PHYSICAL REVIEW B VOLUME 21, NUMBER 8 15 APRIL 1980

Simple parametrized model for Jahn-Teller systems: Vacancy in p-type silicon

G. A. Baraff, E. O. Kane, and M. Schlüter
Bell Laboratories, Murray Hill, New Jersey 07974
(Received 3 October 1979)

We propose a simple and quite general model for charge states and their activated relaxation in localized defects. Applying the model to the analysis of Watkins's electron-paramagnetic resonance and deep-level-transient-spectroscopy observations of vacancies in p-type silicon, we obtain two constraints on the model's three parameters. This allows us to conclude (1) that the vacancy is indeed an "Anderson negative-U" system as we had earlier calculated, and (2) that the Jahn-Teller stabilization energy of the V^0 neutral vacancy cannot substantially exceed 0.8 eV.

I. INTRODUCTION

The detailed electronic structure of the isolated vacancy in silicon is as fundamental a problem for understanding deep-level defects as were the isocoric donor levels in that same material for understanding shallow defects. Much of what is known about it has been obtained from Watkins's[1] observations of the pressure and temperature dependence of the electron-paramagnetic-resonance (EPR) signal strength and its thermally activated decay. His studies make it clear that Jahn-Teller distortions of atoms near the vacancy are important in determining the structure of this particular point defect. We have recently proposed a model for this system,[2] based on Watkins's ideas and certain extensions of them made necessary by the results of our self-consistent calculations of the silicon vacancy,[3] and showed that, within our computational accuracy (0.1–0.2 eV), our calculations agree with Watkins's experiments in p-type material. One of the interesting consequences of our work was the prediction that the vacancy in silicon might exhibit Anderson "negative-effective-U" behavior,[4] in which Jahn-Teller relaxation overcomes the Coulomb-energy cost when a second electron is added to the state. In such a system, capacitance-transient experiments would characteristically reveal the liberation or capture of *two holes* for each vacancy.

In this present work, we approach the problem from a complementary and more general direction. We make *no* theoretical estimates of model parameters, but instead focus on general features of the kinetics and statistical mechanics of the model itself, describing it for the first time in adequate detail. In the end, we fit its parameters to Watkins's experiments. The resulting fit, while not unique, does put an upper limit of about 0.21 eV on the Jahn-Teller stabilization energy of the V^+ center in silicon. This limit is appreciably smaller than what Watkins's experiments led him to infer. It must be understood, however, that his inferences were based on theoretical estimates of the motion of atoms adjacent to the vacancy in response to externally applied strain.[5] By contrast, the upper bound we obtain here depends in no way on knowledge of lattice stiffness or response, but is based directly on the thermal experiments themselves.

The essential physics is that there are states associated with the vacancy which, when they are filled, cause the lattice to deform in response to the presence of the added localized charge. The resulting problem—interaction of a localized state with a deformable imperfect lattice—is an old one, and a series of important papers describe its solution. Huang and Rhys[6] calculated the nonradiative decay rate of the F center in the alkali halides, making the simplifying assumption that all of the phonon modes involved in the transition had the same frequency. Lax[7] critically analyzed the approximations commonly made in studying problems of this type and developed a formalism which could be used when the phonon modes involved in the interaction were distributed over a continuous spectrum of frequencies. Kubo and Toyazawa[8] independently developed essentially the same formalism as did Lax for dealing with the phonon continuum, and discussed its application to radiative transitions in useful detail. Their formalism is still widely used for studying specialized aspects of the problems.[9]

The papers cited above considered transitions in which the electron changes from one localized state to another. The idea that the same formalism could describe nonradiative capture and recombination, i.e., transitions between *de*localized states and the localized state on the defect, was contributed by Henry and Lang,[10] who calculated activated cross sections for these processes using the simplified form of theory in which a single coordinate describes the response of the lattice. This same sort of model is used by Pons and

Makram-Ebeid[11] for calculating the effect of electric field on phonon-assisted tunneling rate of electrons from deep levels in GaAs.

Our intent here is to describe how the activation energy and the occupation of the localized state depends on the Fermi energy of the system. We also wish to describe a situation in which the state of interest may be occupied by zero, one, or two electrons. We abandon, at the outset, any attempt to obtain actual cross sections or to describe transitions in which tunneling occurs. This allows us to describe the lattice in *completely* classical terms, a description which is simple enough to make clear what the new features are and yet direct enough to suggest how the rigorous quantum-mechanical treatments should be extended to include the ideas we are proposing.

In the next section of the paper, we describe the basic model for the silicon vacancy in the simplest possible way, using a single coordinate and its conjugate momentum to describe the lattice, and building in the essential features of the states which we believe are associated with the vacancy in p-type silicon. At the end of the section we explain why we think our model may be applicable to a large number of other deep levels in addition to the vacancy. In Sec. III we describe the statistical properties of the model and in Sec. IV we discuss fitting the parameters of the model to experiment.

II. THE SIMPLEST CLASSICAL MODEL

Our basic model for the vacancy state in p-type silicon builds on the fact that self-consistent calculations of the neutral *undistorted* vacancy reveal that a three-fold orbitally degenerate state of T_2 symmetry resides in the gap, and that for neutrality, this state must contain two of the six electrons it could accommodate. Although mechanically stable when empty (the V^{++} charge state), occupying that state with one electron, yielding V^{+}, or with two electrons, yielding V^{0}, produces a situation which is unstable with respect to Jahn-Teller distortions. These split the T_2 level into a singly degenerate low-energy state of B_2 symmetry and a doubly degenerate higher-energy state of E symmetry. In p-type material, only the lower-energy state can be occupied. This picture is, aside from the reference to V^{++}, exactly what Watkins has proposed for the vacancy in p-type silicon.[1]

We treat the lattice motion classically, which is all that is necessary to get activation energies. We expand to second order in the single lattice-displacement parameter. This neglect of anharmonic effects is a simplifying assumption, justified at this stage only by the lack of compelling evidence that they are needed here.

The coordinates of our model are Q, the amplitude of the Jahn-Teller distortion, P, the momentum conjugate to Q, and N_T, the number of electrons (0, 1, or 2) occupying the B_2 level. In order to discuss the statistical occupation of the vacancy, we must include an electron reservoir as part of the system. We specify transfer energies as follows: μ is the energy to transfer an electron from the top of the valence band to the reservoir, $\epsilon_1(Q)$ is the energy to transfer an electron from the top of the valence band to the B_2 level when the level is initially empty and the lattice distortion is Q, and $\epsilon_2(Q)$ is the energy required to transfer a second electron to the B_2 level when it already contains one electron. Equilibrium transitions in which only N_T changes, leaving occupation numbers for all other states of the system fixed, imply compensating changes in the occupancy of the Fermi reservoir. The total energy of this part of the system $E(Q,P,N_T)$ is expanded to second order in Q. Because the state of the crystal at $Q=0$ is of Γ_1 symmetry and the distortion Q is *purely* symmetry lowering, the linear term in the energy is contributed only by ϵ_1 and ϵ_2. Thus, we have

$$E(Q,P,N_T=0)=P^2/2M+\tfrac{1}{2}kQ^2, \tag{1a}$$

$$E(Q,P,N_T=1)=P^2/2M+\tfrac{1}{2}kQ^2+\epsilon_1(Q)-\mu, \tag{1b}$$

$$E(Q,P,N_T=2)=P^2/2M+\tfrac{1}{2}kQ^2+\epsilon_1(Q)+\epsilon_2(Q)-2\mu. \tag{1c}$$

Because of e-e repulsion, we expect $\epsilon_2(Q)>\epsilon_1(Q)$, but we do not expect that this difference will be strongly dependent on Q. Therefore, we shall set $\epsilon_2(Q)=\epsilon_1(Q)+U$, where U is a constant. We expand $\epsilon_1(Q)$ to first order so that $\epsilon_1(Q)=\epsilon_L-VQ$. These last two assumptions render the quadratic Q dependence of the system indepencent of N_T. Although we do not expect the spring constants to depend *strongly* on N_T, the assumption of neglecting the N_T dependence altogether should, at this stage, be regarded as a plausible simplification which will undoubtedly be reexamined when it becomes possible to calculate total energies with sufficient accuracy. Inserting the forms for $\epsilon_1(Q)$ and $\epsilon_2(Q)$ into (1) gives

$$E(Q,P,N_T)=P^2/2M+\tfrac{1}{2}k(Q-N_TV/k)^2+E(N_T)-N_T\mu, \tag{2}$$

where

$$E(0)=0,\quad E(1)=\epsilon_L-E_{JT},\quad E(2)=2\epsilon_L+U-4E_{JT}, \tag{3}$$

and $E_{JT} \equiv V^2/2k$.

A useful configuration-coordinate diagram which one might consider for this system is that in which the potential-energy part of the total energy, namely $E(Q, P=0, N_T)$, is plotted as a function of Q while N_T labels the three curves. Such a plot corresponds to the usual configuration-coordinate diagram in many ways. As we shall see below, the minimum energy on each curve determines the equilibrium occupation of each state N_T, and the lowest of the three minima identifies the ground state of the system. Because $E(Q,P,N_T)$ depends on μ, however, the state N_T for which the system energy is lowest depends on μ, as is illustrated in Fig. 1 (top and middle) using parameters as calculated in Ref. 2.

We let $E_{N_T}(\mu) = E(N_T) - N_T\mu$ denote the minimum energy on the N_T curve. By plotting $E_{N_T}(\mu)$ vs μ as done in Fig. 1 (bottom), we can determine which N_T is the ground state for any value of μ and thus determine the μ dependence of the equilibrium charge state of the vacancy.

We should note at this point that our model is so simple that it is actually quite general. It assumes a degenerate level at an energy ϵ_L above the valence band which necessarily produces a lattice distortion when occupied (Jahn-Teller theorem).

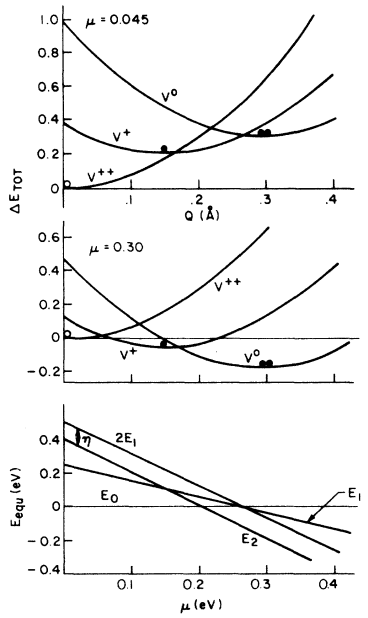

FIG. 1. Total potential energy $E(Q,N_T)$ as a function of distortion Q for three different charge states ($N_T = 0,1,2$, i.e., V^{++}, V^{+}, V^{0}). Two situations for different Fermi-level position are shown (top and middle figure). The bottom panel shows the equilibrium energies as a function of Fermi-level position. The parameters used in the plots are calculated in Ref. 2.

We postulate a repulsive energy U between two electrons occupying the level. Electrons in semiconductors are enough delocalized that one does note expect the model parameters to depend strongly on occupancy; hence the Jahn-Teller energy for double cocupancy is just four times E_{JT}, the distortion energy for single occupancy. By the same argument we expect U to be reasonably small. If only energy is measureable, our model is described by three parameters ϵ_L, U, and E_{JT}, which are few enough to be fitted by a small number of experiments. The model may well be applicable to a great many deep levels in semiconductors. Note that the well-known effective-mass acceptor levels are formally of this type. However, the spread-out character of their wave function is usually thought to reduce the Jahn-Teller coupling. For heavy impurities such as In in Si which are bound to about the same extent as the vacancy, spin orbit splitting can quench substantial Jahn-Teller distortion. Light elements such as O in GaP might be another example for strongly Jahn-Teller distorted systems.

III. STATISTICAL TREATMENT

Now we turn to the statistical description of the system. The equilibrium probability distribution is the canonic one,[12]

$$\mathcal{P}(Q,P,N_T) = C\exp[-E(Q,P,N_T)/k_BT] \tag{4}$$

with the normalization constant C being chosen so that $\sum_{N_T}\int \mathcal{P}\,dQ\,dP = 1$. The probability that the vancancy is occupied by N_T electrons is given by intergrating (4) over all phase space QP. Introducing $\omega^2 \equiv k/M$, we obtain

$$\mathcal{P}(N_T) = (2\pi k_BTC/\omega)\exp\{-[E(N_T)-N_T\mu]/k_BT\} \tag{5a}$$

$$= \frac{\exp\{-[E(N_T)-N_T\mu]/k_BT\}}{\sum_{N_T'}\exp\{-[E(N_T')-N_T'\mu]/k_BT\}}\,. \tag{5b}$$

An alternative statement of (4) and (5) is

$$\mathcal{P}(Q,P,N_T) = \left(\frac{\omega}{2\pi k_BT}\right)\mathcal{P}(N_T)\exp[-\epsilon(P,Q,N_T)/k_BT], \tag{6a}$$

where

$$\epsilon(P,Q,N_T) \equiv P^2/2M + \tfrac{1}{2}k(Q - N_TV/k)^2\,. \tag{6b}$$

This latter form is useful when the lattice has achieved thermal equilibrium while the occupation distribution $\mathcal{P}(N_T)$ has not yet had time to equilibrate to the form given by Eq. (5). This condition may occur at low temperatures when charge relaxation is a thermally activated process.

Consider some of the general features of such a process. Suppose that the vacancy in the state

Q, P, N_T undergoes a transition to the state Q', $P', N_T' = N_T + 1$. There is some other state in the system [call it $\psi_R(r), \epsilon_R$] which supplies that extra electron. It can do so if it is filled, which brings in its average occupation factor $f_O(\epsilon_R)$. There is an inherent rate for the transition, which we denote by $M_R^2(QPN_T; Q'P'N_T')$, using the subscript R to denote the other state involved. According to microscopic reversibility, M_R^2 is the same for $QPN_T \to Q'P'N_T'$ as for $Q'P'N_T' \to QPN_T$. The one-electron transition conserves total energy, which means that $E_{TOT} = E'_{TOT}$, where

$$E_{TOT} = \epsilon(P, Q, N_T) + E(N_T) + \epsilon_R , \tag{7a}$$

$$E'_{TOT} = \epsilon(P'Q'N_T') + E(N_T') \equiv \epsilon' + E(N_T'). \tag{7b}$$

Thus $R_R(N_T \to N_T')$, the total rate of transitions $N_T \to N_T'$ via the state R, is obtained by multiplying the initial probability $f_O(\epsilon_R)\mathcal{P}(Q, P, N_T)$ by the inherent rate M_R^2 and a delta function to conserve energy, then integrating over all initial phase space QP and all final phase space $Q'P'$:

$$R_R(N_T \to N_T') = \left(\frac{\omega}{2\pi k_B T}\right)\mathcal{P}(N_T) f_O(\epsilon_R) \times \int dQ\, dP\, dQ'\, dP' \exp(-\epsilon/k_B T) \times M_R^2(QPN_T; Q'P'N_T')\,\delta(\epsilon' + \Delta E - \epsilon), \tag{8}$$

where

$$\Delta E \equiv E(N_T + 1) - E(N_T) - \epsilon_R . \tag{9}$$

For the reverse transitions, the role of primed and unprimed states is interchanged, the sign of ΔE is reversed, and the state ψ_R must initially be empty.

$$R_R(N_T' \to N_T) = \left(\frac{\omega}{2\pi k_B T}\right)\mathcal{P}(N_T')[1 - f_O(\epsilon_R)] \times \int dQ'\, dP'\, dQ\, dP \exp(-\epsilon'/k_B T) \times M_R^2(Q'P'N_T'; QPN_T)\,\delta(\epsilon - \Delta E - \epsilon') . \tag{10}$$

The integral in (10) is equal to $\exp(\Delta E/k_B T)$ times the integral in (8); this follows by using the δ function to replace ϵ' by $\epsilon - \Delta E$ and invoking the microscopic reversibility symmetry of M_R^2.

When the situation is an equilibrium one, the rates of forward and back transitions via the state ψ_R are equal. Equating the rates in (8) and (10) gives

$$\frac{\mathcal{P}(N_T')}{\mathcal{P}(N_T)} = \frac{f_O(\epsilon_R)}{1 - f_O(\epsilon_R)} \exp(-\Delta E/k_B T) . \tag{11}$$

Since in equilibrium, the occupation probability for the state ψ_R is given by the Fermi function

$$f_O(\epsilon_R) = \{\exp[(\epsilon_R - \mu)/k_B T] + 1\}^{-1}, \tag{12}$$

combining (9), (11), and (12) gives

$$\frac{\mathcal{P}(N_T' = N_T + 1)}{(N_T)} = \frac{\exp\{-[E(N_T + 1) - \mu]/k_B T\}}{\exp[-E(N_T)/k_B T]} , \tag{13}$$

which is exactly what the statistical distribution (5) predicts.

Now let us specify other general features of the transition probability M_R^2, arguing on the basis of a classical picture which is closely related to the Born-Oppenheimer adiabatic approximation. We may suppose that the actual electronic transition is of short time duration, and that the force $F(t)$ it exerts on the lattice is an impulsive one $[\Delta P = \int_{\Delta t \to 0} F(t)dt \neq 0,\ \Delta Q = \int_{\Delta t \to 0} P/M dt = 0]$. This introduces a $\delta(Q - Q')$ into the integrand. Considering also the relative mass of the electron and the lattice, even the change in P is likely to be small. As an approximation to the small range of P' that can result from a given P, we insert a $\delta(P - P')$ times a $\xi(P)$, which represents the range of P', into the integrand. This reduces the integral in (8) to the form

$$\exp(-x_R^2 E_{JT}/k_B T) \int dP \exp(-P^2/2Mk_B T)\Phi_R(P), \tag{14}$$

where

$$x_R \equiv \tfrac{1}{2} + \Delta E/2E_{JT} \tag{15}$$

and where $\Phi_R(P) = \xi(P)M_R^2/V$, the arguments of M_R^2 being fixed in accordance with the three delta functions.

Since the exponential in the integrand of (14) is sharply peaked about $P = 0$, we can replace $\Phi_R(P)$ in the integrand by $\Phi_R(0)$ and complete the evaluation of the integrand using[13]

$$\int dP \exp(-P^2/2Mk_B T) = 2\pi M k_B T^{1/2} .$$

The expression for (8) is therefore

$$R_R(N_T \to N_T + 1) = \left(\frac{k}{2\pi k_B T}\right)^{1/2} \mathcal{P}(N_T) f_O(\epsilon_R)\Phi_R(0) \times \exp(-x_R^2 E_{JT}/k_B T) \tag{16a}$$

and the corresponding expression for (10), where the integral differs by a factor of $\exp(\Delta E/k_B T)$, is

$$R_R(N_T + 1 \to N_T) = \left(\frac{k}{2\pi k_B T}\right)^{1/2} \mathcal{P}(N_T + 1)[1 - f_O(\epsilon_R)]\Phi_R(0) \times \exp(-y_R^2 E_{JT}/k_B T), \tag{16b}$$

where

$$y_R^2 E_{JT} = x_R^2 E_{JT} - \Delta E . \tag{17}$$

Again, a configuration-coordinate diagram is useful, this time, to describe the transition process. In Fig. 2 we plot the potential energy part of E_{TOT} [Eq. (7a) with P set equal to zero] vs Q

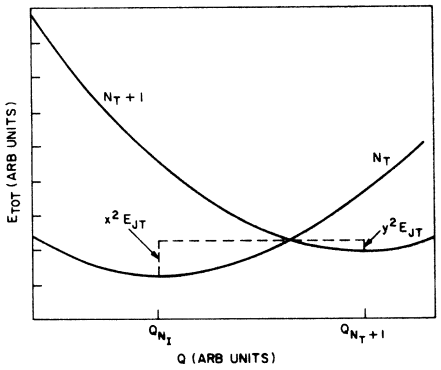

FIG. 2. Total potential energy $E(Q,N_T)$ as a function of Q for two different charge states N_T, N_T+1. The characteristic activation energies x^2E_{JT} and y^2E_{JT} are indicated as defined in the text.

and, on the same axes, the potential energy part of E'_{TOT} [Eq. (7b) with P set equal to zero and Q' set equal to Q]. Then one finds that, just as in the standard treatment of thermal-activated crossings, the energy $x_R^2E_{JT}$ which appears in the exponent of (16a) is the energy difference between the crossing point and the minimum of the initial-state curve (N_T), while the energy $y_R^2E_{JT}$ which appears in the exponent of (16b) is the difference between the crossing point and the minimum of the initial-state curve (N_T+1).

Equation (16) contains another temperature-dependent factor, namely $f_O(\epsilon_R)$, whose temperature dependence may be sufficiently exponential-like, that it alters the activation energy from the $x_R^2E_{JT}$ or $y_R^2E_{JT}$ which appears explicitly in Eq. (16). For example, suppose that at low temperature, $(\epsilon_R-\mu)/k_BT$ is large enough that we can approximate $f_O(\epsilon_R)$ by $\exp[-(\epsilon_R-\mu)/k_BT]$ and $[1-f_O(\epsilon_R)]$ by 1 if $\epsilon_R>\mu$, while for $\epsilon_R<\mu$, we approximate $f_O(\epsilon_R)$ by 1 and $[1-f_O(r)]$ by $\exp[-(\mu-\epsilon_R)/k_BT]$. Then the activation energies associated with (16a) and (16b), respectively, appear to be

$$(N_T-N_T+1)_{ACT}=x_R^2E_{JT}+(\epsilon_R-\mu)\theta(\epsilon_R-\mu), \tag{18a}$$

$$(N_T+1-N_T)_{ACT}=y_R^2E_{JT}+(\mu-\epsilon_R)\theta(\mu-\epsilon_R), \tag{18b}$$

where $\theta(z)=1,0$ for $z>0,<0$.

The actual transitions $N_T\to N'_T$ will proceed via all states ψ_R which can supply or accept the electron needed. Thus, the total rates are

$$R(N_T-N_T+1)=\left(\frac{k}{2\pi k_BT}\right)^{1/2}\mathcal{P}(N_T)\times\sum_R[f_O(\epsilon_R)\Phi_R(0)\exp(-x_R^2E_{JT}/k_BT) \tag{19a}$$

and

$$R(N_T+1-N_T)=\left(\frac{k}{2\pi k_BT}\right)^{1/2}\mathcal{P}(N_T+1)\times\sum_R[1-f_O(\epsilon_R)]\Phi_R(0)\times\exp(-y_R^2E_{JT}/k_BT).$$

Since $f_O(\epsilon_R)$ is the Fermi function (12), while y_R is given by (17) and ΔE by (9), this second line can be rewritten as

$$R(N_T+1-N_T)=\left(\frac{k}{2\pi k_BT}\right)^{1/2}\mathcal{P}(N_T+1)\times\exp\{[E(N_T+1)-\mu-E(N_T)]/k_BT\}\times\sum_R f_O(\epsilon_R)\Phi_R(0)\exp(-x_R^2E_{JT}/k_BT). \tag{19b}$$

Note that the summations in (19a) and (19b) are identical.

For the vacancy in p-type silicon, there are only two types of states ψ_R energetically available to contribute to the transition rate. One of the states ψ_R might be an acceptor whose energy lies low in the gap. For this state to be involved, the inherent rate $M_R^2(QPN_R-Q'P'N'_T)$ must not vanish. But the quantum analog of this inherent rate is essentially the matrix element of the nuclear kinetic-energy operator taken between the Born-Oppenheimer adiabatic states. From this, we can infer that the rate will depend strongly on the overlap between the wave function ψ_R and the vacancy wave function ψ_{B2}, i.e., the wave functions of the two states whose occupancy changes during the transition. Thus, direct transfer of an electron between an acceptor and the vacancy can be important only if the acceptor concentration is great enough that a given vacancy has an acceptor sufficiently nearby.

At low acceptor densities, it is far more likely that the state ψ_R involved in the transition is one of the valence-band states. Let us consider the sum in (19) where the states contributing are those in the valence band. At low temperatures and μ in the gap, we can replace $f_O(\epsilon_R)$ by 1, and can consider the exponential to be a sharp enough function of ϵ_R so that we can replace $\Phi_R(0)$ in the summation by $\Phi_{\bar{R}}(0)$, where $\bar{R}$ denotes the energy which contributes most strongly to the sum. Assuming a density of states $\rho(\epsilon)$ at the top of the valence band, we have

$$\sum_R f_O(\epsilon_R)\Phi_R(0)\exp(-x_R^2E_{JT}/k_BT)\approx\Phi_{\bar{R}}(0)\int_{-\infty}^{0}\rho(\epsilon)d\epsilon\exp[-(a-\epsilon)^2/A], \tag{20a}$$

where

$$x_R^2E_{JT}\equiv(a-\epsilon)^2/A \tag{20b}$$

and where, from (9) and (15),

$$a = E(N_T + 1) + E_{JT} - E(N_T), \tag{21a}$$

$$A = 4E_{JT} k_B T. \tag{21b}$$

The method of evaluating the integral in (20a) depends on the sign of a (a numerical evaluation is always possible). If $a > 0$, then the main contribution to the integral comes from ϵ near 0, i.e., from the top of the valence band. An approximate evaluation is then obtained by expanding the exponent to lowest order, after assuming a density of state $\rho(\epsilon) = c\sqrt{-\epsilon}$ near the top of the valence band:

$$\int_{-\infty}^{0} \rho(\epsilon) d\epsilon \exp[-(a-\epsilon)^2/A] = c\int_0^\infty x^{1/2} dx \exp[-(a^2 + 2ax)/A] = \frac{1}{2\pi}\left(\frac{\pi A}{2a}\right)^{3/2} c \exp(-a^2/A).$$

Then (19) becomes

$$\frac{R(N_T \to N_T + 1)}{\Phi(N_T)} = \frac{k^{1/2}\Phi_0(0)c}{\{1 + [E(N_T+1) - E(N_T)]/E_{JT}\}^{3/2} k_B T} \times \exp(-\alpha/k_B T) \tag{22a}$$

$$\frac{R(N_T + 1 \to N_T)}{(N_T + 1)} = \frac{k^{1/2}\Phi_0(0)c}{\{1 + [E(N_T+1) - E(N_T)]/E_{JT}\}^{3/2} k_B T} \times \exp(-\beta/k_B T), \tag{22b}$$

where $\Phi_0(0)$ denotes $\Phi_R(0)$ evaluated for a state at the top of the valence band and where α and β, the activation energies for the charge relaxation processes, are given by

$$\alpha = \left(\frac{1}{2} + \frac{E(N_T+1) - E(N_T)}{2E_{JT}}\right)^2 E_{JT}, \tag{23a}$$

$$\beta = \left(\frac{1}{2} + \frac{E(N_T) - E(N_T+1)}{2E_{JT}}\right)^2 E_{JT} + \mu. \tag{23b}$$

The activation energy α is exactly the difference in energy from the initial-state minimum (N_T) to the crossing of the $N_T \rightarrow N_T + 1$ configuration-coordinate curves, calculated as in Fig. 2 with the state ψ_R taken at the top of the valence band. Similarly, $\beta - \mu$ is exactly the difference in energy between the initial-state minimum ($N_T + 1$) to the crossing of the $N_T \rightarrow N_T + 1$ curves. The extra energy μ is the added activation energy needed to produce the hole which is captured in the $N_T + 1 \rightarrow N_T$ transition.

In summary, we use (3) in (23) to obtain the four activation energies associated with this system:

$$(0 \to 1)_{ACT} = \left(0 - \frac{\epsilon_L}{2E_{JT}}\right)^2 E_{JT}, \tag{24a}$$

$$(1 \to 2)_{ACT} = \left(1 - \frac{\epsilon_L + U}{2E_{JT}}\right)^2 E_{JT}, \tag{24b}$$

$$(1 \to 0)_{ACT} = \left(1 - \frac{\epsilon_L}{2E_{JT}}\right)^2 E_{JT} + \mu, \tag{24c}$$

$$(2 \to 1)_{ACT} = \left(2 - \frac{\epsilon_L + U}{2E_{JT}}\right)^2 E_{JT} + \mu. \tag{24d}$$

Clearly, when an activation energy less than μ is observed, the associated process must involve hole emission from the vacancy to the valence band, and not hole capture.

Let us digress briefly to suggest what happens in this classical model of the lattice if, instead of a single lattice coordinate Q and its momentum P, there is a continuum of lattice deformations Q_i, with their associated momenta P_i, spring constants k_i, and masses M_i. In that case, the lattice part of the energy in (1) and (2) would become a generalized quadratic form which could, in principle, be diagonalized. One would still arrive at the same distribution function (5b) but the distribution function (6a) would take the form

$$\Phi(\{Q\}, \{P\}, N_T) = \Phi(N_T) \prod_{j=1}^{N} \left(\frac{\omega_j}{2\pi k_B T}\right) \times \exp[-\epsilon(Q_j, P_j, N_T)/k_B T],$$

where ω_j is the frequency of the jth normal mode, $j = 1, \ldots, N$.

Although there is a $(1/T)^N$ dependence to this distribution function instead of the $1/T$, there will be more integrations over the coordinates when we construct rates, as in Eqs. (8), (10), and (16), and the result will be, as in (16), an overall $T^{-1/2}$ prefactor. Kubo and Toyazawa show that the quantity analogous to the $x_R^2 E_{JT}$ or $y_R^2 E_{JT}$ activation energy which appears in (16) is the difference between the lowest energy at which the two configuration-coordinate potential-energy hypersurfaces cross and the energy minimum of the initial-state hypersurface. Their proof will apply here when the algebra is generalized to include more than one normal coordinate.

IV. FITTING THE PARAMETERS TO EXPERIMENT

The first experiment to be fitted is the activation energy for the decay of the V^+ EPR signal in In-doped silicon ($\mu = 0.16$ eV) which Watkins has measured as 0.057 eV,[1] and which he ascribes to hole emission (a transition $N_T = 1 \to 2$ via capture of an electron from the top of the balence band). Once we postulate the existence of V^{++}, another channel of V^+ decay, namely $V^+ \to V^{++}$ must be considered. However, the activation energy for this channel (24c) is greater than μ because a hole must be present in the valence band before it can be captured by the vacancy. The numbers just quoted rule out this channel for V^+ decay. The

possibility of V^+ decay via direct transfer between the vacancy and an In acceptor should also be small, and so we are led to accept Watkins's assignment of the 0.057 activation energy to V^+ decay via hole emission. From (24b), we have $[(\epsilon_L + U)/2E_{JT} - 1]^2 E_{JT} = 0.057$ eV. This condition is conveniently regarded as fixing $(\epsilon_L + U)$ as a function of E_{JT}, and we exhibit this in columns 1 and 2 of Table I.

The second experiment to be fitted is the activation energy of 0.13 eV which Watkins observes by deep-level-transient-spectroscopy (DLTS) measurements in p-type silicon.[14] Watkins has suggested to us privately that the transition being observed here might be the $V^{++} \rightarrow V^0$ transition which we had proposed,[2] and has suggested that at these DLTS temperatures, after the initial $V^{++} \rightarrow V^+$ transition, the subsequent $V^+ \rightarrow V^0$ transition is so rapid that only the $V^{++} \rightarrow V^+$ step is observed as an activated process in DLTS. This suggestion seems reasonable, and from (24a), we obtain $\epsilon_L^2/4E_{JT} = 0.13$. This condition is conveniently regarded as fixing ϵ_L as a function of E_{JT}, and we exhibit this in column 3 of Table I.

Columns 4, 5, and 6 of Table I give certain quantities which can be calculated from entries in columns 1, 2, and 3. Column 4 gives the Coulomb repulsion term $U = (\epsilon_L + U) - \epsilon_L$. Column 5 gives $\eta \equiv 2E_{JT} - U$. If $\eta > 0$, then there is no value of μ for which V^+ is the ground state of the system,[2] that is, the equilibrium configuration changes directly from V^{++} to V^0 as the Fermi energy μ is raised. Notice that $\eta > 0$ for all entries in the table. This implies that we do, in fact, have an "Anderson negative-U" system in this entire range of E_{JT} values. Column 6 gives $E(0/++) = \epsilon_L - 2E_{JT} + U/2$, which is the value of μ at which the ground state changes from V^{++} to V^0 as μ is raised.

TABLE I. Characteristic energies of Jahn-Teller distorted vacancies obtained by fitting two independent experiments to the three parameters of the model (see text). The magnitude of E_{JT} (column 1) is to be regarded as the remaining independent variable. Values below the dotted line can be ruled out by the existence of a deep-level-transient-spectroscopy (DLTS) signal in gallium-doped silicon.

1 E_{JT}	2 $\epsilon_L + U$	3 ϵ_L	4 U	5 η	6 $E(0/++)$
0.0500	0.2068	0.1612	0.0455	0.0545	0.0840
0.1000	0.3510	0.2280	0.1230	0.0770	0.0895
0.1500	0.4849	0.2793	0.2056	0.0944	0.0821
0.2000	0.6135	0.3225	0.2911	0.1089	0.0680
0.2500	0.7387	0.3606	0.3782	0.1218	0.0497
0.3000	0.8615	0.3950	0.4666	0.1334	0.0283
0.3500	0.9825	0.4266	0.5559	0.1441	0.0046
0.4000	1.1020	0.4561	0.6459	0.1541	−0.0210

Consider now the implications of being able to observe a $V^{++} \rightarrow V^+$ transitions in DLTS. The essence of a DLTS experiment is that the charge state of the deep level should alter (producing an observable change in the capacitance of a pn junction) when all the mobile carriers are swept away from the deep level. This means that its charge state under equilibrium conditions (as determined by the Fermi energy μ) and its charge state under depletion (when all mobile carriers are swept away) *cannot* be the same. Watkins's experiments make it clear that the state is a hole trap, close to the valence band, so that in depletion, it is occupied with electrons, just as all the nearby valence band states are fully occupied by electrons when there are no holes present. For this level to be observed, it must therefore be in the state V^+ or V^{++} during that brief time at the beginning of the experiment when its occupation is determined by μ. This requires that the value of μ be *less* than $E(0/++)$ in the sample in which the DLTS signal is observed. Watkins reports seeing the 0.13-eV activation energy in p-type silicon samples doped with boron ($\mu = 0.045$), aluminum ($\mu = 0.057$), and gallium ($\mu = 0.065$). We conclude then that $E(0/++) > 0.065$ eV. According to Table I, this implies a value of E_{JT} which is less than 0.21 eV, and establishes, thereby, allowable ranges of ϵ_L and U.

Watkins also describes an attempt to measure the excitation energy for the state V^+ in boron-doped silicon by observing the strength of the EPR signal from V^+ at a finite temperature and comparing it with the strength of the EPR signal when the number of vacancies in the V^+ state has been enhanced by flooding the sample with infrared photons.[1] The excitation energy for V^+ [i.e., the amount by which the energy of the system in the state V^+ exceeds the energy of the ground state V^{++} when the Fermi energy μ is less than $E(0/++)$] is, in our model,

$$E_{ex} = E_1(\mu) - E_0(\mu) = \epsilon_L - E_{JT} - \mu .$$

For a value of $\mu = 0.045$ eV and values of ϵ_L and E_{JT} from the allowed part of Table I ($E_{JT} < 0.21$ eV) we find $0.066 < E_{ex} < 0.077$. If Watkins's measurement does indeed determine the excitation energy, then one would conclude that $E_{ex} = 0.006$ eV. The discrepancy, of 0.06 to 0.07 eV, between our prediction and his measurements, would serve as a measure of the validity of the assumptions, such as those spelled out at the beginning of this paper, on which our model is based.

There is interestingly a qualitatively new feature which should be considered in redoing the excitation energy experiment. If the equilibrium charge

state in the boron- , or aluminum- , or gallium-doped samples is V^{++} as we believe, then the mechanism for generating V^{+} by flooding the sample with radiation cannot be capture of the holes produced by ionizing the acceptors. It must be instead direct optical excitation of an electron from the top of the valence band to the vacancy. This process requires a photon energy $h\omega > \epsilon_L$, and so measuring the optical threshold for V^{+} production will give a direct measurement of one of the parameters of our model. In In-doped silicon, on the other hand, where $\mu = 0.16$, the value $E(0/++) < 0.084$ indicated in Table I predicts that the ground state of the vacancy is V^{0}. This can be converted to V^{+} by hole capture, so that any illumination capable of creating holes in the valence band should be effective in generating the V^{+} in an In-doped sample.

Our conclusions are summarized in Table II. Also included in Table II are the estimates of the parameters which we had determined earlier.[2] The largest discrepancy is in the value of ϵ_L, which we had determined in part from the Green's function self-consistent calculations and in part from a guess that the atoms near the vacancy might relax outward by an amount comparable to that on the silicon (111) surface. It is likely that our guess of an outward relaxation and the rise of 0.1 eV, in the level position from 0.32 to 0.42 eV, was too generous an estimate. The main point is that we have obtained a rather convincing demonstration that the Anderson effective U for this system is indeed negative, and that we have obtained a rather convincing limit on E_{JT} of about 0.21 eV without relying on estimates of lattice response and the uncertainties that they introduce. We hope that by presenting this analysis, we will stimulate further experiments which can uniquely fix the parameters of this model. We also suggest a search for other systems which may be treated in a similar way, in accord with the discussion at the end of Sec. II.

TABLE II. Comparison of calculated model parameters (from Ref. 2) with bounds obtained by fitting to experimental data (from Refs. 1 and 14).

Model calculations (Ref. 2)	Parameter fit to experiments (Refs. 1 and 14)
$E_{JT} = 0.17$ eV	$E_{JT} \leq 0.21$ eV
$\epsilon_L = 0.42$ eV	$\epsilon_L \leq 0.33$ eV
$U = 0.25$ eV	$U < 0.33$ eV
V^{+} metastable	V^{+} metastable

ACKNOWLEDGMENTS

We would like to thank C. H. Henry and D. V. Lang for their suggestions on an earlier account of this work and G. D. Watkins for discussing his work with us.

[1] G. D. Watkins, in *Lattice Defects in Semiconductors—1974*, edited by F. A. Huntley, Institue of Physics and Physical Society Conference Series No. 23 (Institute of Physics and Physical Society, London, 1975), p. 1, and references therein.

[2] G. A. Baraff, E. O. Kane, and M. Schlüter, Phys. Rev. Lett. 43, 956 (1978).

[3] G. A. Baraff and M. Schlüter, Phys. Rev. 19, 4965 (1979).

[4] P. W. Anderson, Phys. Rev. Lett. 34, 953 (1975).

[5] L. P. Larkins and A. M. Stoneham, J. Phys. C 4, 143 (1971); 4, 154 (1971).

[6] K. Huang and A. Rhys. Proc. R. Soc. A204, 406 (1950).

[7] M. Lax, J. Chem. Phys. 20, 1752 (1952).

[8] R. Kubo and Y. Toyazawa, Prog. Theor. Phys. 13, 160 (1955).

[9] See, for example, M. Kusunoki, Phys. Rev. B 20, 2512 (1979) and references therein.

[10] C. H. Henry and D. V. Lang, Phys. Rev. 15, 989 (1977).

[11] D. Pons and S. Makram-Ebeid, J. Phys. (Paris) (unpublished).

[12] If spin degeneracy is important, the $N_T = 1$ probability should be twice what we have written here since the $N_T = 1$ state, unlike the other two, can be occupied in two ways, spin up and spin down. The resulting formalism is unnecessarily complicated for what we want to study here, however.

[13] If $\Phi_R(P)$ goes as P^n as $P \to 0$, the dependence on $k_B T$ will go as $(k_B T)^{(n+1)/2}$ instead of as $(k_B T)^{1/2}$. This change affects the prefactor but not the activation energy in all that follows.

[14] G. D. Watkins, J. R. Troxell, and A. P. Chatterjee, in *Defects and Radiation Effects in Semiconductors—1978*, edited by J. H. Albany, Institue of Physics and Physical Society Conference Series No. 46 (Institute of Physics and Physical Society, London, 1979), p. 16.