusivity of the vacancy is orders of magnitude higher than that of B , a conservation of lattice sites would not be possible were it not for the cross term in the phenomenological formulation. The relations (3.3) and (3.4) reduce the l_{ij} scheme of such a system to only one sensible coefficient, i.e.,

$$J_v = -l_{vv}[(\partial\mu_v/\partial x) - (\partial\mu_B/\partial x)] \\ = -l_{vv}(1 + N_v/N_B)(\partial\mu_v/\partial x). \quad (3.21)$$

In the L_{ij} scheme, therefore, the diagonal elements L_{vv} and L_{BB} cannot assume independent values. For a symmetrical L_{ij} scheme, it is easily shown that

$$L_{BB} = L_{vv} = -L_{Bv} = -L_{vB}. \quad (3.22)$$

Thus, the determination of L_{vv} leaves the other coefficients completely defined. By eliminating either one driving force using the Gibbs-Duhem relation, one obtains

$$J_v = -L_{vv}(1 + N_v/N_B)(\partial\mu_v/\partial x), \quad (3.23)$$

$$J_B = -L_{BB}(1 + N_B/N_v)(\partial\mu_B/\partial x). \quad (3.24)$$

The above equations will, of course, ensure conservation of lattice sites. One can see from (3.24) that it is the cross term (i.e., $-L_{Bv}\partial\mu_v/\partial x = -L_{BB}(N_B/N_v)\partial\mu_B/\partial x$), and not the diagonal term, that actually dominates J_B . This important fact is seldom recognized. If one expresses the above equations in terms of diffusivity and concentrations (let us temporarily regard γ_v and γ_B as unity, and $L_{ii} = D_i N_i / kT$)

$$J_v = -D_v \left(1 + \frac{N_v}{N_B}\right) \frac{\partial N_v}{\partial x} \cong -D_v \frac{\partial N_v}{\partial x}, \quad (3.25)$$

$$J_B = -D_B \left(1 + \frac{N_B}{N_v}\right) \frac{\partial N_B}{\partial x} \cong -D_v \frac{\partial N_B}{\partial x}, \quad N_B \gg N_v, \quad (3.26)$$

where $D_B = D_v N_v / N_B$ [from (3.22)]. Since $\partial N_B / \partial x = -\partial N_v / \partial x$ in this system, J_B would be a negligible fraction of J_v if the cross term [represented by N_B/N_v in the parentheses in Eq. (3.26)] is neglected.

B. Determination of Phenomenological Coefficients

Next we seek to determine the phenomenological coefficients from random-walk theory, by adopting a suitable atomistic model. One may identify the relationship between the diagonal elements of the phenomenological coefficients and the diffusivities by writing Fick's first law in a *noninterfering system* in the following

modified form:

$$J_i = -D_i(1 + \partial \ln \gamma_i / \partial \ln N_i) \partial N_i / \partial x. \quad (3.27)$$

Then,

$$L_{ii} = (D_i N_i / kT). \quad (3.28)$$

One notes that D_i as defined in Eq. (3.27) will be identical with that obtained from random-walk theory which ignores the interference between atomic jumping processes in neighboring microscopic regions, which collectively constitute a volume element in the continuum phenomenological formulation. The interference is taken care of by the interference coefficients (i.e., the off-diagonal elements) in accordance with appropriate constraints, as discussed in the preceding section. This diffusivity, however, may in general be different from the experimental diffusivity in an interfering system, where the interference has not been separated out by independent measurements (such as the Kirkendall effect, etc.).

From the theory of random walk, the diffusivity of a linear process is given by²²

$$D = \frac{\langle \mathbf{R}^2 \rangle}{6\tau}, \quad (3.29)$$

where $\langle \mathbf{R}^2 \rangle$ is the mean-square displacement of the collection of particles of the species concerned in time τ , and θ is the fraction of these particles that are simultaneously and independently executing random walk. For vacancy, one can express quite generally that

$$D_v = \left\langle \left| \sum_i^{n_v} \mathbf{r}_i \right|^2 \right\rangle / 6 \left\langle \sum_i^{n_v} \frac{1}{\omega_i} \right\rangle, \quad (3.30)$$

where \mathbf{r}_i is the i th vacancy jump vector, and $1/\omega_i$ is the time between the $(i-1)$ th and the i th vacancy jumps, i.e., ω_i is the "frequency" of the i th vacancy jump. In a vacancy mechanism of diffusion, it is postulated that either an impurity atom or a solvent atom can move only by exchanging sites with a neighboring vacancy. We may elect to study the random walks of various species in the system by focussing our attention on the trajectories of vacancies. The succession of vacancy jumps is composed of two interdispersed series of impurity jumps \mathbf{r}_j and solvent atom jumps \mathbf{r}_k . Clearly,

$$\sum_i^{n_v} \mathbf{r}_i = - \left(\sum_j^{n_A} \mathbf{r}_j + \sum_k^m \mathbf{r}_k \right) = -(\mathbf{R}_A + \mathbf{R}_B), \quad (3.31)$$

where n_v , n_A , and m are the number of vacancy jumps, the number of impurity jumps, and the number of solvent atom jumps, respectively, and \mathbf{R}_A and \mathbf{R}_B are the respective impurity and solvent displacement vectors associated with the trajectory of the particular vacancy under consideration. We may adopt a physical

²² For example, see the review article by S. Chandrasekhar, Rev. Mod. Phys. 15, 1 (1943).

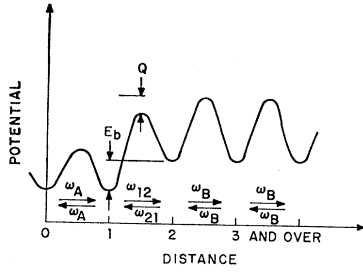


FIG. 4. Schematic diagram of vacancy potential as a function of distance from the impurity atom at the origin, and the four distinguishable vacancy jump frequencies; 1, 2, and 3 denote the coordination order of the sites.

picture as given in Fig. 4, in which the vacancy potential energy is shown to vary with the distance from an impurity at 0. 1, 2, and 3 denote, respectively, the first, second, and third coordination of the impurity atom; E_b denotes the vacancy-impurity binding energy ($E_b = E_\alpha + E_\beta$). Four jump frequencies, ω_A , ω_{12} , ω_{21} , and ω_B can be distinguished,²³ the meaning of which can be read from Fig. 4. Two approximations are made, that the perturbation of the vacancy potential by the presence of the impurity atom terminates at site 2, and that the lowering of the saddle point energy is equal to $\frac{1}{2}E_b$, so that

$$\omega_{21}/\omega_B = \omega_B/\omega_{12} = \exp(E_b/2kT). \quad (3.32)$$

From Fig. 4, one also sees that

$$m = n_{12} + n_{21} + n_B, \quad (3.33)$$

where n_{12} , n_{21} , and n_B are the numbers of vacancy jumps of the types associated with frequencies ω_{12} , ω_{21} , and ω_B , respectively, since there are N_v random-walk trajectories for N_A of impurity atoms $\theta_A = N_v/N_A$, and one has

$$D_A = \left\langle \left| \sum_j^{n_A} \mathbf{r}_j \right|^2 \right\rangle \frac{N_v}{N_A} / 6 \left\langle \frac{n_A}{\omega_A} + \frac{n_{12}}{\omega_{12}} + \frac{n_{21}}{\omega_{21}} + \frac{n_B}{\omega_B} \right\rangle. \quad (3.34)$$

The relationships between the numbers of various types of vacancy jumps can be deduced with reference to the model given in Fig. 4. It is logical to assume that the number of vacancies jumping from site s to site t is directly proportional to the probability of vacancy occupancy at site s and to the jump frequency ω_{st} . One

²³ Strictly speaking, there are, as mentioned earlier, two types of vacancies, v^0 and v^- , and there are four types of vacancy-impurity pairs, A^0v^0 , A^0v^- , A^+v^0 , and A^+v^- . Hence, there will be four different frequencies for each of types ω_A , ω_{12} , and ω_{21} , and two different frequencies for type ω_B . Therefore the average of random walk squared displacements will be dependent on the relative concentrations of these vacancies and vacancy-impurity pairs. Noting that the vacancy level is relatively deep, and assuming that either under conditions of nondegeneracy or of the formation of impurity band, the impurity is completely ionized, we may consider only one type of vacancy, v^- , and one type of vacancy-impurity pair, A^+v^- . Furthermore, such a simplified treatment is entirely adequate for the present purpose of the determination of the diffusivities, the phenomenological coefficients and their interrelations, inasmuch as it is self-consistent.

can relate the probability of vacancy occupancy at the site of type s to the number of sites belonging to type s and to the potential energy at the site of type s . Let p_1 and p_2 represent the probabilities of vacancy occupancy at sites 1 and sites 2, and p_s , the total probability of vacancy occupancy at sites 3 and beyond. Then, if Z_1 and Z_2 are the numbers of the first and of the second coordination sites, respectively, one has

$$p_s/p_2 = (N_s - N_A - Z_1 N_A - Z_2 N_A) / Z_2 N_A. \quad (3.35)$$

In particular, we shall restrict our discussion to diamond lattice where $Z_2 = Z_1(Z_1 - 1)$. Hence,

$$p_s/p_2 = [N_s - (Z_1^2 + 1)N_A] / Z_1(Z_1 - 1)N_A. \quad (3.36)$$

We may drop the subscript 1 from Z_1 . Similarly,

$$p_1/p_2 = [1/(Z-1)] \exp(E_b/kT). \quad (3.37)$$

Then,

$$\begin{aligned} \frac{n_{21}}{n_B} &= \frac{\omega_{21}}{\omega_B} \frac{p_2}{(Z-1)p_2 + Zp_s} \\ &= \frac{(Z-1)N_A}{N_s - 2ZN_A} \exp(E_b/2kT), \end{aligned} \quad (3.38)$$

$$\frac{n_{12}}{n_{21}} = \frac{(Z-1)p_1\omega_{12}}{p_2\omega_{21}} = 1, \quad (3.39)$$

$$\frac{n_A}{n_{12}} = \frac{\omega_A}{(Z-1)\omega_{12}} = \frac{1}{Z-1} \frac{\omega_A}{\omega_B} \exp(E_b/2kT), \quad (3.40)$$

or

$$n_A = \frac{\omega_A}{\omega_B} \frac{1}{N_s - 2ZN_A} \exp(E_b/kT). \quad (3.41)$$

From Eqs. (3.38)–(3.41), one obtains

$$\begin{aligned} \tau &= \langle n_A/\omega_A + n_{12}/\omega_{12} + n_{21}/\omega_{21} + n_B/\omega_B \rangle \\ &= (n_B/\omega_B) [1 + (N_A/N_s) [Z \exp(E_b/kT) - (Z+1)]] \\ &\quad \times [N_s / (N_s - 2ZN_A)]. \end{aligned} \quad (3.42)$$

Hence, Eq. (3.34) becomes

$$D_A = \frac{1}{6} r^2 \omega_A f_A (N_v/N_s) \exp(E_b/kT) / [1 + (N_A/N_s) [Z \exp(E_b/kT) - (Z+1)]], \quad (3.43)$$

where the correlation factor f_A has the usual definition

$$f_A = \left\langle \left| \sum_j^{n_A} \mathbf{r}_j \right|^2 \right\rangle / n_A r^2. \quad (3.44)$$

Equation (3.43) can be written

$$D_A = \frac{1}{6} r^2 \omega_A f_A (N_{Av} / ZN_A), \quad (3.45)$$

in which one has made the substitution of N_{Av} . From Eq. (2.19) it is seen that if $E_A > E_F > E_v$, and hence

$\xi \gg 1$ and $\xi \gg 1$, one has

$$N_{Av} = (N_v N_A / N_s) \alpha \beta / [1 + (N_A/N_s) [\alpha \beta - (z+1)]], \quad (3.46)$$

where one can identify that [q.v. Eq. (2.4) and Eq. (2.25)]

$$\alpha\beta = Z \exp[(E_\alpha + E_\beta)/kT]. \quad (3.47)$$

Actually, Eq. (3.45) is more general, and not restricted to the condition of $E_A > E_F > E_v$ (see Appendix for an alternative method of derivation), while Eq. (3.43) is restricted, since in its derivation, Boltzmann statistics is employed to calculate the probability of site occupancy.

The derivation of the diffusivity of the solvent D_B can be similarly obtained. Following the present approach, however, it is necessary to introduce a correlation factor f_B for solvent diffusion which we define as

$$f_B = \langle \left| \sum_k^m \mathbf{r}_k \right|^2 \rangle / m r^2, \quad (3.48)$$

where m has been defined in Eq. (3.33), which is a sum of all types of vacancy jumps involving exchanges with B atoms. The final result is

$$D_B = \frac{1}{6} r^2 \omega_B f_B \frac{N_v}{N_B} \times \frac{1 + (N_A/N_S)[2(Z-1) \exp(E_b/2kT) - 2Z]}{1 + (N_A/N_S)[Z \exp(E_b/kT) - (Z+1)]}. \quad (3.49)$$

A "vacancy-solvent complex" has not been defined so that the above expression shall be left as it is, which is subjected to the restriction of the validity of Boltzmann statistics. It is noted that one can distinguish three different types of vacancy-solvent pairs, i.e., the pair in which both the vacancy and the interested B atom are surrounded by other B atoms, the pair in which the interested B atom is neighboring to an impurity atom, and the pair in which the vacancy is neighboring to an impurity atom. The diffusivity of vacancy is similarly given by

$$D_v = \frac{1}{6} r^2 \omega_B f_v \left[1 + \frac{N_A}{N_S} \left(\frac{\omega_A}{\omega_B} (Z-1) \exp(E_b/kT) + 2(Z-1) \exp(E_b/2kT) - 2Z \right) \right] / \left[1 + \frac{N_A}{N_S} [Z \exp(E_b/kT) - (Z+1)] \right]. \quad (3.50)$$

The determination of correlation factor f_A for impurity diffusion has been discussed in the literature.^{24,25} In

²⁴ A. D. LeClaire and A. B. Lidiard, *Phil. Mag.* **1**, 518 (1956).

²⁵ J. R. Manning, *Phys. Rev.* **116**, 819 (1959). It is noted that Manning's treatment of the correlation factor of impurity diffusion by means of the concept of equivalent fraction of returning vacancy becomes incorrect when applied to diamond lattice under the general condition that the jump frequencies of a vacancy at the second coordination sites are anisotropic. For a detailed analysis of this subject, see Ref. 26.

particular, the correlation factor of impurity diffusion in diamond lattice is discussed in detail by Hu.²⁶ However, the correlation factors f_B and f_v are not as easily determined. The only work known in literature on f_v ²⁷ is based on the assumption that the vacancy concentration is maintained at local equilibrium. Unfortunately, the validity of this assumption must be seriously questioned even in the presence of internal vacancy sources and sinks (such as dislocations), since the calculation of correlation factor in a random walk is dominated by trajectories in the immediate neighborhood of the originating site, and hence will not be significantly affected by dislocations of typical density in semiconductors. Moreover, the presumption of a local vacancy equilibrium automatically defeats the purpose of the phenomenological formulation which is here to find the diffusion-induced vacancy transients. The association of the vacancy with the solute atom also must not be neglected.

C. Tight-Binding Approximation

Consider a system in which the vacancy-impurity binding energy is high. We shall see that in this system the diffusivities D_B and D_v can be determined quite simply without requiring the evaluation of f_B and f_v . Of course, f_A can be evaluated easily using the conventional methods,²⁴⁻²⁶ but it is unnecessary for this special case. All we need is a self-consistent relationship between D_A , D_B , and D_v (or L_{AA} , L_{BB} , and L_{vv}) for a given atomistic model, from which one can obtain the cross coefficients L_{Av} , L_{Bv} , and L_{AB} , so that one can investigate, for example, the important question of vacancy transient behavior, the so-called "chemical pump" effect, which is to be discussed in Sec. 3 D.

One imagines the system to consist of two subsystems, each of which being a collection of microscopic regions. System I is a collection of microscopic regions, in each of which there is an isolated vacancy. In a microscopic region of system II, there is a vacancy strongly bound to an impurity atom. One need not consider regions having no vacancy, for no event of random walk will take place there. In a region of type II, one imagines the vacancy to move around the impurity atom by traversing the first coordination sites without dissociating from the impurity atom. At times the vacancy also exchanges sites with the impurity atom, thus effecting the displacement of the impurity atom. It should be noted that, strictly speaking, such a process, i.e., the diffusion of an impurity atom and a vacancy as a paired complex, cannot occur in a diamond lattice,^{25,26} although it could be a likely mechanism in close-packed lattices. Without going into detail, let us assume however, that

²⁶ S. M. Hu, *Phys. Rev.* **177**, 1334 (1969).

²⁷ J. R. Manning, in *Lattice Defects and Their Interactions*, edited by R. R. Hasiguti (Gordon and Breach, Science Publishers, Inc., New York, 1967), p. 267 ff.

this is a fairly close approximation.²⁸ It is apparent that in region II, during the same time τ , the vacancy traverses a random path which is given by a fluctuation of one-jump displacement \mathbf{r} superimposed on the path of the impurity atom. The displacements in unit time are related by

$$\mathbf{R}_v = \mathbf{R}_A + \langle \mathbf{r} \rangle = \mathbf{R}_A, \quad (3.51)$$

since the average of fluctuations \mathbf{r} is zero. The diffusivity of vacancy in region II is hence

$$D_v(\text{II}) = \langle \mathbf{R}_v^2 \rangle_{\text{II}} / 6\tau = D_A (N_A / N_{Av}) \\ = \frac{1}{6} r^2 \omega_A f_A (1/Z). \quad (3.52)$$

In region I, the vacancy motion is completely random. Hence,

$$D_v(\text{I}) = \frac{1}{6} r^2 \omega_B = \langle \mathbf{R}_v^2 \rangle_{\text{I}} / 6\tau. \quad (3.53)$$

The vacancy diffusivity in the ensemble is then

$$D_v = \langle \mathbf{R}^2 \rangle_{\text{I+II}} / 6\tau = [(N_v - N_{Av}) / N_v] D_v(\text{I}) \\ + (N_{Av} / N_v) D_v(\text{II}) \quad (3.54) \\ = \frac{1}{6} r^2 [\omega_B - (N_{Av} / Z N_v) (Z\omega_B - \omega_A f_A)].$$

It is clear that if the tightly bound vacancy-impurity pair can be regarded as a single particle, the successive unit displacements of the center of the pair will be uncorrelated, just like a divacancy. Therefore in region II, there will be two B atoms each making a displacement equal to $-\mathbf{R}_{Av}$ from the conservation of lattice sites. One obtains

$$D_B = \frac{1}{6} r^2 (N_v / N_B) [\omega_B - (N_{Av} / Z N_v) \\ \times (Z\omega_B - 2\omega_A f_A)]. \quad (3.55)$$

With the above results and the relationships given in Eqs. (3.18)–(3.20), one can show that for tight-binding approximation,

$$L_{BB} = L_{AA} + L_{vv}, \quad (3.56)$$

$$L_{AB} = -(r^2 \omega_A f_A / 6kT) (N_{Av} / Z) = -L_{AA}, \quad (3.57)$$

$$L_{Av} = 0, \quad (3.58)$$

$$L_{Bv} = -(r^2 N_v / 6kT) [\omega_B - (N_{Av} / Z N_v) \\ \times (Z\omega_B - \omega_A f_A)] = -L_{vv}. \quad (3.59)$$

D. Chemical Pump Effect

A knowledge of vacancy transient behavior is important to the understanding of diffusion via vacancy mechanism. We have seen earlier [Eq. (2.32)] that the equilibrium vacancy concentration is a function of the Fermi level, and hence is a function of impurity concentration which varies with space and time. The question, then, is how does the vacancy concentration respond to the new equilibrium condition as the dif-

²⁸ For example, the presence of Coulombic interaction between a negative vacancy and a positive-impurity ion may tend to keep the vacancy in the neighborhood of the impurity, although this is not in the same sense as the tight binding one would define for a close-packed lattice.

fusion of impurity proceeds? Barring an excessively high density of dislocations, it is logical to suggest that vacancies will flow to the location having a higher rate of Fermi level variation from neighboring regions. If there exists a disparity between the rates of exchange of the vacancy with the impurity atom and with the solvent atom, one may expect an "extraction effect," which will affect the impurity distribution. In other words, the vacancy flux due to a nonvanishing vacancy chemical potential gradient will induce an impurity flux relative to the solvent flux (or any reference frame). Conversely, the diffusion of impurity due to an impurity chemical potential gradient may also induce a vacancy flux, thus causing a departure of local vacancy concentration from its equilibrium value and affecting the local impurity diffusivity. It is clear that this problem should be approached by formulating a complete set of phenomenological equations.

Using a kinetic approach, Seitz showed that⁷ if the diffusivity of the impurity is higher than the self-diffusivity of the solvent, a chemical pump effect will occur such that a vacancy flux will be induced in the opposite direction of the impurity flux. His results may be expressed as follows:

$$J_v = (D_A - D_B) (\partial N_A / \partial x) - D_v (\partial N_v / \partial x). \quad (3.60)$$

The equation above says that even if there is initially no vacancy concentration gradient, there nevertheless will be a vacancy flux due to the presence of an impurity concentration gradient as long as $D_A \neq D_B$. Unfortunately, his analysis did not take into account the correlation effect, which concept was introduced two years later after his work.²⁰ In this section, we show that the correlation factor has an important consequence on the validity of Seitz's analysis.

We first consider the case of tight-binding approximation (which can exist at least in close-packed lattices). We set up a complete set of phenomenological equations for the three species A , B , and v . Using the results of Eqs. (3.56)–(3.59), together with the Gibbs-Duhem relation to eliminate $\partial \mu_B / \partial x$, one obtains, after appropriate manipulations, the following flux expressions:

$$J_A = -L_{AA} [1 + (N_A / N_B)] (\partial \mu_A / \partial x), \quad (3.61)$$

$$J_B = [L_{AA} + (N_A / N_B) (L_{AA} + L_{vv})] (\partial \mu_A / \partial x) \\ + L_{vv} (\partial \mu_v / \partial x), \quad (3.62)$$

$$J_v = -L_{vv} [(N_A / N_B) (\partial \mu_A / \partial x) + \partial \mu_v / \partial x]. \quad (3.63)$$

Equation (3.63) suggests that even if the vacancy chemical potential gradient is initially zero, there can be a vacancy flux, which is given by $-L_{vv} (N_A / N_B) \partial \mu_A / \partial x = D_B \partial N_B / \partial x$, and is hence in the same direction as J_A . Moreover, D_A is not involved at all. These are in apparent disagreement with Seitz's results as given in Eq. (3.60). An examination of Eqs. (3.43) and (3.55) shows that D_A can be considerably larger than D_B , which relation is most easily seen at very small N_A . This is primarily

a consequence of the strong vacancy-impurity association, such that $N_{Av} > ZN_A N_v / N_S$. On the other hand, the effect of the ratio ω_A / ω_B , which may be greater than unity, will be partially compensated by the fact that f_A steadily decreases with the increase of the ratio ω_A / ω_B .²⁶

Next we can show that if there were no correlation effect, our general expressions for D_A , D_B , and D_v as given in Eqs. (3.45), (3.49), and (3.50) would reduce the phenomenological equations to the results of Seitz's, even in the presence of vacancy-impurity association which was neglected in his analysis. Consider the phenomenological equation for J_v . We let $f_A = f_B = f_v = 1$ in the expressions (3.45), (3.49), and (3.50). Then L_{ii} 's are obtained from the D_i 's using Eq. (3.28). These are then substituted into Eqs. (3.18) to (3.20) to obtain the cross coefficients L_{Av} and L_{Bv} . After appropriate manipulations, one obtains

$$L_{Av} = (-r^2 \omega_A / 6kT)(N_{Av} / Z) = -L_{AA}, \quad (3.64)$$

and similarly

$$L_{Bv} = -L_{BB}. \quad (3.65)$$

L_{vv} remains defined by itself. Substituting L_{Av} and L_{Bv} from above into the phenomenological equation for J_v , one obtains

$$J_v = -L_{Av}(\partial \mu_A / \partial x) - L_{Bv}(\partial \mu_B / \partial x) - L_{vv}(\partial \mu_v / \partial x) \\ = (D_A - D_B)(\partial N_A / \partial x) - D_v(\partial N_v / \partial x), \quad (3.66)$$

where we have let $\gamma_A = \gamma_B = \gamma_v = 1$, since this assumption is implied in Seitz's analysis, and we have also utilized the Gibbs-Duhem relation. It can be seen that Eq. (3.66) is the same as Eq. (3.60).

Consider the paradox of the diffusion of a tracer in its isotope lattice. If the tracer concentration is low and each tracer atom can be regarded as isolated, the tracer diffusivity will be only an f_A fraction of the lattice self-diffusivity.²⁰ Then, according to Eq. (3.60), one would expect a vacancy flux. On the other hand, we see that the vacancy movements must be completely random since the vacancy does not distinguish a tracer atom from its neighboring isotope atoms. Thus, one expects no net vacancy flux under a uniform vacancy concentration.

We thus conclude that the chemical pump effect, which may or may not exist for a certain system, does not arise simply because of the disparity between D_A and D_B but depends on the correlated random walk.

4. DIFFUSION IN SEMICONDUCTORS

Having determined the phenomenological coefficients and the chemical potentials, we should now be able to solve the problem of diffusion in semiconductors based on the phenomenological equations, from which continuity equations are easily obtainable. Because both L_{ij} and μ_j are complicated functions of N_j 's, one must anticipate solutions of the continuity equations by

numerical method. In the present paper, however, we confine our discussion to the case of strong vacancy-impurity association under the condition that $E_A \gtrsim E_F \gtrsim E_v$. We also assume that the system is moderately dilute (i.e., $\alpha \beta N_A / N_S \cong N_{Av} / N_v \ll 1$). We introduce the intrinsic diffusivities D_A^* , D_B^* , and D_v^* . These are defined as the diffusivities at infinite dilution of N_A , with vacancy at its thermal equilibrium concentration. From Eqs. (3.43), (3.54), and (3.55), one has

$$D_A^* = \frac{1}{6} r^2 \omega_A f_A (\alpha \beta N_v^* / Z N_S), \quad (4.1)$$

$$D_B^* = \frac{1}{6} r^2 \omega_B (N_v^* / N_S), \quad (4.2)$$

$$D_v^* = \frac{1}{6} r^2 \omega_B. \quad (4.3)$$

Then, it follows that the extrinsic diffusivities are given by

$$D_A = D_A^* (N_v / N_v^*), \quad (4.4)$$

$$D_B = D_B^* (N_v / N_v^*) \\ \times [1 - (N_{Av} / Z N_v)(Z - 2f_A \omega_A / \omega_B)] \quad (4.5)$$

$$\cong D_B^* (N_v / N_v^*),$$

$$D_v = D_v^* [1 - (N_{Av} / Z N_v)(Z - f_A \omega_A / \omega_B)] \cong D_v^*. \quad (4.6)$$

Under typical conditions, N_v is very small in comparison with N_A , so that γ_v and γ_A as given in Eqs. (2.35) and (2.36) are in effect functions of N_A only. Then J_A and J_v as given by Eqs. (3.61) and (3.63) can be expressed as

$$J_A = -D_A^* (N_v / N_v^*) (1 + N_A / N_B) (1 + \partial \ln \gamma_A / \partial \ln N_A) \\ \times (\partial N_A / \partial x) \\ \cong -D_A^* (N_v / N_v^*) (1 + \partial \ln \gamma_A / \partial \ln N_A) \\ \times (\partial N_A / \partial x), \quad (4.7)$$

$$J_v = -D_v^* (N_v / N_S) [1 + \partial \ln \gamma_A / \partial \ln N_A \\ + (N_S / N_A) (\partial \ln \gamma_v / \partial \ln N_A)] (\partial N_A / \partial x) \\ - D_v^* (\partial N_v / \partial x) \\ \cong -D_v^* (N_v / N_A) (\partial \ln \gamma_v / \partial \ln N_A) (\partial N_A / \partial x) \\ - D_v^* (\partial N_v / \partial x), \quad N_S / N_A \gg 1. \quad (4.8)$$

Equation (4.8) shows that even when $\partial N_v / \partial x$ is initially zero, a vacancy flux will occur because of the lowering of the vacancy potential energy (and hence vacancy activity coefficient γ_v) in the region doped with A . From Fig. 2, one sees that $\partial \ln \gamma_v / \partial \ln N_A$ (represented by the slopes of these curves) is always negative, implying that its contribution to the vacancy flux is always opposite to the flux of A , as indicated by Eq. (4.8). Note that this statement is in disagreement with the discussions following Eq. (3.63) in Sec. 3 D, in which it is suggested that a vacancy flux will be induced in the same direction as J_A for the tight-binding case. The reason is that in Sec. 3 D in which Seitz's analysis of chemical pump is examined, we have assumed an electrically neutral system, such that $\gamma_v = 1$ and $\gamma_A = 1$. This is to say that we were considering in that case only the first term (which is unity) in the parentheses of Eq. (4.8). Since, in the semiconductor

system considered in the present section, $\partial \ln \gamma_v / \partial \ln N_A$ is of the order of -1 , and $\partial \ln \gamma_A / \partial \ln N_A$ is also of the order of 1 , the third term in the brackets of Eq. (4.8) dominates, and the first two terms become negligible in comparison. Physically, one imagines this effect as charged vacancies being driven out by the internal electric field created by virtue of $N_A(x)$. In a normal diffusion process in which $\partial N_v / \partial x$ is initially zero, the term involving $\partial \ln \gamma_v / \partial \ln N_A$ dominates at the beginning period of the process, causing a net vacancy efflux (an influx, on the other hand, will be observed in the diffusion of an acceptor). Steadily, as $\partial N_v / \partial x$ continues to build up, $\partial N_A / \partial x$ continues to decrease, and, at least in the region of small x , one may expect a reversal of the vacancy flux. The reverse flux should be in general extremely small, such that a quasi-equilibrium condition will be reached for the vacancy in the region of small x fairly quickly.

The subject of vacancy transients and its effect on diffusion will be treated in detail in a subsequent paper. Herein we limit our discussions to the case in which a quasiequilibrium condition for the vacancy concentration can be assumed. Under such a condition, the vacancy concentration is given by Eq. (2.33) or Eq. (2.32) (or $N_v = N_v^e = N_v^* \gamma_v^{-1}$), which is to a good approximation a function of N_A only. Then, Eq. (4.7) reduces to a very simple form:

$$J_A = -D_A^* \gamma_v^{-1} (1 + \partial \ln \gamma_A / \partial \ln N_A) (\partial N_A / \partial x). \quad (4.9)$$

Equation (4.9) does not explicitly contain variables or functions of N_v . Hence, Eq. (4.9), and the continuity equation derived from it, can be solved alone irrespective of the vacancy flux.

When the Fermi level is evaluated under the conditions of nondegeneracy and local charge neutrality, one has, from (2.26), (2.27), (2.35), and (2.36),

$$\gamma_v^{-1} \cong \xi / (1 + \xi^*) = (\xi^* / 1 + \xi^*) \times [N_A / 2n_i + ((N_A / 2n_i)^2 + 1)^{1/2}] \quad (4.10)$$

and

$$\partial \ln \gamma_A / \partial \ln N_A \cong (\partial / \partial \ln N_A) \ln (\xi^* / \xi) = [N_A / (N_A^2 + 4n_i^2)^{1/2}]. \quad (4.11)$$

If we substitute (4.10) and (4.11) into (4.9), and define $D_A^0 = D_A^* \xi^* / (1 + \xi^*)$, we obtain the same result as given earlier by Hu and Schmidt.⁶ These authors showed through numerical computations how impurity profiles in Si through diffusion are affected by the variation of equilibrium vacancy concentrations.

It should be noted that under the assumption of quasiequilibrium vacancy concentration, Eq. (4.9) is no longer restricted to tight-binding systems, and is of general validity provided N_A is relatively dilute.

A final remark is in order. In typical semiconductor devices, the junction depth is often of the order of $\lesssim 10^{-4}$ cm, and the dislocation density is often of the order of $\lesssim 10^8$ cm⁻². It can be shown that under such conditions, the dislocations are not effective vacancy

source, and a general vacancy undersaturation will occur. However, the magnitude of undersaturation is very small, and its effect on the impurity diffusion is similarly small.

ACKNOWLEDGMENTS

This paper, which aims at a more rigorous extension of the author's previous paper, (Ref. 6) owes its birth to the criticisms and suggestions of Dr. G. R. Gunther-Mohr, particularly in regard to the question of vacancy transients and the phenomenological formalism. The author is much indebted to Dr. H. Rupprecht, Dr. D. Jepsen, and Dr. K. Weiser for many valuable discussions and suggestions.

APPENDIX: ALTERNATIVE DERIVATION OF IMPURITY DIFFUSIVITY

To avoid the restriction of Boltzmann statistics, as discussed in Sec. 3 B, the impurity diffusivity may be derived alternatively using the following argument.

One may assume that the vacancy spends an overwhelming portion of the time period $1/\omega_i$ staying at the site before surmounting the activation energy barrier. One can then identify the portion of time τ_{Av} in which, during its journey of random walk, a vacancy stays as a nearest neighbor to a specified impurity atom, and the portion of time τ_A in which the vacancy stays away from this impurity atom. (The vacancy may stay as a nearest neighbor to some other impurity atoms, notwithstanding.) By an inspection of Fig. 4, one sees that

$$\tau_{Av} = n_A / \omega_A + n_{12} / \omega_{12}, \quad (A1)$$

$$\tau_A = n_{21} / \omega_{21} + n_B / \omega_B. \quad (A2)$$

A fundamental postulate in statistical thermodynamics states that the time average of a state variable of a subsystem is equal to the ensemble average of that variable. Thus, one has

$$\tau_{Av} / \tau_A = (n_A / \omega_A + n_{12} / \omega_{12}) / (n_{21} / \omega_{21} + n_B / \omega_B) = N_{Av} / (N_A - N_{Av}). \quad (A3)$$

Since $n_{12} / n_A = (Z - 1) \omega_{12} / \omega_A$ from Eq. (3.41), it follows that

$$\tau_{Av} = Z n_A / \omega_A, \quad (A4)$$

$$\tau_A = Z n_A (N_A / N_{Av} - 1) / \omega_A. \quad (A5)$$

Therefore,

$$\tau = \tau_{Av} + \tau_A = Z n_A N_A / N_{Av} \omega_A. \quad (A6)$$

One thus obtains

$$D_A = \langle |\sum_j^{n_A} \mathbf{r}_j|^2 \rangle / 6\tau = \frac{1}{6} r^2 \omega_A f_A (N_{Av} / Z N_A), \quad (A7)$$

which is Eq. (3.45), and in which f_A is the same as defined in Eq. (3.46). Note that in this method, it is not necessary to weigh the diffusivity with the ratio of the number of vacancy trajectories to the number of impurity atoms present.