PHYSICAL REVIEW B VOLUME 10, NUMBER 4 15 AUGUST 1974

Enthalpy of vacancy migration in Si and Ge

J. A. Van Vechten
Bell Laboratories, Murray Hill, New Jersey 07974

(Received 5 September 1973)

Values for the enthalpy of vacancy migration ΔH_m in Si and Ge of 0.33 and 0.2 eV, respectively, have been deduced from EPR and ir absorption experiments at low temperature. However, at high temperature, quenching and diffusion experiments yield values for ΔH_m of 1.2 and 1.0 eV. There are corresponding discrepancies in the values for the entropy of vacancy migration ΔS_m, although ΔS_m values are not accurately determined in these experiments. It is here demonstrated that both empirical conclusions are correct and, in particular, that various suggestions of a complex species to account for the high-temperature migration are inconsistent with any reasonable thermodynamic analysis. It is concluded that the reason for this large change in ΔH_m and ΔS_m with temperature is that the predominant mode of single-vacancy migration, i.e., the transition state, at high temperatures is different than that at low temperatures. One may explain this change in the dominant transition state for migration and account for both sets of values for ΔH_m by resort to the macroscopic cavity model of the vacancy which was introduced by Phillips and Van Vechten. In that model there are two contributions to the enthalpy of a vacancy—a contribution due to the breaking of covalent bonds (short-ranged forces) and a larger contribution due to long-ranged forces. Both of these contributions have previously been determined from data completely independent of the vacancies. The high-temperature values of ΔH_m can be explained by assuming that, in addition to the bond-breaking term, there is a contribution from long-ranged forces. The latter is calculated by assuming that the lattice is undistorted beyond the nearest-neighboring atoms to the vacancy and that the anisotropy of these long-ranged forces at the vacancy is the same as is observed for the bulk crystal. For this mode ΔS_m is large. The low-temperature value of ΔH_m is accurately given as just the energy required to break the additional bonds necessary for vacancy migration. Thus, the contribution from long-ranged forces observed at high temperatures is avoided at low temperatures by a correlated motion of the atoms around the vacancy which minimizes the enthalpy of the vacancy at the saddle point and thus also minimizes ΔH_m. For this mode ΔS_m is small and apparently even negative. The macroscopic model also provides a description of the atomic displacements and Jahn-Teller distortion about the vacancy when it is at a lattice site.

I. INTRODUCTION

One of the longest and most perplexing controversies in solid state physics is over the value of the enthalpy of vacancy migration, ΔH_m, in Si and Ge.[1–9] This is a problem of considerable practical importance because vacancy migration is usually the dominant mechanism for atomic diffusion and for the growth of dislocations. Also, vacancy donor and acceptor states are often effective in determining the electrical conductivity of a semiconductor. Therefore, vacancy migration is crucial to the fabrication of many devices and vacancy migration to and across a junction may be an important failure mechanism. Thus, there are practical as well as academic reasons for wishing to understanding why values of ΔH_m determined in the high-temperature experiments are 4–5 times the values determined at low temperature.

The various experiments[10–17] from which information on vacancy migration has been deduced will be reviewed in Sec. II. As these experiments are rather diverse, a certain amount of background discussed seems to be warranted.

The most conventional parameter which may be used to describe the results of the various experiments is the diffusion coefficient or diffusivity D_v of the single vacancies. As always, D_v is defined by

$$\vec{J}_v = -\underline{D}_v \cdot (\overrightarrow{\text{grad}}[V]), \qquad (1.1)$$

where $\vec{J}_v$ is the flux of single vacancies and $[V]$ is their concentration. $\underline{D}_v$ is, in general, a symmetric second-rank tensor, but for a cubic lattice, such as that of Si and Ge, it may be replaced by the scalar quantity D_v. If all other mechanisms for the self-diffusion of the host atoms may be neglected, the relation between D_v and the self-diffusion coefficient D_s is[18]

$$D_s = [V] D_v / n_H , \qquad (1.2)$$

where n_H is the concentration of host lattice sites.

In this paper the results of absolute rate theory are used to provide an expression for D_v in terms of a thermodynamic description of the state of the vacancy at a lattice site and at the saddle point of its path between lattice sites.[19–21] The state of the vacancy at the saddle point will be denoted as the "transition state." This expression is[2,9,19]

$$D_v = \tfrac{1}{3} d_0^2 \nu e^{\Delta S_m / k} e^{-\Delta H_m / kT} , \qquad (1.3)$$

where $\frac{1}{3}$ is a geometrical factor appropriate to the diamond lattice, d_0 is the interatomic spacing, ν is usually described as the frequency with which the vacancy attempts to jump from one lattice site to another,[19,20] k is Boltzmann's constant, and T denotes temperature. ΔS_m is the entropy of vacancy migration, which is defined as

$$\Delta S_m = \Delta S_{v'} - \Delta S_v . \qquad (1.4)$$

In Eq. (1.4), $\Delta S_{v'}$ and ΔS_v denote the entropy of formation of the vacancy in the transition state and at a lattice site, respectively. The enthalpy of vacancy migration, ΔH_m in Eq. (1.3), is similarly defined as

$$\Delta H_m = \Delta H_{v'} - \Delta H_v , \qquad (1.5)$$

where $\Delta H_{v'}$ and ΔH_v denote the enthalpy of vacancy formation in these two states. (Note that for all experiments considered in this paper, the pressure was sufficiently low that the various enthalpies are effectively equal to the corresponding internal energies.)

The reader should note that from vacancy migration experiments alone, i.e., from a knowledge of $D_v(T)$ alone, one can not determine ν and ΔS_m separately. One can only distinguish the pre-exponential factor D_0 from the $e^{-\Delta H_m/kT}$ factor

$$D_v(T) = D_0 e^{-\Delta H_m/kT} , \qquad (1.6)$$

where

$$D_0 = \tfrac{1}{3} d_0^2 \nu e^{\Delta S_m/k} . \qquad (1.7)$$

At present, the theory of ν and of ΔS_m is not sufficently developed that one might specify either with confidence.[19,20] However, it is a common practice to assume that ν is the Debye frequency ν_D

$$\nu \cong \nu_D = k\Theta/h \qquad (1.8)$$

where Θ is the Debye temperature and h is Planck's constant.

It should be evident that ν_D is actually an approximate upper bound to the plausible values of ν. This upper bound should be approached only if the mechanism by which the vacancy migrates is simple and requires no correlated motion of the neighboring atoms. If the migration mechanism requires the correlated motion of a large number of neighboring atoms, then ν could be orders of magnitude less than ν_D. Therefore, we should expect to replace Eq. (1.8) with

$$\nu = \gamma \nu_D , \qquad (1.9)$$

where

$$\gamma \approx 1 \qquad (1.10)$$

for simple migration and

$$\gamma \ll 1 \qquad (1.11)$$

for complex or correlated migration.

The reader may question whether this discussion of ν is meaningful in view of the crude state of the theory for ΔS_m and the fact that diffusion experiments can determine only the value of D_0. The answer is that, if the premise of the transition state is valid, then the actual concentration of vacancies in the states V and V' will obey the law of mass action[22] in thermal equilibrium

$$[V']/[V] = e^{\Delta S_m/k} e^{-\Delta H_m/kT} . \qquad (1.12)$$

The ionization spectrum of V' is presumably different from that of V so that one may be able to distinguish these two states of the single vacancy in optical experiments. Particularly if ΔS_m is large (~ 10 k) as is often surmised,[2,4,9] $[V']$ may become appreciable even at moderate temperatures in samples, wherein $[V]$ is large as a result of irradiation. Indeed, there are occasional reports of a "new state of the vacancy" in Si or Ge samples which have been irradiated near room temperature.[16,23] Such phenomena should be systematically investigated in hopes of obtaining a good estimate of ΔS_m independent of ν. However, this separation is not strictly necessary to the development of this paper.

In this context, one may conclude that a better definition of ν than that cited after Eq. (1.3) can be cast. Instead of the "frequency with which the vacancy attempts to jump from one lattice site to another," ν might better be described as a factor proportional to the inverse of the time required for the transformation $V \rightleftarrows V'$. Equations (1.9)–(1.11) would obtain with this definition also.

Now, the empirical situation in Ge, i.e., the result of the experiments described in Sec. II, is illustrated in Fig. 1. The situation in Si is qualitatively similar. The two lines representing high-temperature experiments on the $\log_{10} D_v$ versus inverse temperature plot are drawn from the conclusion of Hiraki's analysis of his high temperature vacancy annealing studies[10]

$$D_v(\mathrm{Ge}) = 2\exp[-(1.2\ \mathrm{eV})/kT]\ \mathrm{cm}^2/\mathrm{sec}, \qquad (1.13)$$

and from a comparison of Logan's data[11] on the van't Hoff plot of the vacancy concentration in Ge, i.e., $\log_{10}[V]$ vs $1/kT$, with the Letaw *et al.* data[12] for the self-diffusion coefficient D_s,

$$D_v(\mathrm{Ge}) = 3\exp[-(1.0\ \mathrm{eV})/kT]\ \mathrm{cm}^2/\mathrm{sec}. \qquad (1.14)$$

For the sake of clarity, both of these experimental lines are shown with dashed extrapolations beyond the range of reliable data. The line representing low-temperature experiments, which has large error bars on the absolute magnitude factor D_0,

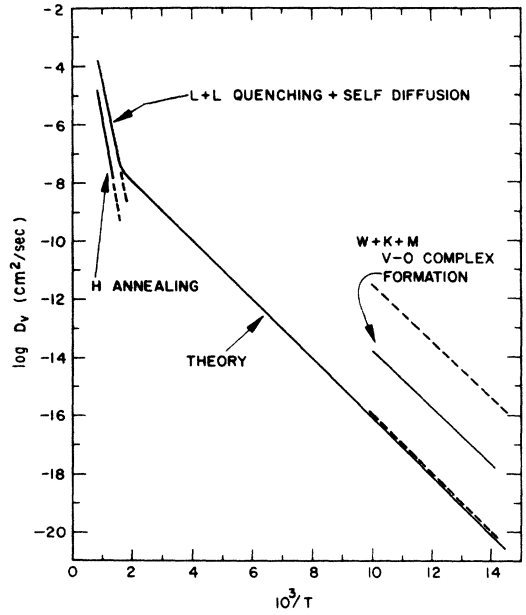


FIG. 1. Empirical diffusivity as a function of temperature $D_v(T)$ for single vacancies in Ge. The empirical situation with vacancies in Si is qualitatively similar. The enthalpy of vacancy migration ΔH_m is proportional to the slope of $\log_{10} D_v$ vs T. The labels on the experimental lines denote the workers and their method. Thus, H denotes Hiraki (Ref. 10); L + L denotes Logan (Ref. 11) and Letaw *et al.* (Ref. 12); W + K + M denotes Whan (Ref. 13) and Klontz and MacKay (Ref. 14).

$$D_v(\mathrm{Ge}) = D_0 \exp[-(0.2\ \mathrm{eV})/kT] \qquad (1.15)$$

is the result of Whan's infrared-absorption (irA) experiments[13] that identified the low-temperature defect transformation process observed by Klontz and MacKay[14] as the formation of vacancy-oxygen impurity complexes. The quoted value of $\Delta H_m = 0.2$ eV is that deduced by Klontz and MacKay,[14] but it can also be deduced from the annealing curves presented in Fig. 2 of Ref. 13.

The theory used to analyze the low-temperature data on the formation of complexes[13–15] from single vacancies and impurities (interstitial oxygen) is known as diffusion limited kinetics. The requisite result is[24]

$$1/\tau = 4\pi r_0 D_v [I], \qquad (1.16)$$

where τ is the time constant for the exponential decay of free single vacancies (above their equilibrium concentration, which is negligible at low temperatures), r_0 is the capture radius of the impurity I, i.e., the radius within which the vacancy migrates directly to I and forms the complex, and $[I]$ denotes the concentration of the impurity. It is a common practice to assume

$$r_0 = d_0 . \qquad (1.17)$$

With the assumption of Eq. (1.17) and the data in Ref. 13, one obtains the value of D_v indicated by the upper limit of the error bars in Fig. 1 ($D_0 \approx 4\times 10^{-2}$ cm²/sec). However, it seems clear that d_0 is the minimum plausible value for r_0 and, particularly if either or both the vacancy or the impurity is charged, r_0 could be orders of magnitude larger, so that D_v would be orders of magnitude less.[24] (It will be shown that the charge state of the vacancy affects ΔH_m by an amount of order 0.1 eV and can not be the source of the discrepancy of interest.) The actual situation obtaining in Ge under the conditions of the experiments of Refs. 13 and 14 is somewhat unclear. However, in Si the situation is well established and Watkins has identified the process in which a neutral single vacancy migrates to and complexes with a neutral interstitial oxygen impurity. For this process and assuming Eq. (1.17), Watkins found[15] the effective frequency factor in Eq. (1.7) to be $\nu e^{\Delta S_m/k} = 2\times 10^{12}$ sec^{-1} at $T = 160$ °K, or equivalently

$$D_v(\mathrm{Si}) = 3.7\times 10^{-4} \times \exp[-(0.33\pm 0.03\ \mathrm{eV})/kT]\ \mathrm{cm}^2/\mathrm{sec}. \qquad (1.18)$$

If one scales the effective frequency factor, or equivalently D_0, according to the relative Debye temperatures of Si and Ge [$\Theta(\mathrm{Si}) \cong 650$ °K and $\Theta(\mathrm{Ge}) \cong 370$ °K], one estimates

$$D_v(\mathrm{Ge}) = 2\times 10^{-4} \exp[-(0.2\ \mathrm{eV})/kT]\ \mathrm{cm}^2/\mathrm{sec}. \qquad (1.19)$$

Equation (1.19) is shown as the intermediate, solid line representing the low temperature experiments in Fig. 1. The lower limit of the error bars indicated for this low-temperature data in Fig. 1 is arbitrarily chosen to be the same distance from the solid line [Eq. (1.19)] as is the upper limit.

It should be noted that the discrepancy in the slope of D_v vs $1/T$, and therefore in ΔH_m, is well established[1–9] despite the uncertainty in the absolute magnitude of D_v and of D_0. This is true both for Ge [compare Eq. (1.15) with both Eqs. (1.13) and (1.14)] and for Si, where Eq. (1.18) may be compared with the result of high-temperature experiments[9,16,17] (within a few hundred °K of the melting point $T^F = 1685$ °K)

$$D_v(\mathrm{Si}) = 4 \exp(-\Delta H_m/kT)\ \mathrm{cm}^2/\mathrm{sec}, \qquad (1.20)$$

with ΔH_m between 1.0 and 1.5 eV. D_0 is uncertain because of uncertainties in parameters, such as r_0, which are difficult to determine. However, both the high-temperature experiments and the low-temperature experiments will determine ΔH_m correctly, within an accuracy of perhaps 20–30%, provided that these parameters do not change appreciably over the temperature range investigated.

The curve labeled "theory" in Fig. 1 is the interpolation-extrapolation formula

$$D_v = 3\exp[-(1.0\text{ eV})/kT] + 1\times 10^{-6} \times \exp[-(0.2\text{ eV})/kT]\text{ cm}^2/\text{sec}. \qquad (1.21)$$

This corresponds to the hypothesis of this paper that there are two modes of migration or transition states available to the single vacancy and that one state, denoted $V'(T_h)$, dominates experiments at high temperatures because it has a larger entropy and/or frequency factor while the other transition state, denoted $V'(T_l)$, dominates low-temperature experiments because it has a lower enthalpy. The theory of this paper accounts for the two values of ΔH_m but can make only crude estimates of ΔS_m and the true frequency factor ν.

The reader will note the theoretical curve in Fig. 1 passes slightly below the lower limit of the error-bars of the low temperature experiments. Note further that the theoretical Eq. (1.21), assumes that both preexponential factors are independent of T. The magnitude of the first preexponential factor, $D_0(T_h)$, was fixed to agree with Eq. (1.14), i.e., the data of Logan[11] and Letaw *et al.*,[12] as this value seems to be well established in view of the agreement with an entirely different experiment.[10] The low-temperature pre-exponential, $D_0(T_l)$ was chosen so that the knee of the curve would occur just to the right of the limit of the high-temperature data in Fig. 1. This was done so as to minimize the discrepancy with the low temperature data while maintaining the assumption that

$$\frac{dD_0(T_l)}{dT} = 0 \qquad (1.22)$$

and the agreement with high-temperature data. Although the lower limit of the error bars was arbitrarily set, the magnitude of the discrepancy (more than two orders of magnitude) with Eq. (1.19) would seem to cast in doubt the assumption of Eq. (1.22). This situation is unusual because it appears $dD_0(T_l)/dT < 0$, whereas one normally expects[25] $dD_0/dT \geq 0$. We shall see that the hypothesis that $dD_0(T_l)/dT < 0$ in some range in T is plausible within the present theory because the frequency ν might be significantly greater at low temperatures. However, at this point, the author prefers to hypothesize that the low-temperature values of D_0 have been indeed overestimated by more than two orders of magnitude so that the assumption of Eq. (1.22) is indeed correct. As one can see from Fig. 1, the contrary hypothesis would require $dD_v/dT \leq 0$ over some range in T!

Now equations of the form of Eq. (1.21) are not particularly uncommon in the field of atomic diffusion. They usually signify that two entirely separate mechanisms are contributing to the total diffusion. Therefore, several previous authors have concluded[1-9] from the data upon which Fig. 1 was drawn that the species observed at high temperatures T_h is not the same as that observed a low temperatures T_l. Electron-paramagnetic-resonance (EPR) experiments by Watkins and Corbett[15,26,27] in Si at T_l, provide a great deal of microscopic information about the species under observation and have firmly identified the single vacancy in its various charge states. On the other hand, the quenching diffusion, and annealing experiments at T_h provide no microscopic description of the species whatever. Therefore, these previous authors have tried to explain the large values of ΔH_m observed at T_h by proposing that the single vacancies condense into various complex species, such as divacancies, which are then assumed to migrate with a large enthalpy of migration. In Sec. III it is shown that simple thermodynamic arguments against all such hypotheses are extremely persuasive and that the species observed at T_h must surely be the simple single vacancy.

In fairness to Seeger and Chick, it should be noted that they also came to some of the conclusions presented here.[2] They recognized that the divacancy hypothesis could not be correct and noted strong evidence that the ΔH_m of the single vacancy varied with temperature. However, their solution to the problem was to propose that the simple single vacancies condense into a rather loosely defined "extended-defect" state at T_h. One might say that their extended-defect state corresponds to the hypothesis presented here of a transition state or mode of migration that involves the correlated motion of many atoms.

However, there are two important distinctions between the Seeger-Chick extended-defect hypothesis and the correlated motion proposed here. First, Seeger and Chick invoke the complex, extended state to account for high temperature observations, whereas in the present analysis, the correlated motion occurs at low temperatures, as is the case, for example, with superconductivity and superfluidity. Second, Seeger and Chick imply that a sort of phase transition has occurred to the "normal" state of the vacancy, i.e., that which has the vacancy centered on a lattice site at T_l. On the other hand, the present hypothesis is that it is the dominant *transition state* which is different at T_l and T_h, i.e.,

$$V'(T_h) \neq V'(T_l) , \qquad (1.23)$$

and that the state of the vacancy at the lattice site V remains unchanged;

$$V(T_h) = V(T_l) . \qquad (1.24)$$

Within absolute rate theory, the Seeger-Chick hypothesis is not tenable. As by definition within absolute rate theory

$$\Delta H_m(T_h) = \Delta H_{v'}(T_h) - \Delta H_v(T_h) \ , \qquad (1.25)$$
$$\Delta H_m(T_l) = \Delta H_{v'}(T_l) - \Delta H_v(T_l),$$

where $\Delta H_{v'}(T_h)$ and $\Delta H_{v'}(T_l)$ denote the enthalpy of formation of $V'(T_h)$ and $V'(T_l)$ and $\Delta H_m(T_h)$ and $\Delta H_m(T_l)$ are the respective enthalpies of migration, the fact that $\Delta H_m(T_h) > \Delta H_m(T_l)$ requires either (a) $\Delta H_{v'}(T_h) > \Delta H_{v'}(T_l)$ or (b) $\Delta H_v(T_h) < \Delta H_v(T_l)$. But case *b*, the Seeger-Chick hypothesis, is not satisfactory because the more probable state of any system at low temperature will always be that with the lower enthalpy. The present theory assumes case *a*, $\Delta H_v(T_h) = \Delta H_v(T_l)$ as per Eq. (1.24), and that the concentrations $[V'(T_h)]$ and $[V'(T_l)]$ are governed by the law of mass action at all temperatures.

In Sec. IV it is observed that any accurate and rigorous *microscopic* theory of ΔH_m will be even more difficult than such a theory for ΔH_v will be when and if developed. However, as Phillips and Van Vechten have recently shown,[28] the atomic vacancy in tetrahedral semiconductors such as Si and Ge may properly be treated as a *macroscopic* cavity on account of the high valence electron density and consequent short range of the effects which established the surface energy of these materials. Use of this macroscopic cavity model for the vacancy allows a vastly simplified yet accurate analysis[28,29] of ΔH_v. The method is similar to the classical treatment of voids. Therefore, the problem of ΔH_m is treated using the macroscopic cavity model and the absolute rate theory definition of ΔH_m [Eqs. (1.5) and (1.25)]. That is, the enthalpies of formation of the vacancy at a lattice site and in the possible transition states are evaluated and compared.

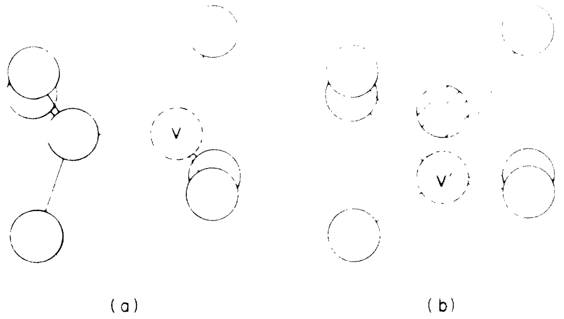
The central feature of this analysis and of the cavity model is that ΔH_v and $\Delta H_{v'}$ may be divided into two contributions.[28] The larger contribution, denoted ΔH_v^m or $\Delta H_{v'}^m$, is due to long-ranged forces that also appear in simple metals,[30,31] such as the metallic, liquid phase of Si and Ge. The smaller contribution, denoted ΔE^b or $\Delta E^{b'}$, is due to the short-ranged and highly directional forces of the covalent bonds in the solid semiconducting phase of Si and Ge. Thus, ΔE^b is related to the difference in cohesive energy between the semiconducting and metallic phases of the same composition[32] and may be regarded as a residual term which corrects ΔH_v^m to account for the "dangling bonds" in the semiconductor. Therefore,

$$\Delta H_v = \Delta H_v^m + \Delta E^b \ , \quad \Delta H_{v'} = \Delta H_{v'}^m + \Delta E^{b'} \ . \qquad (1.26)$$

Considering the state of the vacancy centered at a lattice site, V, it is evident in Fig. 2(a) that there must be four broken bonds "dangling" into the cavity

$$\Delta E_v^b = 4E^b(T) \ , \qquad (1.27)$$

where $E^b(T)$ is the difference in cohesive energy per bond between the semiconducting and metallic phases of the same composition at temperature T. (We shall not be interested in effects at the exterior surface of the sample.) The theory of Ref. 32, which has been checked against experiment by the comparison of calculated and observed pressure-temperature phase diagrams, is used to calculate $E^b(T)$. In calculating ΔH_v^m, it is assumed that the anisotropy of the surface energy, and therefore of the long-ranged forces, is the same on the surfaces of the vacancy cavity as that on a macroscopic cavity in the bulk material. Therefore, the equilibrium shape of V is taken to be the same as that of a macroscopic cavity or of the sample itself[33]—an octahedron. This assumption is supported by the comparison[28,29,34] of the calculated ΔH_v with the results of high temperature quenching experiments and by the observation that the octahedral equilibrium shape of the vacancy leads naturally to the tetragonal Jahn-Teller distortion of the atoms about the vacancy which is observed at low temperatures.[15,26,27,35,36]



(a) (b)

FIG. 2. Schematic diagram of the vacancy in Si or Ge (a) at a lattice site in the normal state V and (b) during migration in the transition state V'. The mobile atom is shown shaded. The purpose of this figure is to illustrate that there are exactly four broken bonds about the vacancy in the normal state V and at least six broken bonds in the transition state V'. Thus migration requires at least the enthalpy of two covalent bonds $2E^b$. By counting the total number of complete bonds in the crystal, the reader may note that a simple migration, i.e., one in which the mobile atom completes no bonds in state V', in fact, requires that three addition bonds be broken for a total of seven. With some lattice distortion the mobile atom may complete one bond so that a total of six for state V' is possible.

Regarding the possible transition states, it is evident that the vacancy can not migrate between lattice sites without breaking at least two more bonds [see Fig. 2(b)]. Therefore, there must be a contribution of $2E^b(T)$ to both $\Delta H_m(T_h)$ and $\Delta H_m(T_l)$. Indeed, empirically,

$$\begin{aligned} 2E^b(\text{Si}, T=0) &= 0.37 \text{ eV} \gtrsim \Delta H_m(\text{Si}, T_l) \\ &= 0.33 \pm 0.03 \text{ eV}, \\ 2E^b(\text{Ge}, T=0) &= 0.22 \text{ eV} \gtrsim \Delta H_m(\text{Ge}, T_l) \\ &= 0.2 \text{ eV} \end{aligned} \tag{1.28}$$

by comparison of the results of Ref. 32 with Refs. 13–15 or Eqs. (1.18) and (1.19). However, it is also evident from the geometry of the lattice that the vacancy can not migrate in the undistorted lattice without distorting its octahedral equilibrium shape. The requisite distortion is analyzed in Sec. IV and the energy of distortion $\Delta H_{v'}^m(T_h) - \Delta H_v^m$ is estimated from the observed anisotropy of the surface energy of bulk Si and Ge. This distortion energy is large (~0.8 eV), because the surface energy of Si and Ge is very anisotropic, and accounts for $\Delta H_m(T_h)$ within experimental accuracy.

However, from the observation that $\Delta H_{v'}^m(T_h) - \Delta H_v^m$ is much larger than either a typical phonon energy or a bond energy

$$\Delta H_{v'}^m(T_h) - \Delta H_v^m \gg E^b(T) \gg k\Theta, \tag{1.29}$$

it is obvious that the transition state with the distorted vacancy in an undistorted lattice (beyond nearest-neighboring atoms) is not the transition state of lowest enthalpy. It is a state of large entropy and thus low free energy at high temperatures. Therefore, the distorted-vacancy–undistorted-lattice state is likely to be the correct high temperature-transition state $V'(T_h)$, but there is also a transition state of much lower enthalpy $V'(T_l)$, which results from relaxing the lattice. Evidently, it can be relaxed down to the value indicated by Eq. (1.28). This requires that the motion of the surrounding atoms be correlated with that of the vacancy so that the entropy of this mode will be low and the frequency factor ν should be much less than ν_D.

For the convenience of the reader, the theory of the analysis of equilibrium shapes[33,37–39] is briefly reviewed and the particular case of the octahedron is worked out in Appendix A.

Readers familiar with atomic migration in metals or in ionic solids, such as the alkali halides or AgBr, will note that the nature of the vacancy migration process described here for Si and Ge is quite different from that generally observed or inferred in those noncovalent materials. This dissimilarity between what is familiar and already fairly well understood and that which is to be investigated here may distract such readers from the logic of the discussion. For the benefit of anyone so distracted, the reasons that the vacancy migration process in a covalent solid should be expected to be strikingly different from the vacancy migration process in a nearly close-packed metal or rocksalt-type crystal are outlined in Appendix B.

It is noted that the application of the macroscopic model to the calculation of vacancy formation energies and Schottky constants in heterpolar semiconductors[29] and a description of the proper thermodynamic analysis of the van't Hoff plot obtained from quenching experiments[34] will be published separately.

II. REVIEW OF EXPERIMENTAL OBSERVATIONS

High-temperature determinations of ΔH_m are principally obtained from either atomic diffusion[9,12,17,40,41] or vacancy annealing[10,16] experiments [Eqs. (1.13) and (1.14)]. These two types of experiments have two features in common: they both involve vacancies that are produced by thermal activation and they both require a rapid quench from temperatures near the melting point T^F to room temperature. However, in the annealing experiment, the purpose of the quench is to provide the initial concentration of frozen-in defects which are then annealed out by raising the temperature to a moderate value ~700 °K. On the other hand, in the atomic diffusion experiment, the purpose of the quench is to determine the concentration of vacancies $[V]/n_H$ in Eq. (1.2) actually present at temperatures near T^F for which the atomic diffusion is observed. Therefore, these two types of experiments may be considered independent of one another and the agreement between the conclusions drawn from them constitutes a valid cross check.

Now, there are two types of (single) native defects which may be present in elemental crystals, vacancies and interstitials.[18] Therefore, perhaps the first point that should be reviewed is the evidence that the high-temperature experiments actually involve vacancies rather than interstitials or a mixture of vacancies and interstitials.

An obvious point is that more energy is required to produce a Frenkel pair, i.e., one vacancy and one interstitial, simultaneously in the bulk of the sample than to produce either separately at an external surface, void, a dislocation, or at some other source. The enthalpy to create the Frenkel pair $\Delta H_v + \Delta H_i$, where ΔH_i is the enthalpy of formation of a host interstitial, is determined in electron irradiation damage experiments. One determines the minimum electron energy required to produce the pair and calculates the maximum energy that can be transferred to the atom which creates the vacancy by becoming the interstitial.[42] In Si and Ge these enthalpies (energies) are[42,43]

$$\Delta H_v(\mathrm{Si}) + \Delta H_i(\mathrm{Si}) = 11 \text{ eV}, \tag{2.1}$$

$$\Delta H_v(\mathrm{Ge}) + \Delta H_i(\mathrm{Ge}) = 16 \text{ eV}. \tag{2.2}$$

These values are several times the enthalpies of formation of the native defect observed in the quenching experiments described below. A proper analysis[34] of these quenching experiments[10,11,16] shows

$$\Delta H_v(\mathrm{Si}) = 2.4 \pm 0.2 \text{ eV}, \tag{2.3}$$

$$\Delta H_v(\mathrm{Ge}) = 2.0 \pm 0.2 \text{ eV}, \tag{2.4}$$

assuming that the observed species is indeed the vacancy rather than the interstitial. Consequently, the enthalpy of formation of the other member of the Frenkel pair, the interstitial, is

$$\Delta H_i(\mathrm{Si}) = 8 \text{ eV}, \tag{2.5}$$

$$\Delta H_i(\mathrm{Ge}) = 14 \text{ eV}. \tag{2.6}$$

The ratio of the concentration of interstitials to vacancies in equilibrium in Si will be of the order $\exp(-5 \text{ eV}/kT^F) = 1\times 10^{-15}$ or less and in Ge the ratio will be far less.

The evidence that the one member of the Frenkel pair which is produced by thermal activation is indeed the vacancy and not the interstitial is not as clear cut as one might hope. It seems that the native defects have conventionally been assumed to be vacancies on the basis of the analogy with the case in metals where it is firmly established that the dominant defect is the vacancy.[44] However, a few scientists have questioned this convention[2,45] on the grounds of the discrepancy in ΔH_m measurements discussed here and the failure to establish a completely successful correlation between the ionization levels of vacancies and vacancy complexes identified by EPR and irA in electron irradiated samples and the defects found in quenched samples. (The conventional explanation for the noncorrelation of ionization levels is that they are shifted by strain fields in the quenched samples.) There have also been observations of prismatic dislocation loops and stacking faults of an interstitial nature produced in Si by the precipitation of oxygen impurities.[46] That situation is clearly not the equilibrium condition considered here since the driving force for the interstitial loop formation is the high degree of strain generated by the precipitants. In any event, the most convincing experiment, the simultaneous measurement of sample dimension and lattice constant as a function of temperature to show that lattice sites are added to the sample as temperature rises while the number of atoms remains fixed,[44] has not yet been reported for Si or Ge. This sort of experiment will probably be required to distinguish the migrating vacancy state $V'(T_h)$ or $V'(T_l)$ from an interstitial. Although $[V(T_h)]$ is much less in Si and Ge near T^F ($\lesssim 10^{16}/\mathrm{cm}^3$) than in metals, it may be that x-ray and laser technology have now developed to the point that this experiment is practical.

However, the author believes the conventional assumption that the native defect is the vacancy is correct because of the following observations. First, the Cu-decoration phenomenon,[47–49] wherein Cu is diffused into the sample and observed to precipitate into crystalline defects upon quenching, is explained by the hypothesis that the interstitial Cu atoms precipitate onto the excess vacancies upon quenching. If one were to assume that the native defects were interstitials, no plausible explanation is apparent. Second, electron microscopy has established[50] that the dislocation loops formed in nonirradiated GaAs are produced by vacancies and no evidence of interstitials can be found. The analogy between the predominantly covalent GaAs and Si or Ge is much more convincing than the analogy with metals. Third, in low-temperature irradiation experiments, one finds that it is difficult to retain a Si or Ge interstitial as these have a great propensity to replace group-I, -II, or -III substitutional impurity atoms on a lattice site,[15,26,27,51–56] so that the impurity becomes the interstitial. This occurs even at 2 °K and apparently involves an athermal migration mechanism.[56] Although vacancies are known to associate with impurities,[26,27] no such drastic mechanism by which vacancies are expelled from the sample has been observed. It seems likely that the motivation for the expulsion of host interstitials is related to the large enthalpy of formation proposed by Eqs. (2.5) and (2.6) and that the same expulsion mechanism would prevent any appreciable concentration of host interstitials from appearing at high temperatures or in quenched samples.

There are also theoretical grounds for believing that the thermally activated native defect is the vacancy and not the interstitial. The macroscopic model,[28,29] which predicts values of ΔH_v in good agreement with the assignment of Eqs. (2.3) and (2.4), might overestimate ΔH_v somewhat, but clearly could not underestimate it by a factor of 4 or more. (That is, ΔH_v can not be much greater than the product of the empirical surface energy times the surface area of a cavity having one atomic volume.) On the other hand, one may make a plausibility argument for the magnitude of the ΔH_i values shown in Eqs. (2.5) and (2.6) by a simple extension of an argument originally proposed[57] by Blount. According to Blount, the difference in energy between the valence band and the levels produced by the host interstitial is roughly half the difference in energy between the bonding and antibonding atomic orbitals, provided atomic relaxation is neglected. If we take the average bonding to antibonding energy gap to be the dielectrically determined average band gap[58] E_g, where $E_g(\mathrm{Si}) = 4.7$ eV and $E_g(\mathrm{Ge}) = 4.3$ eV, and assume

$$\Delta H_i \approx \tfrac{1}{2} Z E_g \ , \tag{2.7}$$

where $Z = 4$ is the valence of the interstitial atom, then we estimate

$$\Delta H_i(\mathrm{Si}) \approx 9.4 \text{ eV}, \tag{2.8}$$

$$\Delta H_i(\mathrm{Ge}) \approx 8.6 \text{ eV} \ . \tag{2.9}$$

This estimate of $\Delta H_i(\mathrm{Si})$ is somewhat greater than the experimental value [Eq. (2.5)] as ought to be expected on account of the atomic relaxation about the interstitial. Presumably, $\Delta H_i(\mathrm{Ge})$ is larger than the estimate of Eq. (2.9) because of the ten d-shell electrons which must also be accommodated for a Ge interstitial in Ge. This argument seems to explain why the group-IV interstitial replaces group-I, -II, or -III substitutional impurities but not group-V substitutional impurities.[55]

Having established within a reasonable doubt that vacancies and not interstitials are produced by thermal activation, let us review the quenching procedure.[10,11,16] One starts with a lightly doped sample of known carrier concentration and heats it to a temperature T_h within a few hundred degrees of T^F. The sample is maintained at T_h long enough for an equilibrium concentrated of vacancies (in all charge states) $[V(T_h)]$ to diffuse from the vacancy sources (external surfaces, dislocations, or voids) throughout the crystal. It is found[11] that 1 min is usually sufficient for this; one does not wish to prolong the heating due to the danger that fast diffusing impurities, especially Cu, may contaminate the surface, migrate into the bulk, and affect an error in the vacancy count.[59] (In both Refs. 10 and 11 it was established that the Cu concentration was less than $10^{13}/\mathrm{cm}^3$, so that this could not have affected their results.)

Next, the crystal is quickly cooled to room-temperature freezing in most of $[V(T_h)]$. This may be accomplished in about 0.1 sec. by dropping the sample into cold ethylene glycol[11] or in about 1 sec. by radiative cooling.[10,16] Both cooling rates give essentially the same result in Ge,[10,11] where they may be compared. However, when the cooling rate is reduced so that an interval of two minutes elapses, the sample is found to have returned to its initial condition.[11] The last point is taken as convincing evidence that no impurities have been introduced from or rejected to the surface.

After quenching, $[V(T_h)]$ is measured by detecting the change in extrinsic carrier concentration with a Hall effect experiment at room temperature or below. It is observed that vacancies act as either donors or acceptors depending on the position of the Fermi level, so their effect is to reduce the net concentration of the majority carrier.

It is not clear that the defects measured at room temperature by Hall effect are simple (i.e., not complexed by impurities) single vacancies *a priori*. Indeed, it is likely that most vacancies have associated with 0 or other impurities by the time that the Hall effect experiment can be performed. However, the effect of association is only to shift the ionization energy of the donor or acceptor level somewhat.[60] The number of donor or acceptor centers is not changed by association of the vacancies and the Hall-effect measurement in the extrinsic sample determines only the number of frozen-in donor or acceptor centers. As long as this number is linearly proportional[61] to $[V(T_h)]$, the slope of the empirically determined van't Hoff plot

$$\ln\{[V(T_h)]/n_H\} = \Delta S_{\mathrm{expt}}/k - \Delta H_{\mathrm{expt}}/kT \tag{2.10}$$

will yield the correct value for the enthalpy to be used in connection with atomic diffusion experiments, even if the entropy term is in error. (ΔH_{expt} should be distinguished[34] from ΔH_v because the samples are generally intrinsic and the vacancies are generally ionized at T_h. Therefore, the ionization of the vacancy and its interaction with thermally activated carriers, as well as the temperature dependence of the band gap, must be considered in order to determine ΔH_v from ΔH_{expt}.)

Once the sample has been quenched and the number of frozen-in donor or acceptor centers determined, one may periodically reheat and recool these samples to a series of moderate temperatures T_i ($T_i \sim 700$ °K, $i = 1, 2, \ldots$), to determine the rate at which these frozen-in defects anneal out. Assuming the distance a vacancy must diffuse to find a vacancy sink (an external surface, dislocation, or void) is independent of T_i,

$$\Delta H_m = T_1 T_2 k \ln(t_2/t_1)/(T_1 - T_2), \tag{2.11}$$

where t_i is the time required to anneal out $\frac{1}{2}$ the excess concentration of vacancies above the thermal equilibrium value at T_i. One may question whether or not the observed centers and the diffusing species are actually simple, single vacancies, but, as is shown in Sec. III, the annealing rate is dominated by the species which diffuses most rapidly. Thus, $\Delta H_m(T_i)$ can not be significantly less than the value determined from Eq. (2.11), although it may be larger. Moreover, if the dominant defect species present at room temperature is the divacancy, as may be the case[27] in Si, while migration at T_i is dominated by the single vacancy, then we shall see in the next section that Eq. (2.11) will yield $\frac{1}{2}\Delta H_m$ rather than ΔH_m.

Using this vacancy defect annealing technique, Elstner and Kamprath estimated

$$\Delta H_m(\mathrm{Si}, T_i \approx 700\ ^\circ\mathrm{K}) = 1.0\ \mathrm{eV}. \tag{2.12}$$

(This value is probably too low because of the divacancy effect just mentioned.) In the same way, Hiraki found

$$\Delta H_m(\mathrm{Ge}, T_i \approx 700\ ^\circ\mathrm{K}) = 1.2\ \mathrm{eV}. \tag{2.13}$$

The author has analyzed the annealing data presented by Logan[11] in this manner and found

$$\Delta H_m(\mathrm{Ge}, T_i \approx 700\ ^\circ\mathrm{K}) = 1.0\ \mathrm{eV}. \tag{2.14}$$

The alternate method for determining ΔH_m at high temperatures is to measure the rate of atomic diffusion as a function of temperature[9] near T^F. This is usually done using radioisotopes of either the host material[12,40,41] (for self-diffusion) or of various impurities.[2,7-9,17] It is found[9] that atomic diffusion may be described by the analog of Eq. (1.3) or (1.6),

$$D_A(T) = D_0(A)\, e^{-\Delta H_d/kT}, \tag{2.15}$$

where A denotes the atomic species. Assuming that the atomic migration is indeed dominated by the vacancy mechanism, the activation enthalpy[9]

$$\Delta H_d = \Delta H_{\mathrm{expt}} + \Delta H_m + \Delta H_b(A), \tag{2.16}$$

where $\Delta H_b(A)$ is the "binding energy" of the vacancy in the nearest-neighbor position to the diffusing species A. $\Delta H_b(A)$ may be greater or less than zero for impurities, depending on the relative charge state of the impurity and the vacancy, and is identically zero for the host species according to the defintion adopted here[62];

$$\Delta H_b(\mathrm{Si}) \equiv 0, \quad \Delta H_b(\mathrm{Ge}) \equiv 0\ . \tag{2.17}$$

With this convention, Eq. (2.14) reduces to Eq. (1.2) for self-diffusion when the proper entropy factor, ΔS_{expt}, is incorporated into $D_0(A)$.

For the case of Ge (Ref. 12)

$$\Delta H_d(\mathrm{Ge}) = 3.0\ \mathrm{eV}, \tag{2.18}$$

whereas $\Delta H_{\mathrm{expt}} = 1.9$ or 2.0 eV according to Hiraki[10] or Logan,[11] respectively. Therefore, Eqs. (2.16) and (2.17) imply $\Delta H_m = 1.1$ or 1.0 eV and the agreement with Eqs. (2.11) and (2.14) is entirely satisfactory. It is concluded that

$$\Delta H_m(\mathrm{Ge}, T_h) = 1.0 \pm 0.2\ \mathrm{eV}. \tag{2.19}$$

Some of the diffusion data in Si, particularly the self-diffusion data,[41] are complicated by effects occurring near external surfaces.[63] (As only short lived isotopes of Si are available for use as radiotracers, all Si self-diffusion data are obtained within about 3 μ of a surface.[41,63]) However, with careful analysis of impurity diffusion data and reasonable estimates of $\Delta H_b(\mathrm{I})$, one obtains[9,17] values for $\Delta H_m(\mathrm{Si}, T \approx 1500\ ^\circ\mathrm{K})$ of between 1.0 and 1.5 eV. In view of the annealing results[16] [Eq. (2.12)], it is concluded that

$$\Delta H_m(\mathrm{Si}, T_h) = 1.2 \pm 0.3\ \mathrm{eV}. \tag{2.20}$$

Note that if surface effects[63] are ignored and the self-diffusion data are taken at face value, then one estimates from Eq. (2.16) that $\Delta H_m(\mathrm{Si}, T_h) \simeq 2.5$ eV.

In order to complete this review of high-temperature evaluations of ΔH_m, one further experiment should be mentioned. Smith and Holland have analyzed[64] the results of a.c. conductivity measurements on a Ge whisker[65] in the range $700 \leq T \leq 1000\ ^\circ\mathrm{K}$. It is found that the resistivity lags the current-produced temperature fluctuations by a factor inversely proportional to frequency below 100 Hz. Smith and Holland interpret this lag as resulting from the diffusion of vacancies in from and out to the surface of the whisker as the temperature fluctuates. They conclude that $D_v \simeq 10^{-3}$ cm^2/sec and essentially independent of T in this range. [One might force a fit[2] to their data with a $\Delta H_m \lesssim 0.2\ \mathrm{eV} = \Delta H_m(\mathrm{Ge}, T_l)$.] Note that this is an order of magnitude greater than the value of $D_v(1000\ ^\circ\mathrm{K})$ shown in Fig. 1. At $T = 700\ ^\circ\mathrm{K}$, the discrepancy is about four orders of magnitude. Whereas a whisker sample is free of dislocations, it may contain voids, defect clusters, or impurity clusters. Thus, the discrepancy in the estimate of the magnitude of the diffusivity may lie with the presumption that the species producing the time dependent resistivity comes from the surface. It is also possible that this species is not the vacancy but a fast diffusing impurity, such as Cu, or that the effect results from the ionization of some impurity which remains in the bulk and has deep levels. There is not adequate information available to resolve this question, but the conclusion[2] that $\Delta H_m(T_h) = \Delta H_m(T_l)$ is unwarranted.

Turning to the low-temperature evaluation of ΔH_m, we find that these are always performed on samples wherein the vacancies were created by irradiation at low temperature. It is preferable that fast (≈ 1.5 MeV) electron irradiation be used as this method minimizes the disruption of the lattice about the vacancies.[26] As mentioned in connection with Eqs. (2.1) and (2.2), the irradiation actually produces vacancy-interstitial pairs, but the interstitials are rapidly expelled from the sample[26,51–56] and seem to have no effect on experiments involving the vacancies.[26]

Now, in order to be observed in an EPR experiment, the simple, single vacancy must be ionized. Therefore, $\Delta H_m(T_l)$ is determined for the neutral single vacancy in Si by alternately annealing the sample at $T_l \approx 160$ °K and illuminating it at about 20 °K with light which ionizes the neutral vacancies so that their concentration may be determined. Given the concentration of impurities which trap vacancies, one analyzes the data using the diffusion limited kinetics theory as described in Sec. I, Eqs. (1.16)–(1.18), and finds[15,26] for the neutral vacancy

$$\Delta H_m(\mathrm{Si}, T_l) = 0.33 \pm 0.03\ \mathrm{eV}. \qquad (2.21)$$

In the irA experiments[13,14] it is not necessary to cycle the temperature because the vacancies may be monitored continuously. However, one is not sure which charge state of the vacancy is being observed because the ionization levels of the vacancy are not firmly established and the infrared spectra of the different charge states have not been distinguished. Moreover, the distribution of charge states may change if the Fermi level changes during the anneal as a result of the concomitant shift in ionization levels. On the other hand, the empirical estimate of ΔH_m does not vary more than about 0.1 eV as a function of the initial position of the Fermi level.[13,14] The value one would obtain from Whan's annealing data, Fig. 2 of Ref. 13, is, if anything, somewhat less than the 0.2 eV found by Klontz and MacKay.[14]

Regarding the possible dependence of ΔH_m on the charge state of the vacancy, it should be clear that the difference in ΔH_m's can not exceed the difference of the energies of the charge states in question relative to the Fermi level. Let us denote the enthalpy of formation of the two charge states of the vacancy, say V^0 and V^+, as ΔH_v^0 and ΔH_v^+ at the lattice site and as $\Delta H_{v'}^0$ and $\Delta H_{v'}^+$ in the respective transition states. As long as the lattice site is the normal, i.e., more probable, position for the vacancy,

$$\Delta E \equiv \Delta H_v^+ - \Delta H_v^0 \qquad (2.22)$$

is the difference of the energies of these charge states relative to the Fermi level. Let us suppose $\Delta E > 0$, so that V^0 is the predominate species in equilibrium. As with Eq. (1.5), we may define

$$\Delta H_m^0 \equiv \Delta H_{v'}^0 - \Delta H_v^0 \ , \qquad (2.23)$$

$$\Delta H_m^+ \equiv \Delta H_{v'}^+ - \Delta H_v^+ \ . \qquad (2.24)$$

However, if $\Delta H_m^0 - \Delta H_m^+ > \Delta E$, then the predominate transition state for V^0 will be $V^{+\prime}$, rather than $V^{0\prime}$, and the observed value of the enthalpy of migration for V^0 will be

$$\Delta H_m(V^0) = \Delta H_{v'}^+ - \Delta V_v^0 \leq \Delta H_m^+ + \Delta E \ . \qquad (2.25)$$

Assuming that one can measure ΔH_m for the minority species, here assumed to be V^+, the same argument shows that it is not possible for this value to exceed ΔH_m^0 nor to be less than $\Delta H_m(V^0) - \Delta E$.

In both Si and Ge four charge states of the vacancy are observed[2,36] within the fundamental band gap, which is never greater than 1.1 eV. Although the levels of these four charge states are not accurately known, it is clear that Eq. (2.25) will limit possible variation of ΔH_m to a few tenths of an eV. Thus, the discrepancy between $\Delta H_m(T_h)$ and $\Delta H_m(T_l)$ can not be explained by the fact that a statistical distribution of vacancy charge states are present in the high-temperature experiments. [As the material is generally intrinsic at T_h in these experiments,[34] the Fermi level is near the middle of the gap and the most probable charge states are V^+(Si) and V^-(Ge).]

Moreover, on theoretical grounds, one might expect

$$|\Delta H_m^+ - \Delta H_m^0| \ll \Delta E \ , \qquad (2.26)$$

because the charge associated with the various charge states is generally spread over many unit cells[66] and thus should not affect either the local electron density or the covalent bonding around the vacancy too severely. One might also guess that the largest ΔH_m would occur for the neutral charge state of the vacancy. This is because the neighboring covalent bonds which must be broken for migration might be expected to suffer the least disruption when the vacancy is neutral.

Indeed, Watkins estimated[15] $\Delta H_m(T_l)$ in a sample with the Fermi level near the conduction band edge, for which V^{--} is taken to be the migrating species, as

$$\Delta H_m(\mathrm{Si}, V^{--}, T_l) = 0.18\ \mathrm{eV}. \qquad (2.27)$$

[Although in this case, the estimate is probably correct, one should be careful to verify that the decay of the simple state concentration of the charged vacancies is really exponential whenever the apparent ΔH_m is approximately half the value of ΔH_m observed for the neutral vacancy (see Ref. 24).]

It is found that the rate of annealing of radiation induced vacancy-impurity complexes below 500 °K is greatly affected by the position of the Fermi level. However, these effects have been analyzed by Kimerling *et al.*, with the aid of observations of the effect of infrared light on the process.[67,68] They demonstrate that it is the binding energy of the complex, i.e., $\Delta H_b(A)$ in Eq. (2.16), which varies with the charge state of the vacancy, while ΔH_m remains sensibly constant.

At high temperatures, it is found that the position of the Fermi level has no significant effect[10–12] on the apparent value of $\Delta H_m(T_h)$ in Ge. (This would follow from the fact that the vacancy annealing and atomic diffusion experiments yield the same value for ΔH_m.) In Si self-diffusion experiments, it is observed that the apparent ΔH_d of Eqs. (2.15) and (2.16) is reduced from its anomalous value of about $\Delta H_{\text{expt}}+2.5$ eV to the expected value of $\Delta H_{\text{expt}}+[\Delta H_m(T_h)\approx 1.3 \text{ eV}]$, when the sample is heavily doped either n type or p type.[41] This probably results from screening of the effects of the external surface which cause the anomalous self-diffusion.[63] Other experiments indicate no such variation of ΔH_m with the Fermi level.[17]

This concludes a brief review of the experimental situation. The author does not pretend to have done justice to such a vast field containing so many excellent works with this rather spotty treatment; the only purpose of this section was to mention the experimental facts necessary to the theoretical argument.

III. SPECIES OBSERVED AT T_h IS IN FACT THE SINGLE SIMPLE VACANCY

It is probably due to the elegance and thoroughness of the EPR experiments[15,26,27] that, to the author's knowledge, no one has seriously attempted to discredit the empirical conclusions regarding $\Delta H_m(T_l)$ just discussed. These experiments provide a great deal of detailed information about the processes observed. The high temperature experiments, on the other hand, yield almost no microscopic information about the processes involved. It is probably for this reason that those who presume that ΔH_m can not vary with temperature have sought to remove the discrepancy between measurements at T_l and at T_h by hypothesizing that the experiments at T_h do not observe vacancies but instead some bound complex, an impurity bound to a vacancy or a divacancy perhaps, which is assumed to migrate with a larger activation enthalpy.[1–9] It is the purpose of this section to show that the condition that vacancies be present in thermal equilibrium in these experiments suffices to discredit all such hypotheses and to support Eqs. (2.19) and (2.20) very strongly.

The argument will be based on the law of mass action.[22] For the formation of divacancies from single vacancies, one has the equilibrium condition

$$V+V\rightarrow V_2,\qquad(3.1)$$
$$[V_2]n_H/[V]^2=K_{2v}=\exp[-(\Delta H_{2v}-T\Delta S^f_{2v})/kT],$$

where ΔS^f_{2v} and ΔH_{2v} are the standard entropy[69] and the enthalpy of formation of the divacancy from monovacancies. (Note that $\Delta H_{2v}<0$.) As in Eq. (1.2), n_H is the concentration of lattice sites. For the formation of the complex C from impurity I and the vacancy, one has the condition

$$I+V\rightarrow C,\qquad(3.2)$$
$$[C]n_H/[I][V]=K_c=\exp[-(\Delta H_b-T\Delta S^f_c)/kT],$$

where ΔH_b is the binding energy, as in Eq. (2.16) and ΔS^f_c is the standard entropy of formation of the complex. (Note $\Delta H_b<0$.)

One should immediately note that the reactions which form divacancies or complexes, Eqs. (3.1) and (3.2) *do not reduce the concentration of simple, single vacancies* $[V]$. As divacancies and complexes are formed, more single vacancies are generated at the vacancy sources so that $[V]$ remains at its equilibrium value. Thus, because the chemical potential of vacuum is constant, the total number of vacancies is not conserved for any of these reactions. Consequently, the rate of atomic diffusion can not be less than the rate produced by a single vacancy mechanism alone, although it could be greater if some other mechanism also contributes significantly. Therefore, the hypothesis that single vacancies tend to condense into divacancies or some other complex species at high temperatures, does not help to resolve the problem of the absolute magnitude of the diffusivity noted in Sec. I [Fig. 1 and Eq. (1.22)]. One would still have to assume either that the low temperature estimates of D_0 are more than two orders of magnitude too large or that the frequency factor ν in Eq. (1.7) decreases with increasing temperature so that the assumption of Eq. (1.22) is violated.

From Eq. (3.1) and the fact that $[V]\lesssim 10^{-6}\,n_H$ is the equilibrium concentration at T_h it follows that

$$[V_2]\ll[V]\qquad(3.3)$$

for all reasonable values[27] of $\Delta F_{2v}=\Delta H_{2v}-T\Delta S^f_{2v}$. The condition that $[V_2]=[V]$ would require that ΔF_{2v} $\Delta F_{2v}=-\Delta F_{\text{expt}}=-\Delta H_{\text{expt}}+T\Delta S^f_{\text{expt}}\approx-\Delta H_{\text{expt}}$. However, because the divacancy is presumed to migrate slowly and yet still dominate the experiment, one must postulate that $[V_2]\gg[V]$, which requires that ΔF_{2v} be an even larger negative quantity. Thus, in order for the divacancy hypothesis to be consistent in Ge, the reaction $0\rightarrow V_2$ must require only 2.0 eV of enthalpy (i.e., the observed ΔH_{expt}), while $0\rightarrow V$ must require more than 2.8 eV (in order for the single vacancies not to dominate experiments). The situation for divacancies in Si would be $0\rightarrow V_2-2.5$ eV, while $0\rightarrow V-(\Delta H>3.7$ eV). These conditions would require that the vacancy-vacancy binding energy exceed 3.6 eV for Ge and 4.9 eV for Si, whereas the observed values[2,27] are about 1.0 eV $<\Delta H_{\text{expt}}$.

The hypothesis that the dominate species for migration and doping at T_h is a complex C of the single vacancy with some impurity I has not been nearly as popular as the divacancy or the extended defect hypotheses.[1-9] For the sake of completeness, however, arguments against such a hypothesis are presented here. (Recall that the Seeger-Chick hypothesis[2] of an "extended defect" consisting of a single vacancy or interstitial and several *host* atoms was discussed and shown to be untenable in Sec. I.)

The condition that migration of C dominate the atomic migration process requires

$$\nu_C \exp\{-[\Delta H_m(C) - T\Delta S_m(C)]/kT\}[C] > \nu \exp[-(\Delta H_m - T\Delta S_m)/kT][V], \quad (3.3)$$

where, as in Eq. (1.3), ν, ΔH_m, and ΔS_m refer to the single vacancy and ν_C, $\Delta H_m(C)$, and $\Delta S_m(C)$ have the corresponding definitions for the complex C. The hypothesis that $\Delta H_m(C)$ exceed ΔH_m by about 1.0 eV for Si and 0.8 eV for Ge requires that

$$\nu_C \exp[\Delta S_m(C)/k][C] \gg \nu \exp(\Delta S_m/k)[V]. \quad (3.4)$$

Indeed, the left-hand side of inequality (3.4) would have to be more than 10^5 times the right-hand side to account for the defect annealing data.[10,16] As noted in Eqs. (1.9) to (1.11), one expects $\nu_C \ll \nu$ when ν represents a simple mode of migration for the single vacancy. Therefore, this factor makes the hypothesis that C dominates even less plausible. The measured value of ΔS_m, whatever the mechanism, at high temperatures is about $7k$, assuming Eq. (1.10) so that $\nu = \nu_D$. Swalin has argued from geometrical consideration that this value of ΔS_m should be expected for the simple migration of the single vacancy.[70] Therefore, $\Delta S_m(C) - \Delta S_m$ should not be large enough to account for Eq. (3.4). Consequently, one must suppose that

$$[C(T^F)] \gg [V(T^F)], \quad (3.5)$$

if he would maintain that C dominates the atomic migration near T^F.

Now if the complex $C = I + V$ is bound at temperatures near T^F, it will certainly be bound at all lower temperatures. One would have to conclude from Eq. (3.5) that the species responsible for the Cu decoration effect[46-49] and for the growth of dislocations is C and not V. If this were the case, then one would expect to find the impurity species I, in conjunction with the substitutional Cu impurities in the Cu-decoration experiment and with the dislocations in electron-microscope experiments.[50] Neither experiment supports these expectations.

One may further note from Eq. (3.2) that the hypothesis that C dominates high temperature experiments requires that $-\Delta H_b$ be large, i.e., equal to or greater than ΔH_{expt}. Furthermore, because the empirical estimates of $\Delta H_m(T_h)$ are insensitive to the position of the Fermi level, as noted at the end of Sec. II, one would have to assume ΔH_b is independent of the charge state of the vacancy. Both these requirements are implausible and contrary to existing observations.[17,26,60,67,68]

Moreover, Eq. (3.5) would imply that the enthalpy of defect formation measured in the quenching experiments would be that of the bound complex C and not of V. Then ΔH_v would have to exceed this empirical value by the (large) binding energy $-\Delta H_b$. But the empirical value of ΔH_v is in good agreement with the macroscopic model[28,29,71] and, as was noted in Sec. II with regard to question of interstitials, this model certainly should not underestimate the true value of ΔH_v.

Therefore, the author is confident that the usual conclusion

$$[C(T^F)] \ll [V(T^F)], \quad (3.6)$$

is correct. The only possible reservation might be that a complex of interstitial oxygen and a vacancy might form at the surface of the sample and diffuse in and out as the vacancy is assumed to do. The high-temperature experiments were performed in the presence of air. Thus, it might be worthwhile to repeat them with oxygen-free samples in vacuum.

Of course, when $[V]$ is much larger than the equilibrium value at the prevailing temperature, then, according to Eqs. (3.1) and (3.2), $[V_2]$ and $[C]$ may become large compared with $[V]$. This condition obtains in the EPR and irA experiments and also in the quenched samples at room temperature or below.

Consider now whether the annealing experiments actually measure ΔH_m for the single vacancy if the center observed in the Hall-effect measurement is in fact a bound complex of the vacancy. The point here is that for all reasonable binding energies, the rate of association and dissociation of the complexes is sufficiently rapid at the annealing temperature that Eqs. (3.1) and (3.2) always prevail. For example, using the estimate that for $P-V$ in Si,[26] $\Delta H_b = 1.0$ eV, taking $kT = 0.05$ eV for $T = 580$ °K, and taking the dissociation attempt frequency $\nu_d = 10^{13}\ \text{sec}^{-1} \approx \nu_D$, one estimates the dissociation time t_d to be

$$t_d = \exp(\Delta H_b/kT)/\nu_d = 5\times 10^{-5}\ \text{sec}. \quad (3.7)$$

Annealing times, t_i, in Eq. (2.11) are typically several minutes.[10,11,16] Therefore, the rate of change in concentration of the various complex species is simply related to the rate of change in concentration of the most rapidly diffusing component, which is assumed to be the single vacancy. It follows from Eqs. (3.1) and (3.2) that

$$\frac{d\ln[V_2]}{dt} = \frac{2d\ln[V]}{dt} \tag{3.8}$$

and

$$\frac{d\ln[C]}{dt} = \frac{d\ln[V]}{dt} \ . \tag{3.9}$$

Thus, Eq. (2.11) will yield the correct value of ΔH_m from annealing data if the observed center is either the single vacancy or a complex containing one vacancy; if the observed center is a divacancy or contains two vacancies, then the value calculated from Eq. (2.11) is $\Delta H_m/2$. These experiments will not overestimate ΔH_m unless ΔH_b is so large that the rate of anneal is governed by the rate of dissociation [Eq. (3.7)].

Is it possible that some impurity in these samples has such a large binding energy that the rate of dissociation governs the rate of anneal at the relevant temperatures[10,11,16] ($T_l \approx 700$ °K)? This conjecture would require a $\Delta H_b \approx -2$ eV. If this were the case, then C would be the predominate species near T^F and the objections raised against the hypothesis that C mediates atomic diffusion would apply. Moreover, the values of ΔH_m estimated from Eq. (2.11) would then be $-\Delta H_b$; thus the observed rate of defect annealing is too rapid for this hypothesis to be correct.

In view of all this evidence from experiments which have been repeated and confirmed several times, it seems one must conclude that

$$\begin{aligned} &\Delta H_m(\mathrm{Si}, T_h) \approx 4\Delta H_m(\mathrm{Si}, T_l), \\ &\Delta H_m(\mathrm{Ge}, T_h) \approx 5\Delta H_m(\mathrm{Ge}, T_l) \ . \end{aligned} \tag{3.10}$$

Furthermore, this must occur because the single vacancy has at least two modes of migration, or transition states, which may predominate at different temperatures.

IV. MACROSCOPIC MODEL CALCULATIONS OF ΔH_m

In previous theoretical work,[70–74] the question of ΔH_m has generally been treated as an adjunct to the question of ΔH_v. With the exception of Swalin's Morse potential treatment,[70] which is the one most often quoted by experimentalists, these discussions use wave mechanical methods and treat the vacancy in a manner analogous to that used to treat impurity atoms. As was discussed in Ref. 28, such methods are not satisfactory for use as a predictive tool[75] in high electron density solids, like Si, Ge, or Al, even when the discussion is confined to ΔH_v. Reasons for this are that (a) the results calculated by such methods are so sensitive to the values of the crystal potential assumed that one may calculate any value he desires[76–78] for ΔH_v, including values $\Delta H_v < 0$ which would imply that the crystal is unstable; (b) unlike the impurity problem, for which the electron density is high everywhere, the electron density is very low at the center of the vacancy so that the nearly-free-electron screening approximation and the crystal potential, which was deduced from the bulk, are most probably unrealistic when used to treat vacancies with these methods; (c) they neglect correlation terms which are apparently large[30,31]; (d) the relaxation of the atoms, in particular, the Jahn-Teller distortion, about the vacancy is so large[35,36] that these calculations of ΔH_v and ΔH_m, which either neglect the distortion or treat it as a small perturbation, can not be realistic. (For the negatively charged state of the vacancy in Si, it has been shown that the observed ground state can not be obtained by linear perturbation theory.[35,36]) Quantum calculations of impurity states[66] have yielded good agreement with optical spectra, but similar calculations of the ionization spectrum of vacancies have not been presented despite the importance of this property to the conductivity of the material.

If these problems with the microscopic, wave-mechanical treatments are severe for ΔH_v, they are clearly worse for ΔH_m because the difference between $\Delta H_{v'}$ and ΔH_v is only (10–20)% of ΔH_v at low temperature [see Eqs. (1.18) and (1.19)]. Moreover, none of these discussions has given any clue as to why ΔH_m should vary with temperature.

In view of this lack of success in accounting for $\Delta H_m(T_l)$ and $\Delta H_m(T_h)$ in Si and Ge encountered when the problem is considered from the microscopic point of view, let us consider it from the point of view of the macroscope model introduced in Ref. 28.

Let us begin by reviewing the justification for treating an atomic vacancy as a macroscopic cavity. Recall that it was assumed in Ref. 28 that the volume of the vacancy at lattice site V_v was the atomic volume Ω,

$$V_v = \Omega = \tfrac{4}{3}\pi r_w^3 = \tfrac{1}{8}a^3 \ , \tag{4.1}$$

where r_w is the Wigner-Seitz atomic radius and a is the lattice constant. The dimensions of the cavity implied by Eq. (4.1), e.g., r_w, are large on the scale of effects which dtermine the cohesive energy and surface energy when the electron density is as high as it is in Si and Ge.

The reasons for this are that as noted in Sec. I, Eq. (1.26), there are two primary contributions[28–32,71] to the surface energy per unit area E_S. Here, these will be denoted E_S^m and E_S^b,

$$E_S = E_S^m + E_S^b \ . \tag{4.2}$$

E_S^m is the surface energy per unit area which would obtain for a simple metal of the same electron density. In fact, E_S^m is determined from observation of the metallic, liquid phase of the semiconductor,[71] but its value for liquid Si and Ge and some 50 other liquid metals is accounted for in terms of the long-ranged electron-electron correlation energy in Refs. 30 and 31. The effect giving rise to E_S^m may be visualized as the spilling of the bulk electron density into the vacuum exterior to the sample or in the center of the vacancy. As the electron density decays exponenitally,

$$\rho(r) = \rho_0 e^{-k_s r} , \tag{4.3}$$

where k_s is the linearized Thomas-Fermi screening wave number,[30] and as $k_s r_w$ equals 3.5 and 3.6 for Si and Ge, respectively, we see that

$$\rho(r_w) \cong 0.03 \rho_0 . \tag{4.4}$$

Thus, the overlap of charge density spilling from the opposing surfaces of the vacancy is quite negligible provided Eq. (4.1) is approximately correct. The remaining contribution to E_S : E_S^b , is the result of the directional, covalent bonds in the semiconducting phase which are broken when a surface, either exterior or about a vacancy, is created;

$$E_S^b = N_S^b E^b , \tag{4.5}$$

where N_S^b is the number of broken bonds per unit area and E^b is the energy per bond of the directional, covalent bonding. The values of E^b are calculated as a function of composition, temperature and pressure in Ref. 32, where it is shown that the calculation gives a good account of the observed pressure-temperature phase diagrams, alloying properites, and impurity distribution coefficients of these semiconductors. Covalent bonding results from the constructive interference of orbitals centered on the bonded atoms so that the effects producing E_S^b decay with distance even more rapidly than those producing E_S^m.

It is important to understand that E^b is the energy per covalent bond of the solid over and above that which would obtain if the solid were a simple metal and not the total cohesive energy per bond of the solid. (The semiconductor is viewed as a perturbed state of the metal for which a gap has appeared at the Fermi level.[32]) This is in contradistinction to a widely used model in which all the cohesive energy is ascribed to the covalent bonds so that the heat of atomization ΔH_a is just two bond energies.[79] In the present model,

$$\begin{aligned} 4.6\ \text{eV} &= \Delta H_a(\text{Si}) \gg 2E^b(\text{Si}, T=0) = 0.37\ \text{eV}, \\ 3.9\ \text{eV} &= \Delta H_a(\text{Ge}) \gg 2E^b(\text{Ge}, T=0) = 0.22\ \text{eV}, \end{aligned} \tag{4.6}$$

just as

$$\Delta H_a \gg \Delta H_f \sim 0.5\ \text{eV}, \tag{4.7}$$

where ΔH_f is the heat of fusion, which accompanies the transformation from the semiconducting phase to the metallic liquid phase.[32] (The notion that all the surface and cohesive energy may be accounted for in terms of nearest neighbor bonds may be refuted by the observation that when crystals are prepared in a shape other than their equilibrium shape, a spontaneous bending of the crystal is observed.[80] Such bending is too slight to change the number of broken bonds on any face and clearly strains the bulk lattice. Therefore, this effect can not be explained with an additive bond model. It must result from long-ranged forces.)

It will here be assumed that the volume of the vacancy does not change as it migrates. Thus, the volume of the vacancy at the saddle point between lattice sites V_v' is the same as the volume at the lattice site

$$V_v' = V_v = \Omega . \tag{4.8}$$

In view of the importance of Eqs. (4.1) and (4.8) to the values of ΔH_m which will be calculated, it seems worthwhile to rationalize these assumptions as well as possible. It must be admitted that one can not deduce the microscopic features of the vacancy from macroscopic observations and that the size of the vacancy is not conclusively given from microscopic observations.[81] However, it is quite clear that the macroscopic volume to be associated with the vacancy is exactly Ω. This is because when a number of vacancies N_v is induced by diffusion from an exterior surface into an initially perfect crystal containing N_A atoms, the number of lattice sites in the crystal increases from N_A to $N_A + N_v$ and the macroscopic dimensions of the crystal expand.[44] As these lattice sites are equivalent, the volume per vacancy is $\Omega = \frac{1}{8}a^3$ irrespective of the sense of the local strain about the vacancy. Furthermore, as the vacancy migrates from one lattice site to another in a time comparable to an average phonon period, the macroscopic dimensions of the sample can not change during migration because the migration time is obviously not long enough for any phonons from the vacancy to reach the surface. Therefore, it seems that the most reasonable assumption that can be made regarding V_v and V_v' is that they are equal to each other and to Ω, i.e., Eqs. (4.1) and (4.8).

At this point, some readers may protest that because the unperturbed diamond lattice is so "open," one can not center a volume Ω on a lattice site with any reasonable shape and maintain the notion that the atoms are more-or-less hard spheres with radius equal to the (tetrahedral) covalent radius,

$$r_c = \tfrac{1}{2} d_0 = \tfrac{1}{8} 3^{1/2} a , \tag{4.9}$$

without forcing the nearest-neighboring atoms around the vacancy back. (See Fig. 3.) Thus, because

$$r_w \cong 1.43 r_c, \qquad (4.10)$$

it seems the atoms around a vacancy must be displaced *away* from the center of the vacancy, i.e., the local strain field must be compressional, if the volume of the vacancy is to be as large as Ω. This sense of the displacement is that predicted by Pauling's bond order relations and will seem natural to readers with a chemical background. This may be contrary to the expectations of other readers to whom it may seem more likely that nearest neighboring atoms should be displaced toward the center of the vacancy as if they were spilling into the void.

In order to counter the objection that the sense of the displacement proposed here seems unlikely, the somewhat circumstantial evidence which supports the contention that the displacement is away from the vacancy will be summarized here. In the first place, it is often noted that the bulk lattice constant a as measured by x-ray diffraction, varies with $[V]$.[44] In the case of tetrahedral semiconductors,[82,83] *a increases* linearly with $[V]$,

$$a([V]) = a(0) + \alpha[V], \qquad (4.11)$$

with $\alpha > 0$. Note that the value of a measured by x-ray diffraction is determined by the ordered regions of the crystal and not by any disordered region which may exist about the vacancy. This behavior seems analogous to the linear variation of virtual lattice constants in substitutional alloys[84] (Vegard's law); if so, then the strain field must be compressional. Secondly, from observations of the interstitial atoms created concomitantly with vacancies by electron irradiation, it has been concluded that there is a barrier against the direct annihilation of vacancies and interstitials.[51–54] As the vacancy-interstitial pair is created from a neutral atom, the vacancy and interstitial should either be both neutral or in ionization states of opposite charge. Therefore, the simple coulombic interaction should either be negligible or attractive. It seems the barrier against annihilation must arise from interaction of the strain fields about the two defects. Therefore, either both strain fields are compressional or both are dilational and it seems most likely that an interstitial atom produces a compressional strain field. Thirdly, this contention is in agreement with the conclusion of the most reliable molecular-orbital calculations available[81] (principally for diamond). Fourthly, if it is granted that the vacancy is so large that one may ascribe to its surfaces the properties of the exterior surfaces, then it would seem it is the (111) surfaces which are relevant because the four nearest-neighboring atoms (in the undistorted lattice) are in the [111] bonding directions from the vacancy. For the exterior (111) surfaces, it has been concluded from both experiment[85] and theory[86] that the displacement of the surface atoms is into the bulk and not toward the vacuum. These four pieces of evidence that neighboring atoms are displaced away from the vacancy so that a volume Ω can be accommodated, seem very convincing.

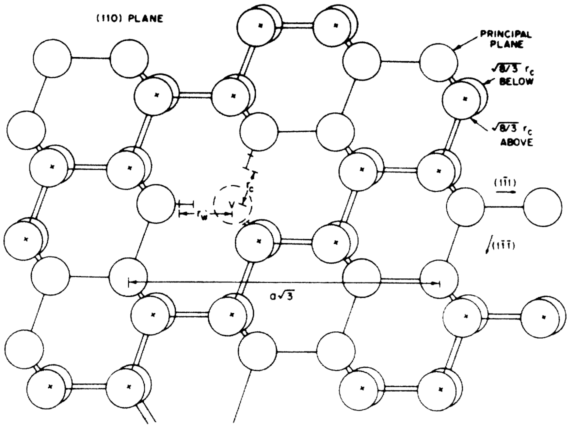


FIG. 3. Undistorted diamond crystal structure. Three (110) planes are shown—the principal plane wherein the vacancy V resides and will migrate and the next plane above and below. A body diagonal of the unit cube, which has length $a\sqrt{3}$, is indicated. The covalent radius, r_c and the Wigner-Seitz atomic radius r_w are compared.

Let us now consider the question of the shape of the vacancy cavity. In Ref. 28, it was simply assumed that the vacancy was spherical but that the surface energy appropriate to a (111) macroscopic surface[87] obtained at all points of the vacancy surface. That is not a self-consistent assumption because the macroscopic surface energy of Si and Ge must be grossly anisotropic, i.e., $E_S(ijk)$ is not independent of the Miller indexes (ijk) of the surface plane, in order to account[33,37–39] for the observed octahedral equilibrium shape of these crystals. In fact, the (111) surface, which is the only surface to appear on the equilibrium shape,[88] must be the surface of lowest (free) energy per unit area and, in particular,

$$\begin{aligned} E_S(110) &\geq (1.5)^{1/2} E_S(111), \\ E_S(100) &\geq 3^{1/2} E_S(111). \end{aligned} \qquad (4.12)$$

(See Appendix A.) If, in accord with the premise that the vacancy is so large that one may ascribe all the properties of the exterior surfaces to its surfaces, we assume the same anisotropy of surface energy obtains on the vacancy, then the equilibrium shape of this effectively macroscopic cavity will also be an octahedron.[33] In this case, it will cost energy comparable to ΔH_v to distort the vacancy from its equilibrium shape even if the volume of the vacancy and the number of broken bonds remain constant.

Let us begin to consider whether or not Eq. (4.12) should apply to the vacancy by generalizing Eq. (4.2) to account for the anisotropy of the macroscopic surface energy per unit area. Thus,

$$E_S(ijk) = E_S^m(ijk) + E_S^b(ijk) , \tag{4.13}$$

where, again, $E_S^m(ijk)$ is the contribution from long-ranged metallic forces, which are taken to be due to the long-ranged part of the electron-electron correlation,[30,31] and $E_S^b(ijk)$ is the contribution from short-ranged directional covalent forces. It is here assumed that $E_S^b(ijk)$ is simply the number of broken covalent bonds per unit area of surface $N_S(ijk)$ times the energy per bond of the semiconducting phase relative to the simple metallic reference phase[32] at temperature T (and a pressure of one atmosphere) $E^b(T)$,

$$E_S^b(ijk) = N_S(ijk)\, E^b(T) . \tag{4.14}$$

[By assuming Eq. (4.14), one neglects the effects of whatever distortion may exist among the unbroken bonds of atoms near the surface. In the present model, the appropriate correlation of $E_S^b(ijk)$ would be the order of a typical phonon energy, 30 meV per surface atom, whereas two $E^b(T)$ is of order 300 meV. The effect of such a distortion on $E_S^m(ijk)$ may be larger, especially if it alters the local charge density,[89] but we shall manage to avoid having to face that problem in the present model. If one were instead to ascribe all the cohesive energy of the solid to the bonds, then the problem of distortions would be very important[70] and difficult to manage properly.]

By simple geometry, one finds

$$E_S^b(110) = (1.5)^{1/2} E_S^b(111), \qquad E_S^b(100) = 3^{1/2} E_S^b(111) . \tag{4.15}$$

Therefore, the anisotropy of the long-ranged contribution $E_S^m(ijk)$ must be at least as severe as that of the total, $E_S(ijk)$, if Eq. (4.12) applies. Therefore,

$$E_S^m(110) \geq (1.5)^{1/2} E_S^m(111) , \qquad E_S^m(100) \geq 3^{1/2} E_S^m(111) . \tag{4.16}$$

To the author's knowledge, no derivation of Eq. (4.16), and therefore of the equilibrium shape of diamond, Si and Ge, has yet been presented.[90] A straightforward derivation in terms of the theory of Refs. 30 and 31 would be very difficult as it would require a knowledge of the full dielectric tensor[91] $\epsilon(G+q, G'+q, \omega_{sp})$ for all reciprocal lattice vectors G and plasmon wave vectors q at the surface plasmon frequency $\omega_{sp}(q)$. However, for the purposes of this paper, we may simply take the octahedral equilibrium shape and Eq. (4.16) to be the observed fact for the macroscopic exterior surfaces.[88]

In order to ascertain if it is reasonable to assume that Eqs. (4.1) and (4.16) apply to a single atomic vacancy, let us consider what one would predict for the nature of the vacancy at a lattice site using this model. In Fig. 3 we see the undistorted diamond structure. Three (110) planes are shown—the principal plane, where the vacancy is assumed to appear and to migrate, and one plane $\sqrt{\frac{8}{3}}\, r_c$ above the principal plane, and one plane $\sqrt{\frac{8}{3}}\, r_c$ below it, which contain some of the atoms that are bonded to those in the principal plane. All atoms bonded to those in the principal plane lie in one of these three planes. As can be seen, the atoms above and below the principal plane occur in pairs one behind the other. The initial position of the vacancy is taken to be the site labeled V in Fig. 3. No atomic adjustment has yet occurred. The magnitude of the Wigner-Seitz atomic radius r_w, which would be the radius of the vacancy if it were to be spherical, is compared with the covalent radius r_c. While it is not clear that the macroscopic concept of the vacancy surface can be given a rigorous microscopic definition, this might be thought of as the surface where the electron density falls to some critical fraction (near $\frac{1}{2}$) of the value that would obtain in the perfect crystal at the corresponding point.

In Fig. 4 we see the vacancy V as a perfect octahedral cavity of volume Ω centered at the lattice site. The extent of this octahedral cavity is also compared with r_c. If we assume that the surface of the vacancy must remain one covalent radius from any of the neighboring atoms, then by simple

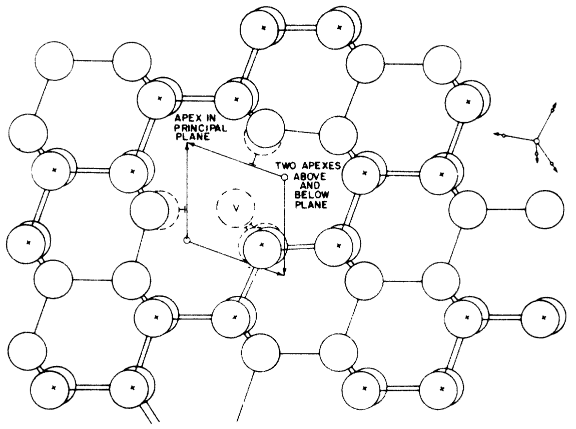


FIG. 4. Vacancy as a perfect regular octahedron of volume Ω. The symmetry preserving back displacement of the surrounding atoms is indicated in the insert. The magnitude of this displacement consistent with the assumed volume Ω is indicated.

geometry, each of these will be displaced away from the center of the vacancy by $0.211\, r_c$. (This distortion does not change the symmetry of the center; the Jahn-Teller distortion has not yet been introduced.) This criterion for determining the atomic displacement is certainly oversimplified, but it seems to be the most plausible first approximation. For Si this calculated back displacement is 0.25 Å, which may be compared with the theoretical value for the back displacement at an unreconstructed (111) surface obtained by Appelbaum and Hamann,[86] 0.33 Å, and with the value obtained by Florio and Robertson[85] from a somewhat crude analysis of low-energy-electron-diffraction data 0.16 Å. (EPR experiments do not determine how great the back displacement around a vacancy may be, but it has been concluded that the displacement is indeed away from the vacancy.[81]) It can be seen that the octahedron fits into the lattice well when centered at the lattice site and can not be displaced without distortion.

In Fig. 5 we see the octahedral vacancy cavity with its apexes rounded off. This rounding is expected to occur when the electron density tails spilling from the four faces which meet at each apex interfere [recall Eqs. (4.3) and (4.4)]. As a result of this rounding off, one may expect the four atoms at the surface of the cavity to be displaced toward the rounded apexes. A given surface atom could go toward any of the three apexes for its (111) plane, but one expects the atoms to pair off at opposite ends of the cavity in order to "balance the forces" so that the cavity remains centered at the lattice site with all (111) surfaces. Therefore, one concludes that the intereference of the electron charge density spilling into the vacancy cavity, which has a basically octahedral shape as a consequence of Eq. (4.12), should produce a threefold degenerate distortion of tetragonal symmetry. This evidently accounts for the tetragonal Jahn-Teller distortion which is observed in both the positively and the negatively charged state of the vacancy and is expected to occur in the neutral state.[36] (The neutral state is not observed in EPR.)

In Fig. 5 the tetragonal displacement of the two atoms neighboring the vacancy and lying in the principal plane is indicated. The two atoms neighboring the vacancy and lying above and below the principal plane undergo an analogous displacement toward each other. The magnitude of this tetragonal distortion can be estimated[15] as 7.2° ± 0.5° from measurements of distortion of the hyperfine tensor. (This is the magnitude indicated in Fig. 5). On the basis of the simplest microscopic models of the vacancy, one expects the tetragonal distortion to equal the measured hyperfine distortion and to be comparable in magnitude for all charge states.[36] However, it should be noted that the negative charge state has a trigonal distrotion in addition to the tetragonal distortion.

Now that we have seen that there is reasonable agreement between the macroscopic and microscopic pictures of the vacancy centered at a lattice site, let us consider how the vacancy will migrate to a neighboring lattice site, say the one that is to the left of V in the principal plane. The most naive assumption would be that the atom which is initially at the final site of the vacancy moves directly along the [111] bond axis to the initial site of the vacancy (see Fig. 6). The mobile atom would then jump through the vacancy. This is actually a very unlikely mode of migration for the following reasons. First, the mobile atom is not on the bond axis when the migration begins; it is displaced by the tetragonal Jahn-Teller distortion (see Fig. 5). Therefore, the bond axis, which by summetry, should be a line of local maxima or local minima for the function giving total energy as a function of mobile atom position, should in fact be a line of local maxima. Molecular-orbital calculations by Watkins *et al.*, support this conclusion.[56] Secondly if the mobile atom were to jump through the vacancy, then at the midpoint on the bond axis, it would be effectively separated from the bulk of the crystal. One might say that the mobile atom would be surrounded by the vacuum of the vacancy. Therefore, the activation enthalpy for this mode would be comparable with the heat of atomization

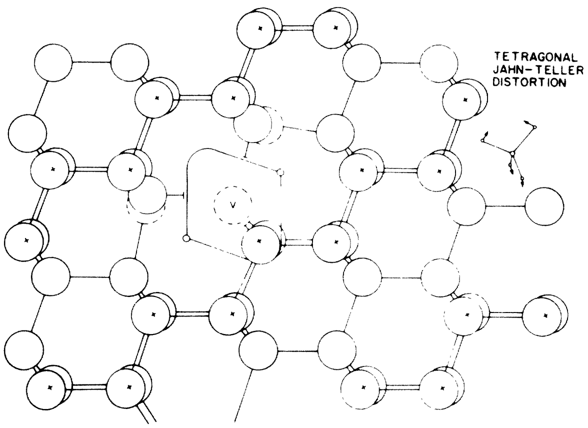


FIG. 5. Vacancy as octahedron with apexes rounded. As the electron density spilling from the various (111) surfaces interfere near each apex, these are rounded. This produces an instability and the surrounding atoms undergo a tetragonal, Jahn-Teller distortion. This distortion, which occurs for all charge states, has the symmetry indicated in the insert and is threefold degenerate. The magnitude indicated is taken from experiment, Ref. 15.

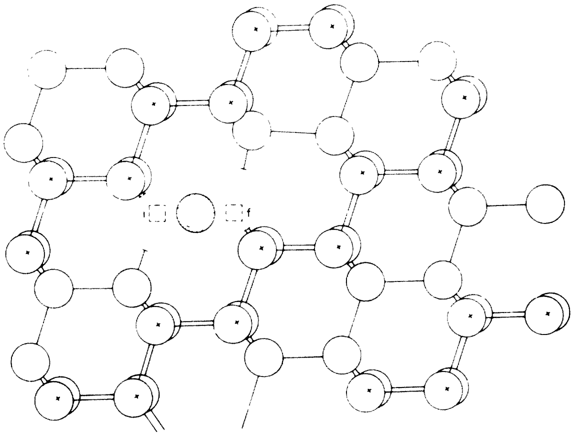


FIG. 6. Mode not taken. Direct motion of the mobile atom (shown shaded) along a bond axis would require a $\Delta H_m \approx \Delta H_a$ or about 4 eV.

ΔH_a in Eq. (4.6), which is much larger than the experimental value of either $\Delta H_m(T_l)$ or $\Delta H_m(T_h)$. We must conclude that the usual mode of vacancy migration is one in which the mobile atom moves around, rather than through, the vacancy so that it is never separated from the region of high electron density in the bulk of the crystal. One might describe this by saying that the mobile atom migrates around the macroscopic surface of the vacancy. Again, the distinction between metal-like forces, which depend primarily on local electron density, and covalent forces, together with the observation that the former make several times as great a contribution to the total energy as the latter, is crucial to our understanding.

Let us now consider the case that the vacancy migrates without distorting the lattice beyond the nearest-neighboring atoms. It is clear from Fig. 5 that this can not be accomplished without distorting the vacancy cavity from its equilibrium shape. From a consideration of the geometry of the Si lattice and from the observation that it is likely that $E_S(110)$ is significantly less than $E_S(100)$ [Eq. (4.16)], the author has concluded that the most likely configuration at the saddle point (subject to the restriction that the rest of the lattice remain undistorted) is the rhombohedral-dodecahedron illustrated in Fig. 7. This is taken to be the high-temperature transition state $V'(T_h)$. If we assume that the mobile atom lies in the principal plane and must not approach closer than the normal bond distance $2r_c$, then the other neighbor in the principal plane, that which is paired with the mobile atom by the tetragonal Jahn-Teller distortion, must be displaced somewhat to the right as indicated in Fig. 7. However, it seems likely that the mobile atom is actually displaced either above or below the principal plane. This is because the principal plane intersects the dedehedron along an edge on the side where the mobile atom is shown. Thus, the true position of the mobile atom is probably above or below the plane and further down toward the center of V' than is illustrated in Fig. 7. Therefore, the displacement of the paired atom need not be as great as is illustrated. The other atoms in Fig. 7 are shown at the same positions that they occupy in Fig. 5, where the vacancy is centered at a lattice site. The fit seems to be quite satisfactory and to support the assumption.

Clearly, one should expect there to be a large number of low-energy distortions[70] of the transition state idealized in Fig. 7 and, perhaps, there are other transition states of approximately the same enthalpy. Such considerations are relevant only to the estimation of ΔS_m and will not be followed here. The point is that ΔS_m and the frequency factor ν of Eqs. (1.7) to (1.11) may be expected to be large for this transition state so that it is a good candidate for $V'(T_h)$.

We must now evaluate $\Delta H_m(T_h) = \Delta H_{v'}(T_h) - \Delta H_v$ for the assumed $V'(T_h)$. As per Eq. (4.13), there will be two contributions—one from the addition covalent bonds which must be broken for migration and one from the metallic forces resulting from the distortion of the equilibrium shape of the va-

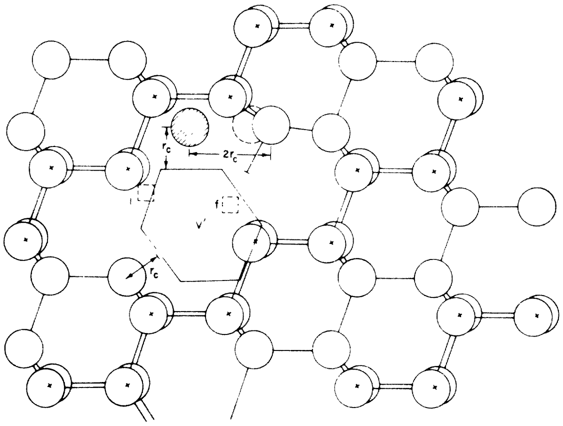


FIG. 7. Migrating vacancy as a regular rhombohedral dodecahedron [12 (110) surfaces]. The vacancy cavity with volume Ω is shown at the midpoint of its migration between the sites indicated by the dashed squares i and f. The mobile atom is shown shaded. The four short sides of the projection of the dodecahedron in the principal plane are the edges of rhombi on the surface of the cavity. The two long sides are diagonals of two other rhombi. The mobile atom is shown in the principal plane for simplicity. However, as it would bear against one of the edges joining surface planes if it were in the principal plane, it is almost certainly displaced above or below that plane. Thus the displacement of the paired atom need not be as great as is illustrated here.

cancy cavity. As was noted in Fig. 2, the minimum number of bonds which must be broken is two. It seems that the simple mode illustrated in Fig. 7 will break three bonds. Consequently, we have

$$\Delta H_m(T_h) = 3E^b(T_h) + \Delta H^m_{v'}(T_h) - \Delta H^m_v = 3E^b(T_h) + f(T_h)\,\Delta H^m_v \;. \tag{4.17}$$

Now $E^b(T)$ decreases slowly[32] with rising T and at the melting point

$$\begin{aligned} 3E^b(\mathrm{Si}, T^F) &= 0.46\ \mathrm{eV}, \\ 3E^b(\mathrm{Ge}, T^F) &= 0.24\ \mathrm{eV}. \end{aligned} \tag{4.18}$$

(At $T=0$, $3E^b$ would amount to 0.57 and 0.33 eV, respectively.)

It is not possible to derive a precise value for the factor $f(T_h)$ in Eq. (4.17) because Eq. (4.16) merely sets a lower bound. However, we may calculate this lower bound in a straightforward manner. Denoting the surface area of the regular rhombohedral-dodecahedron with volume Ω, i.e., $V'(T_h)$, as A_{12} and that of the regular octahedron, i.e., V, as A_8, one finds

$$A_8 = 1.1826 A_\infty, \quad A_{12} = 1.1053 A_\infty, \tag{4.19}$$

where $A_\infty = 4\pi r_w^2$ is the surface area of the Wigner-Seitz sphere. Therefore, from Eq. (4.16), one finds

$$f(T_h) \geq 0.145 \;. \tag{4.20}$$

Using the values of ΔH^m_v obtained from observation of the surface energy of the metallic liquid phase, which yield values for ΔH_v in good agreement with experiment[28,34,71]

$$\begin{aligned} \Delta H^m_v(\mathrm{Si}) &= 2.15\ \mathrm{eV}, \\ \Delta H^m_v(\mathrm{Ge}) &= 1.85\ \mathrm{eV}, \end{aligned} \tag{4.21}$$

one finds $f(T_h)\,\Delta H^m_v(T_h) \geq 0.31$ and 0.27 eV for Si and Ge, respectively. Therefore, we may conclude

$$\begin{aligned} \Delta H_m(\mathrm{Si}, T_h) &\geq 0.77\ \mathrm{eV} \\ \Delta H_m(\mathrm{Ge}, T_h) &\geq 0.51\ \mathrm{eV}. \end{aligned} \tag{4.22}$$

In order to account for the observed values of $\Delta H_m(T_h)$, it is necessary to assume that $f(T_h)$ is larger than the minimum value consistent with the octahedral shape [Eq. (4.20)]. If one takes

$$f(T_h) = 0.40 \;, \tag{4.23}$$

which corresponds to assuming $E^m_S(110) = 1.5\, E^m_S(111)$ instead of Eq. (4.16), and maintaining the assumption that the saddle point shape is the rhombohedral dodecahedron, then one calculates, for $T_h = T^F$,

$$\begin{aligned} \Delta H_m(\mathrm{Si}, T_h) &= 1.32\ \mathrm{eV}, \\ \Delta H_m(\mathrm{Ge}, T_h) &= 0.98\ \mathrm{eV}, \end{aligned} \tag{4.24}$$

which is in excellent agreement with the experimental values 1.2 ± 0.3 eV and 1.0 ± 0.2 eV, respectively.

It seems reasonable that $f(T_h)$ could have a value as large as that assumed in Eq. (4.23). Indeed, it would be strange if Si and Ge both just happened to have the minimum anisotropy of their surface energy consistent with their octahedral equilibrium shapes.[90] The value assumed in Eq. (4.23) is supported by observation of the relative dissolution rate in aqueous etching solutions, where the (110) surface of Ge dissolves about 1.44 times as fast as the (111) surface.[92,93] Phillips has obtained[89] similar values for both Si and Ge from analysis of the vacancy concentrations associated with the surface reconstruction observed on these faces.

Returning to the problem of $\Delta H_m(T_l)$, we recall that it was noted at Eq. (1.28) that experimentally

$$\Delta H_m(T_l) \lesssim 2E^b(T=0) \tag{4.25}$$

and at Eq. (1.29) that

$$f(T_h)\,\Delta H^m_v \gg E^b \gg k\Theta \;. \tag{4.26}$$

From Eq. (4.26) it is clear that the enthalpy of the state $V'(T_h)$, which is illustrated in Fig. 7, can be reduced by distortion of the lattice. A precise specification of the transition state of lowest enthalpy $V'(T_l)$ is probably beyond the scope of the present macroscopic model because it likely involves a complicated distortion of the lattice for several rings about the vacancy. However, it seems clear that the minimum enthalpy distortion should have an enthalpy only a few $k\Theta$ greater than the minimum $\Delta H_v + 2E^b$ imposed by the counting of broken bonds. Thus we calculate

$$\begin{aligned} \Delta H(\mathrm{Si}, T_l) &\gtrsim 2E^b(\mathrm{Si}, T=0) = 0.37\ \mathrm{eV}, \\ \Delta H(\mathrm{Ge}, T_l) &\gtrsim 2E^b(\mathrm{Ge}, T=0) = 0.22\ \mathrm{eV} \;. \end{aligned} \tag{4.27}$$

The entropy of the state $V'(T_l)$ should be low because a complicated correlation of many atoms is required to achieve this minimum. If the third law of thermodynamics may be invoked, the state $V'(T_l)$ should be a singlet with zero entropy.[94] Moreover, the frequency factor in Eqs. (1.7) to (1.11) for the transition to $V'(T_l)$ should clearly be much less than ν_D because a time of several phonon periods will be required to minimize the enthaply. For the values of D_0 for Ge used in Eq. (1.21) (the "theory" curve in Fig. 1), one may assume

$$\nu(T_l) = 3.5\times 10^{-2}\,\nu_D, \quad \Delta S_m(T_l) = -\Delta S_v \approx -4k, \tag{4.28}$$

while at high temperatures for the simple migration process (Fig. 7),

$$\nu(T_h) = \nu_D, \quad \Delta S_m(T_h) = 7.6\,k \;. \tag{4.29}$$

Now the slight discrepancy between Eqs. (4.25) and (4.27) might be due to some relaxation which reduces the energy of the broken covalent bond to a few $k\Theta$ below $E^b(T=0)$ or it might result from experimental or theoretical error. But it might also indicate that $\nu(T_l)$ is in fact decreasing with increasing temperature. This possibility was mentioned in Sec. I, where it was noted that it would help greatly to reconcile low temperature estimates of D_0 with high-temperature measurements. Note that if ν is a function of T, then there is a semantic problem regarding the definition of ΔH_m; the definition (1.5) is not simply proportional to the slope of $\log_{10} D_v(T)$ vs T. In order to reconcile exactly the theoretical estimate $\Delta H_m(\mathrm{Si}, T_l) = 0.37$ eV with the Watkins empirical estimate[15] $\Delta H_m(\mathrm{Si}, T_l) = 0.33$ eV, one must assume that $\nu(T_l)$ decreases by a factor of 2.0 between 140 and 180 °K. Such an effect is plausible in the present model because several steps may be required for the transition to the $V'(T_l)$ state and it is possible that the probability of these intermediate states is not independent of the occupancy of other states. Thus, if the transition is not a Markov process, it may be damped by thermal excitation of phonons, etc.

ACKNOWLEDGMENTS

The author has greatly benefited from the discussions with several colleagues. He is particularly grateful to Dr. W. L. Brown, Dr. H. C. Casey, Dr. C. Herring, Dr. L. C. Kimerling, Dr. R. C. Miller, Dr. J. C. Phillips, Dr. M. D. Sturge, Dr. C. D. Thurmond, and Dr. G. D. Watkins.

APPENDIX A: THEORY OF EQUILIBRIUM SHAPES

As early as 1878 it was realized by Gibbs [95] that the thermodynamically stable shape for a given volume (quantity) of material is that which minimizes the integral of the surface tension, i.e., the surface free energy per unit area $F_S(ijk)$, over the surface,

$$\int F_S(ijk)\, dA \equiv F = \min \ . \tag{A1}$$

If $F_S(ijk)$ is constant, then the equilibrium shape is a sphere. If the observed equilibrium shape is not a sphere, then $F_S(ijk)$ can not be constant.

In 1901, Wulff proposed that the true equilibrium shape may be found as follows.[37] Starting from a fixed point 0, one constructs a spherical plot of $F_S(ijk)$. Thus, one constructs a closed surface about 0 such that in every direction (ijk) from 0, the surface is a distance $r(ijk)$ proportional to $F_S(ijk)$ away. One then constructs planes perpendicular to the radius vector from 0 at each point of this surface. The volume interior to all these planes has the equilibrium shape appropriate to $F_S(ijk)$. (see Fig. 8).

Wulff's theorem is completely general and correct but his proof of it was not.[37] Several authors, including those of Refs. 33, 38, and 39 have since given correct proofs. The most general of these is that given by Herring.[33] However, the proof given by Liebmann,[38] which shows only that the Wulff construction minimizes F among all convex polyhedrons, will be reviewed here because it lends itself more easily to the analysis of the octahedral equilibrium shape of Si and Ge.

For a given convex polyhedron, we may write

$$F = \sum A(ijk)\, F_S(ijk) \ , \tag{A2}$$

where $A(ijk)$ is the area of the (ijk) surface of the polyhedron and the summation extends over all its surfaces. The volume V of this polyhedron is found by adding contributions from pyramids of base $A(ijk)$ and height $r(ijk)$ to be

$$V = \frac{1}{3} \sum r(ijk)\, A(ijk). \tag{A3}$$

When considering variations of the shape of the material, it is convenient to treat the areas of the various surface planes, the $A(ijk)$'s, as the independent variables. At this point, we shall consider variations which change the given $A(ijk)$'s but do not add new surfaces. In order to express dV in terms of $dA(ijk)$, we differentiate (A3)

$$3dV = \sum \left[r(ijk)\, dA(ijk) + A(ijk)\, dr(ijk) \right]$$

and also find by geometrical construction

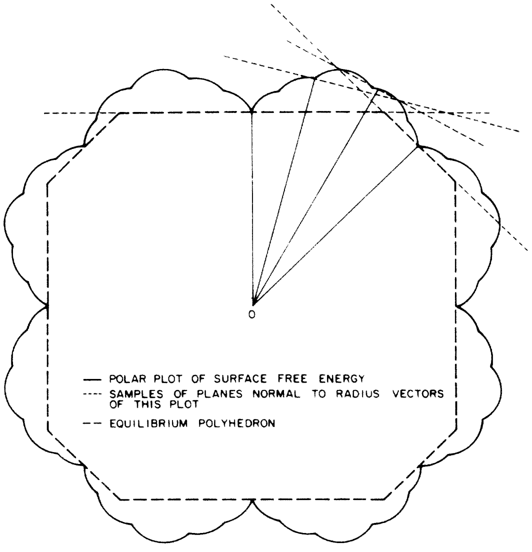


FIG. 8. Schematic illustration of the Wulff construction of the equilibrium shape. Adapted from Herring, Ref. 33.
]

$$dV=\sum A(ijk)\,dr(ijk)\ ,$$

so that

$$2dV=\sum r(ijk)\,dA(ijk)\ . \tag{A4}$$

Because the surface of the material is closed, the sum of the projections onto any axis of an outward normal vector from each surface proportional to the surface area must be zero. Taking the usual x, y, and z coordinate axes and directional cosines of the (ijk) surfaces, $\cos\alpha_x(ijk)$, etc., one has

$$\begin{aligned}&\sum A(ijk)\cos\alpha_x(ijk)=0\ ,\\&\sum A(ijk)\cos\alpha_y(ijk)=0\ ,\\&\sum A(ijk)\cos\alpha_z(ijk)=0\ .\end{aligned} \tag{A5}$$

We may now solve the problem of minimizing F for a convex polyhedron having the given set of surfaces subject to the three constraints of (A5) plus one that V be constant by the well known method of Lagrangian multipliers

$$\begin{aligned}\lambda dF-2dV&+\lambda_x\Big(\sum\cos\alpha_x(ijk)\,dA(ijk)\Big)\\&+\lambda_y\Big(\sum\cos\alpha_y(ijk)\,dA(ijk)\Big)\\&+\lambda_z\Big(\sum\cos\alpha_z(ijk)\,dA(ijk)\Big)=0\ .\end{aligned} \tag{A6}$$

Upon substitution from (A4) and with $dF=\sum F_S(ijk)\,dA(ijk)$, one may separate the terms of the summation into factors of each $dA(ijk)$,

$$\begin{aligned}\lambda F_S(ijk)-r(ijk)&+\lambda_x\cos\alpha_x(ijk)\\&+\lambda_y\cos\alpha_y(ijk)+\lambda_z\cos\alpha_z(ijk)=0.\end{aligned} \tag{A7}$$

Now consider the plane defined by

$$\begin{aligned}x\cos\alpha_x(ijk)&+y\cos\alpha_y(ijk)\\&+z\cos\alpha_z(ijk)-r(ijk)=0,\end{aligned}$$

for which the directional cosines of the normal vector are indeed $\alpha_x(ijk)$, $\alpha_y(ijk)$, and $\alpha_z(ijk)$. The distance from this plane to the point $x=\lambda_x$, $y=\lambda_y$, $z=\lambda_z$ is simply

$$\begin{aligned}R(ijk)=r(ijk)&-\lambda_x\cos\alpha_x(ijk)\\&-\lambda_y\cos\alpha_y(ijk)-\lambda_z\cos\alpha_z(ijk),\end{aligned} \tag{A8}$$

which by (A7) is

$$R(ijk)=\lambda F_S(ijk). \tag{A9}$$

Obviously, the point $0=(\lambda_x,\lambda_y,\lambda_z)$ is the same for all surface planes which appear on the surface of the given polyhedron and this is the point about which the Wulff method, with Eq. (A9), constructs the equilibrium shape. However, it may occur that certain planes do not appear on the equilibrium shape at all, i.e., $A(ijk)=0$ for some (ijk). This is certainly the case for Si and Ge, for which only (111) surfaces appear. Therefore, one must consider which faces may have nonzero area on the equilibrium surface, i.e., for which set of surfaces for the given polyhedron will the minimum free energy be obtained? One may show that the Wulff construction gives answers to this question correctly by selecting first the set of (ijk)'s for which $F_S(ijk)$ is minimum. These surfaces must certainly appear on the equilibrium shape. One constructs a closed surface from these planes which, according to (A9) must be regular in order that F be a minimum for this set of planes. For the set of (111) planes, one obtains the regular octahedron; for the set of (110) planes, one obtains the rhombehedral dedecahedron; for the set of (100) planes, one obtains the cube. Having constructed the regular solid for the minimum $F_S(ijk)$, one may consider the set of $(i'j'k')$ planes corresponding to the second smallest values of $F_S(i'j'k')$ and ask if, in the Wulff construction, the corresponding planes intersect this regular solid. If they do, then one may add the set of $(i'j'k')$ planes to the (ijk) set and follow through the minimization process to (A9). Obviously, the $A(i'j'k')$'s will be nonzero as indicated by the Wulff construction. One may then proceed to the next smallest set $F_S(i''j''k'')$ and so forth. When it occurs that the set of $(i'j'k')$ planes drawn according to the Wulff construction do not interest the previously optimized solid, we clearly have the situation that $F_S(i'j'k')$ is too large for $A(i'j'k')$ to be nonzero on the equilibrium shape.

This concludes the review of Liebmann's proof[38] of Wulff's theorem.[37]

The author has not been able to find the example of octahedral equilibrium shape explicitly analyzed and Eq. (4.12) derived anywhere in the literature. Therefore, that will be done here.

The equation of the regular octahedron is

$$|x|+|y|+|z|=\text{const.}\equiv 3^{1/2}F_S(111). \tag{A10}$$

Thus, we have taken $\lambda=1$ and $R(111)=F_S(111)$. Therefore, the equations of the cube and of the rhombohedral dodecahedron are

$$|x|=F_S(100),\quad y=z=0, \tag{A11}$$

etc., and

$$|x|+|y|=2^{1/2}F_S(110),\quad z=0 \tag{A12}$$

etc. Finally, the condition that the cube face just touch the apex of the octahedron is

$$F_S(100) = 3^{1/2} F_S(111) \tag{A13}$$

and that the face of the dodecahedron just touch the apex of the octahedron is

$$2^{1/2} F_S(110) = 3^{1/2} F_S(111)\ . \tag{A14}$$

It follows from the fact that only (111) surfaces appear on the equilibrium shape of Si and Ge that

$$\begin{aligned} &F_S(100) \geq \sqrt{3}\, F_S(111), \\ &F_S(110) \geq \sqrt{1.5}\, F_S(111)\ . \end{aligned} \tag{A15}$$

This is equivalent to Eq. (4.12) where, as in this paper, the relevant temperature is sufficiently low that we may neglect the distinction between $F_S(ijk)$ and $E_S(ijk)$. In view of Eq. (4.15), it appears likely that the surface entropy per unit area $S_S(ijk)$ is anisotropic to approximately the same extent as the broken bond energy

$$\begin{aligned} &S_S(100) \approx 3^{1/2} S_S(111), \\ &S_S(110) \approx (1.5)^{1/2} S_S(111)\ , \end{aligned} \tag{A16}$$

so that the distinction between F_S and E_S is less important than one might otherwise think.

APPENDIX B: DISTINCTION BETWEEN VACANCY MIGRATION IN COVALENT SOLIDS AND IN METALLIC OR IONIC SOLIDS

Some readers may wonder why the temperature dependence of ΔH_m, and indeed the entire nature of vacancy migration, in covalent solids like Si and Ge is so dramatically different from that observed in metals and in ionic solids, such as the alkali halides or AgBr. It is hoped these few qualitative observations will be useful to such readers.

The primary reasons that the nature of vacancy migration is so different in the covalent solids from that in ionic solids or metals are that the nature of the interatomic forces and of the crystal structure are quite different in the two cases. The structures of metals and of ionic solids are, in general, nearly close packed and the interatomic forces are almost entirely central in nature. The hard sphere model of atoms is usually adequate to describe most properties of such noncovalent solids. In particular, it is probably adequate to describe atomic or vacancy migration in such solids, as Vineyard and Rice have done.[19,20] However, covalent structures, such as that of Si or Ge, are notoriously non-close-packed and the covalent interatomic forces are by definition, noncentral.[96] Strong bond-bending force constants are required to stabilize a diamond-type lattice against shear. Therefore, it should be obvious that a hard sphere model of atoms, which utterly fails to describe the static lattice, can not be appropriate to describe atomic or vacancy migration in covalent solids. Thus, one must resort to some other model such as the bubble or macroscopic cavity model used here, to describe vacancy migration in a covalent solid.

In metals and in ionic solids, there is evidence for only one transition state and ΔH_m seems to *decrease* somewhat with rising temperatures due to the effects of anharmonicity.[25,97] Put simply, atomic migration becomes easier as the lattice lossens up. This is because the thermal expansion of the lattice and the softening of its vibrational modes will affect a decrease in $\Delta H_{V'}$ relative to ΔH_V, i.e., the transition state benefits more than the state at the lattice site. Therefore, $\Delta H_m \equiv \Delta H_{V'} - \Delta H_V$ decreases. This effect is particularly striking in AgBr and other superionic-conductors,[97–99] where ΔH_m falls to a very low value at some critical temperature.

As is well known, there is a direct relation between the coefficient of (linear) thermal expansion α and the anharmonic component of the interatomic forces. The hard sphere model implies an asymmetric minimum in the potential as a function of nearest-neighbor distance. Therefore,

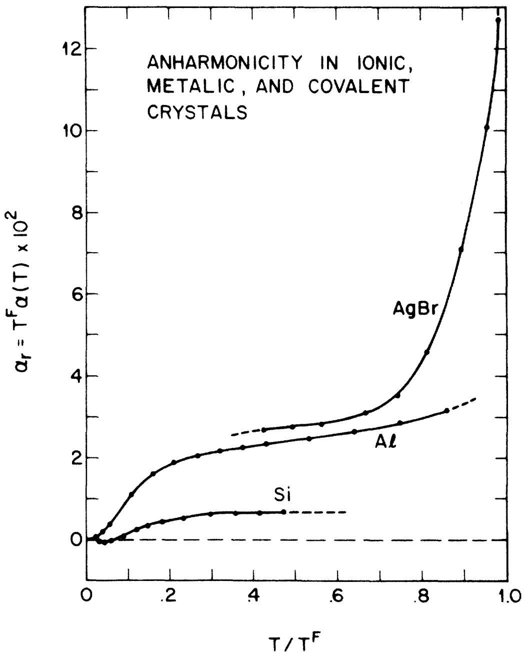


FIG. 9. Relative magnitudes of the anharmonic components in the interatomic forces of various classes of solids are evidenced by the reduced coefficients of linear expansion $\alpha(T/T^F) = \alpha T^F$. T^F is the melting temperature: $T^F(\mathrm{Si}) = 1685\,°\mathrm{K}$; $T^F(\mathrm{Al}) = 933\,°\mathrm{K}$; $T^F(\mathrm{AgBr}) = 705\,°\mathrm{K}$. The points shown are experimental (Refs. 100–102).

solids which may be described by the hard sphere model must have a large anharmonic component in their interatomic forces and large values of α. Because covalent bonding results from the constructive interference between the atomic wave function on neighboring atoms, the anharmonic component and thus, the values of α should be much less in covalent solids than in those which may be described with a hard sphere model.

This point is illustrated in Fig. 9, where the reduced coefficients of thermal expansion $\alpha_r = T^F \alpha(T)$ where T^F is the melting point, is plotted as a function of reduced temperature T/T^F, for an ionic solid[100] AgBr, a typical metal[101] Al, and for Si.[102] As may be seen, α_r is everywhere positive in AgBr and Al and one to two orders of magnitude larger than in Si. The softening of the AgBr lattice above about 500 °K is noted. In Si, $\alpha(T) < 0$ for $T < 130$ °K with a minimum of $-4.6 \times 10^{-7}/°\mathrm{K}$ at 80 °K; at higher temperatures $\alpha(T)$ saturates at $+4.1 \times 10^{-6}/°\mathrm{K}$ a little above the Debye temperature and does not continue to rise.[102] This behavior of $\alpha(T)$ is typical of all tetrahedrally coordinated semiconductors.[103] Indeed, if similar data[104] for ZnO, one of the most ionic compounds[105] to occur in a tetrahedral structure, were to be plotted in Fig. 9, the two curves could not be distinguished on that scale.

From these considerations and from Fig. 9, we should conclude that the anharmonic effects, which would be expected to produce a decrease in ΔH_m for both transition states, are much less in Si and Ge than in metals or ionic solids and may be neglected at this stage in the refinement of a theory for atomic migration in covalent solids. Moreover, the hard-sphere model of atoms, which may lead to an adequate understanding of migration in metals and ionic solids, is certainly unjustified here. We have no reason to expect that experience with migration in metals and rock salt structure solids will be any guide to the problems of atomic migration in tetrahedral semiconductors.

---

[1] A. Seeger, Commun. Solid State Phys. 1, 157 (1968).

[2] A. Seeger and K. P. Chick, Phys. Status Solidi 29, 455 (1968).

[3] A. Seeger, Commun. Solid State Phys. 2, 55 (1969).

[4] D. L. Kendall and D. B. DeVries, in *Semiconductor Silicon*, edited by R. R. Haberect and E. L. Kern (Electrochemical Society, New York, 1969), p. 358.

[5] B. J. Masters, Solid State Commun. 9, 283 (1971).

[6] R. R. Hasiguti, in *Lattice Defects in Semiconductors*, edited by R. R. Hasiguti (Pennsylvania State U. P., University Park, Pa., 1968), p. 131.

[7] V. V. Voronkov, G. I. Voronkova, and M. I. Iglitsyn, Fiz. Tek. Popuprov 6, 20 (1972) [Sov. Phys.-Semicond. 6, 15 (1972)].

[8] R. F. Peart and R. C. Newman, in *Radiation Damage and Defects in Semiconductors*, edited by J. E. Whitehouse (Institute of Physics, London, 1973), p. 170.

[9] H. C. Casey, Jr. and G. L. Pearson, in *Point Defects in Solids*, edited by J. H. Crawford and L. M. Slifkin (Plenum, New York, 1973), Vol. 2.

[10] A. Hiraki, J. Phys. Soc. Jap. 21, 34 (1966).

[11] R. A. Logan, Phys. Rev. 101, 1455 (1956).

[12] H. Letaw, W. M. Portnoy, and L. Slifkin, Phys. Rev. 102, 636 (1956).

[13] R. E. Whan, Phys. Rev. 140, A690 (1965).

[14] E. E. Klontz and J. W. MacKay, J. Phys. Soc. Jap. Suppl. 18, 216 (1963).

[15] G. D. Watkins, J. Phys. Soc. Jap. Suppl. 18, 22 (1963).

[16] L. Elstner and W. Kamprath, Phys. Status Solidi 22, 541 (1967).

[17] R. N. Ghoshtagore, Phys. Rev. B 3, 397 (1971).

[18] N. F. Mott and R. W. Gurney, *Electronic Processes in Ionic Crystals* (Dover, New York, 1964), p. 34.

[19] G. H. Vineyard, J. Phys. Chem. Solids 3, 121 (1957).

[20] S. A. Rice, Phys. Rev. 112, 804 (1958).

[21] S. A. Rice and H. L. Frisch, J. Chem. Phys. 32, 1026 (1960).

[22] C. Kittel, *Thermal Physics* (Wiley, New York, 1969), p. 348.

[23] A. V. Vasil'ev, L. A. Rybakova, and L. S. Smirnov, Fiz. Tekh. Poluprov. 5, 2297 (1971) [Sov. Phys.-Semicond. 5, 2008 (1972)]; A. V. Vasil'ev, *ibid.* 6, 603 (1972) [*ibid.* 6, 522 (1972)]; V. I. Panov, L. S. Smirnov, and E. G. Tishkovskii, *ibid.* 6, 673 (1972) [*ibid.* 6, 580 (1972)].

[24] cf. H. Reiss and C. S. Fuller, in *Semiconductors*, edited by N. B. Hannay (Reinhold, New York, 1959), p. 253. Note that the full equation is

$$\frac{-d(\ln N_v)}{dt} = 4\pi r_0 \, D_v[I] \left(1 + \frac{r_0}{(\pi D_v t)^{1/2}} \right)$$

where N_v is the (excess) concentration of vacancies. The decay does not become exponential until the final term approaches unity. For the values r_0 and D_v estimated by Watkins in Ref. 15, Eqs. (1.17) and (1.18), for $T = 160$ °K in Si, the final term is down to 1.11 after 1 sec. As a typical annealing time is 600 sec, it is commonly assumed that the decay is exponential throughout the experiment. However, if $r_0 \gg d_0$, as would occur if either or both species were charged, and/or D_0 is much less than the values indicated by Eqs. (1.18) and (1.19), and/or a lower temperature is used, one may get into a regime where the final term can not be neglected and the decay is not exponential, Moreover, as it is often difficult to distinguish an $e^{-t/\tau}$ from a $t^{-1/2}$ dependence experimentally, one regime may be mistaken for the other. Note that when the final term is dominant, the slope of the $\ln D_v$ vs $1/T$ will give $\frac{1}{2}\Delta H_m$ rather than ΔH_m.

[25] cf., e.g., K. J. Laidler, *Theories of Chemical Reaction Rates* (McGraw-Hill, New York, 1969), p. 52.

[26] G. D. Watkins and J. W. Corbett, Phys. Rev. 134, A1359 (1964).

[27] J. W. Corbett and G. D. Watkins, Phys. Rev. 138, A555 (1965).

[28] J. C. Phillips and J. A. Van Vechten, Phys. Rev. Lett. 30, 220 (1973).

[29] J. A. Van Vechten, Bull. Am. Phys. Soc. 18, 427 (1973), and J. Electrochem. Soc. (to be published).

[30] J. Schmit and A. A. Lucas, Solid State Commun. 11, 415 (1972); 11, 419 (1972).

[31] J. Schmit and A. A. Lucas, Collect. Phen. 1, 1 (1973).

[32] J. A. Van Vechten, Phys. Rev. B 7, 1479 (1973).

[33] C. Herring, in *Structure and Properties of Solid Surfaces*, edited by G. Gomer and C. S. Smith (University of Chicago Press, Chicago, 1953), p. 5 and especially pp. 24–30.

[34] J. A. Van Vechten and C. D. Thurmond, J. Phys. Chem. Solids (to be published). The subject of this paper is mentioned by F. A. Kröger [in *The Chemistry of Imperfect Crystals* (North-Holland, Amsterdam, 1964), p. 325] and by R. F. Brebrick [J. Solid State Chem. 1, 88 (1969)].

[35] A. D. Liehr, J. Phys. Chem. 67, 389 (1963); 67, 471 (1963).

[36] For an excellent review see M. D. Sturge, in *Solid State Physics*, edited by F. Seitz, D. Turnbull, and H. Ehrenreich (Academic, New York, 1967), Vol. 20, 91 and especially p. 170.

[37] G. Wulff, Z. Kristallpgr. 34, 449 (1901).

[38] H. Liebmann, Z. Kristallogr. 53, 171 (1914).

[39] M. v. Lau, Z. Kristallogr. 105, 124 (1943).

[40] R. N. Ghoshtagore, Phys. Rev. Lett. 16, 890 (1966).

[41] J. M. Fairfield and B. J. Masters, J. Appl. Phys. 38, 3148 (1967).

[42] W. L. Brown and M. Augustyniak, J. Appl. Phys. 30, 1300 (1959).

[43] H. J. Stern and F. L. Vook, in *Radiation Effects in Semiconductors*, edited by F. L. Vook (Plenum, New York, 1968), p. 155.

[44] R. O. Simmons and R. W. Balluffi, Phys. Rev. 117, 52 (1960).

[45] M. L. Swanson, Phys. Status Solidi 33, 721 (1969).

[46] J. R. Patel (private communication).

[47] W. C. Dash, J. Appl. Phys. 27, 1193 (1956).

[48] A. G. Tweet, Phys. Rev. 111, 67 (1958).

[49] T. S. Plaskett, Trans. AIME 223, 809 (1965).

[50] P. Petroff and R. L. Hartman, Appl. Phys. Lett. 23, 469 (1973).

[51] G. K. Wertheim, Phys. Rev. 115, 568 (1959).

[52] E. E. Klontz and J. W. MacKay, J. Appl. Phys. 30, 1269 (1959).

[53] S. Ishino, F. Nakazawa, and R. R. Hasiguti, J. Phys. Chem. Solids 24, 1033 (1963).

[54] V. V. Khainovskaya and L. S. Smirnov, Fiz. Tverd. Tela 9, 2043 (1967) [Sov. Phys.-Solid State 9, 1603 (1968)].

[55] L. C. Kimerling, thesis Massachusetts Institute of Technology, 1969 and to be published.

[56] G. D. Watkins, R. P. Messmer, C. Weigel, D. Peak, and J. W. Corbett, Phys. Rev. Letters 27, 1573 (1971).

[57] E. I. Blount, Phys. Rev. 113, 995 (1959).

[58] J. A. Van Vechten, Phys. Rev. 182, 891 (1969).

[59] A. G. Tweet, J. Appl. Phys. 30, 2002 (1959). Note especially footnote 10 re Logan's experiment, (Ref. 11).

[60] For a review of this subject see H. J. Stein in *Radiation Effects in Semiconductors*, edited by J. W. Corbett and G. D. Watkins (Gordon and Breach, London, 1971), p. 125.

[61] R. R. Hasiguti and S. Motomiya have questioned this assumption [in Ref. 60, p. 39]. They propose that the true value of $\Delta H_{\rm expt}$ is 40% larger than the empirically determined value and use this hypothesis to explain the Ge self-diffusion data with the low-temperature value for ΔH_m. However, vacancy annealing experiments in Refs. 10, 11, and 16 and Eq. (2.10) determine ΔH_m independent of this assumption and find the same value as one gets by subtracting $\Delta H_{\rm expt}$ from the activation enthalpy for self-diffusion. Moreover, one would expect the apparent $\Delta H_{\rm expt}$ to change when the quenching time is varied from 0.1 to 1.0 sec; this was found not to occur (Refs. 10 and 11). Therefore, $\Delta H_{\rm expt}$ can not be seriously in error.

[62] Some authors denote the factor $\exp[-\Delta H_b(A)/kT]$ the "correlation factor" (Refs. 2 and 17). Some, e.g., Ref. 2, also define the correlation factor to be $\frac{1}{2}$ for self-diffusion via the vacancy mechanism in a Si lattice. This factor of $\frac{1}{2}$ results from geometry rather than binding and was included in the $\frac{1}{3}d_0^2$ factor in Eq. (1.3), etc.

[63] R. N. Ghoshtagore, Phys. Rev. Lett. 25, 856 (1970).

[64] R. C. Smith and L. R. Holland, J. Appl. Phys. 37, 4866 (1966).

[65] R. C. Smith, J. Appl. Phys. 37, 4860 (1966).

[66] A. Baldereschi, J. Luminesc. 7, 79 (1973).

[67] L. C. Kimerling, H. M. DeAngelis, and C. P. Carnes, Phys. Rev. B 3, 427 (1971).

[68] L. C. Kimerling and C. P. Carnes, J. Appl. Phys. 42, 3548 (1971).

[69] The standard entropy change ΔS^f in a reaction is that portion of the total which is independent of the number of units which react. Thus ΔS^f results from changes in the vibrational levels, etc., occurring within the reacting species and does not include the entropy of mixing. $\Delta F_{0,2v} = \Delta H_{2v} - T\Delta S_{2v}^f$ is the "standard free energy of formation" of the divacancy.

[70] R. A. Swalin, J. Phys. Chem. Solids 18, 290 (1961).

[71] In Ref. 28 the error of not correcting the surface tension, σ, which is the *free* energy per unit area, for effect of the surface entropy per unit area, $\Delta S_s = (\partial\sigma/dt)_V$, was made. This error is corrected in Ref. 29 and is shown to improve agreement between theory and experiment. However, this correction is only about 15% and is partially compensated by the effect of the difference in density between the liquid and solid phases—an effect also neglected in Ref. 28.

[72] K. H. Bennemann, Phys. Rev. 137, A1497 (1965).

[73] K. N. Sevast'yanov and F. T. Umarova, Fiz. Tverd. Tela 12, 2195 (1970) [Sov. Phys.-Solid State 12, 1756 (1971)].

[74] T. Soma and A. Morita, J. Phys. Soc. Jap. 32, 357 (1972).

[75] Wave mechanical methods may be useful as a descriptive tool in cases, such as Si, where several of the microscopic parameters of the vacancy have been established by experiment and thus can be used to fix the parameters in the calculation. See, e.g., G. D. Watkins, and R. P. Messmer, in *Proceedings of the Tenth International Conference on the Physics of Semiconductors*, edited by S. P. Keller, J. C. Hensel and F. Stern (U. S. AEC, Springfield, Va., 1970), p. 623.

[76] R. Chang and L. M. Falicov, J. Phys. Chem. Solids 32, 465 (1971).

[77] C. A. Coulson and F. P. Larkins, J. Phys. Chem. Solids 32, 2245 (1971).

[78] F. P. Larkins and A. M. Stoneham, Proc. Phys. Soc. Lond. 4, 143 (1971); 4, 154 (1971); 4, 3065 (1971); 4, 3077 (1971).

[79] cf. F. A. Kröger, Ref. 34, p. 446.

[80] cf. C. M. Drum. Philos. Mag. 13, 1239 (1966).

[81] R. P. Messmer and G. D. Watkins, Phys. Rev. B 7, 2568 (1973); G. D. Watkins (private communication).

[82] H. R. Potts and G. L. Pearson, J. Appl. Phys. 37, 2098 (1966).

[83] F. L. Vook, J. Phys. Soc. Jap. Suppl. 18, 190 (1963).

[84] cf. J. C. Woolley, in *Compound Semiconductors*, edited by R. K. Willardson and H. L. Goering (Reinhold, New York, 1962), p. 5.

[85] J. V. Florio and W. D. Robertson, Surf. Sci. 24, 173 (1971).

[86] J. A. Appelbaum and D. R. Hamann, Phys. Rev. Lett. 31, 106 (1973); Phys. Rev. B 8, 1777 (1973).

[87] See J. J. Gilman, J. Appl. Phys. 31, 2208 (1960) for an empirical value for $E_s(111)$ of Si.

[88] cf. J. I. Pankove, R. C. A. Rev. 15, 75 (1954).

[89] J. C. Phillips, Surface Sci. 40, 459 (1973).

[90] Those who ascribe all the cohesive energy to the nearest neighbor bonds sometimes maintain that the equilibrium shape is explained by Eq. (4.15), i.e., the anisotropy of $N_s(ijk)$ corresponds to the minimum required anisotropy of $E_s(ijk)$ [Eq. (4.12)]. As noted with regard to Ref. 80, this notion is refuted by the observations of crystal bending. Moreover, the premise of bond additivity is untenable for these crystals, as may be seen from Eq. (4.7).

[91] J. A. Van Vechten and R. M. Martin, Phys. Rev. Lett. 28, 446 (1972), and references therein.

[92] H. C. Gatos and M. C. Lavine, Prog. Semicond. 9, 13 (1965).

[93] W. K. Zwicker and S. K. Kurtz, in *Semiconductor Silicon* 1973, edited by H. R. Huff and R. R. Burges (Electrochemical Society, New York, 1973), p. 315.

[94] C. Kittel, Ref. 22, p. 48.

[95] J. W. Gibbs, Trans. Conn. Acad. 3, 343 (1878); *Collected Works* (Longmans, Green and Co., New York, 1928), pp. 219–331.

[96] R. M. Martin, Phys. Rev. B 1, 4005 (1970).

[97] C. P. Flynn, Z. Naturforsch 26, A 99 (1971).

[98] B. B. Owens and G. R. Argue, J. Electrochem. Soc. 117, 898 (1970).

[99] S. Geller and B. B. Owens, J. Phys. Chem. Solids 33, 1241 (1972).

[100] W. Zieten, Z. Phys. 145, 125 (1956).

[101] *American Institute of Physics Handbook*, edited by D. E. Gray (McGraw-Hill, New York, 1963), p. 4–66.

[102] H. Ibach, Phys. Status Solidi 31, 625 (1969).

[103] N. A. Goryunova, *The Chemistry Diamond-Like Semiconductors* (M.I.T. Press, Cambridge, Mass. 1965), p. 207.

[104] H. Ibach, Phys. Status Solidi 33, 257 (1969).

[105] J. A. Van Vechten, Phys. Rev. 187, 1007 (1969).