PHYSICAL REVIEW B VOLUME 13, NUMBER 4 15 FEBRUARY 1976

Effect of electron-hole pairs on phonon frequencies in Si related to temperature dependence of band gaps

V. Heine*
Cavendish Laboratory, University of Cambridge, Cambridge, England

J. A. Van Vechten*
IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598

(Received 30 June 1975)

The direct and indirect gaps between valence and conduction bands in semiconductors usually decrease with temperature. This effect is related by thermodynamic identities to the influence of electron-hole pairs on the lattice vibration frequencies. We show that the surprisingly large magnitude of the effect in Si and similar semiconductors is related to the sensitivity of the transverse-acoustic modes to covalent bonding. We are able to account for the magnitude of the effect from zero to the melting temperature. We also account for anomalous temperature variation in HgTe and related cases and mention other applications of the theory.

The theory of the temperature dependence of the various direct and indirect band gaps in solids can be cast in two different forms. The more usual considers the effect of the lattice vibrations in giving a temperature-dependent electronic band structure,[1-3] and involves four terms,[4] namely, the Debye-Waller factors in the Fourier expansion of the periodic part of the potential,[2] the Fan intraband terms,[1] the corresponding interband contributions which have hardly ever[3] been considered, and the thermal expansion. Almost all past calculations have been restricted to the expansion plus either the Fan terms[1] or the Debye-Waller factors,[2] but all four effects should be included together.[4] The situation is analogous to x-ray scattering with the Debye-Waller factor at reciprocal-lattice vectors plus incoherent scattering from thermal disorder at other wave vectors. Calculations in this theory are complicated by the presence of the four terms and partial cancellations[4] between them.

We have therefore returned to the equivalent[4] formulation of the theory by Brooks,[5] which is a simple consequence of thermodynamics and expresses the temperature dependence of the gap as the effect of the electrons on the phonons instead of vice versa. Its advantage is that Debye-Waller, Fan, and interband terms can be replaced by a single entity, thus allowing simple models and physical insights to be applied, and interesting extensions to be made, e.g., to carriers bound to vacancies and other traps in semiconductors.[6,7]

We will use here the results of the bond charge model[8,9] of calculating the phonon frequencies in tetrahedral structures to determine the effect of electron-hole, n-p, pairs on these frequencies. We treat the case of the fundamental gap E_{cv} of Si explicitly. We calculate $E_{cv}(T)$ without empirical adjustment from very simple considerations and find the result never differs from experiment[10] by more than 4.0 meV between $T=0$ and 300 K. This result could easily be improved if a more detailed account were made of the Fan terms in order to obtain the T^2 variation at low T.[1,4] (See below.) The variation of E_{cv} above 300 K is shown to be dominated by an anharmonicity in the TA phonons. We relate this to the competing β-tin phase[11] that may be reached through a large TA distortion. Thus, we are also able to calculate dE_{cv}/dT at high temperature with good accuracy and no empirical adjustment.

To be precise, Brooks[5] relates $E_{cv}(T)$ to the change in the lattice vibration frequencies from ω_i to ω_i' when an n-p pair is excited across any particular gap. The gap is a difference in chemical potential,[5] given by

$$E_{cv}(T)-E_{cv}(0)=\sum_i[f(\omega_i',T)-f(\omega_i,T)]$$
$$\approx kT\sum_i \ln\left(\frac{\omega_i'}{\omega_i}\right) \quad \text{for } kT>\hbar\omega_i, \tag{1}$$

where $f(\omega,T)$ is the standard formula[12] for the free energy of an oscillator of frequency ω. It is convenient to define the corresponding standard entropy ΔS_{cv} of the fundamental gap, i.e., of creating an n-p pair:

$$\Delta S_{cv}(T)=-\frac{\partial E_{cv}(T)}{\partial T}\ . \tag{2}$$

In the harmonic approximation, $f(\omega,T)$ becomes linear in T at high temperatures and ΔS_{cv} tends to a constant $\Delta S_{cv}(\infty)$ as is evident in the last line of (1).

The crux of the present paper is to note how surprisingly large the temperature dependence is for diamond-type semiconductors. ΔS_{cv} for Si is $2.9k$ (k is Boltzmann's constant) at room temperature, increasing to over $5k$ at the melting point,[13,14] which is typical of other gaps in Si and in other similar materials.[15] We can appreciate how large this is from a simple estimate. The valence band consists of bonding states whereas the conduction band is antibonding, and we might suppose that one n-p pair destroys the equivalent of one bond, i.e., that the force constants of all of the $2N_A$ bonds in the crystal are reduced by a fraction $1/2N_A$, where N_A is the total number of atoms:

$$(\omega_i')^2 = \omega_i^2(1 - 1/2N_A). \tag{3}$$

Substitution in (1) for each of the $3N_A$ normal modes gives an entropy $\Delta S_{cv}(\infty)$ of only $\frac{3}{4}k$!

The much larger observed values of ΔS_{cv} force us to conclude that each n-p pair softens the lattice vibrations by as much as knocking out several bonds!

The explanation of these large values of ΔS_{cv} is that the transverse acoustic (TA) modes depend critically on the covalent bonding, without which the TA modes and the whole diamond structure go unstable.[8] The stabilizing feature is the bond charge $Z_b e$ heaped up at the center of each bond, given according to Phillips[16] by $Z_b = 2/\epsilon_0 = \frac{1}{6}$ for Si where ϵ_0 is the optical dielectric constant. Thus, one hole in the most bonding states at the top of the valence band could destroy the Z_b in six bonds, which is our qualitative explanation for the large ΔS_{cv}.

More quantitatively, we note that in the simplest version[8] of the simple bond charge model (SBCM), the frequencies of the TA modes at wave vector q, $\omega_{\text{TA}}(q)$, are linear functions of Z_b. This is because the forces which stabilize these modes arise from the interaction between the bonds[8,9]:

$$\omega_{\text{TA}}^2(q) = F(q)\,\epsilon_0 Z_b^2, \tag{4}$$

where $F(q)$ is the dispersion function which we shall take from experiment.[17] When an n-p pair is excited across E_{cv} of Si, Z_b is decreased by $(85 \pm 10\%)$ because the conduction band state, near X_1, contributes about 5% of its charge density to Z_b while the states at the top of the valence band, $\Gamma_{25'}$, contribute about 90% of their charge density to Z_b.[18] Therefore, the shift in TA-phonon frequencies due to the excitation of n_p n-p pairs is calculated as

$$\omega_{\text{TA}}'(q) = \omega_{\text{TA}}^0(q)\,(1 - f_{cv}\epsilon_0 n_p/4N_A), \tag{5}$$

where N_A is the number of atoms in the sample and $f_{cv} = 0.85 \pm 0.10$[18] is a bond charge reduction factor for the fundamental gap. (A similar reduction factor f_{if} can be obtained for the gap between any initial state i and final state f by reference to the charge density calculation.[18]) It is important to note that $\omega_{\text{TA}}^0/\omega_{\text{TA}}' - 1.0 < 10^{-3}$ even at the melting point T^f, because n_p never exceeds 2×10^{19} $\text{cm}^{-3} \approx 1\times10^{-4}$ of the valence charge density in semiconductors.[14] Therefore, our approach is well within the limits of validity of perturbation theory even near T^f. This cannot be said for the alternate approaches,[1-3] because E_{cv} varies a factor of 2 in Si, a factor of 3 in Ge, and goes through zero in cases such as InSb and InAs.[14]

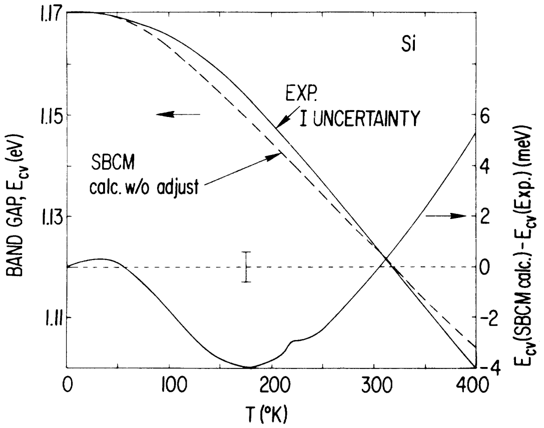

FIG. 1. Temperature dependence of the indirect band gap in Si (left scale): difference between experiment, Ref. 10, and theory using the simplest bond charge model (Ref. 8) and Eq. (5) without adjustment for Fan T^2 terms, Eq. (6) (expanded scale at right). The error bars indicate the approximate experimental uncertainty. The empirical phonon spectrum, Ref. 17, is used in the calculation.

Considering just the contribution of the N_A TA normal modes to ΔS_{cv}, one calculates $\Delta S(\infty) = (2.55 \pm 0.26)k$. Using the empirical $F(q)$,[17] the $E_g(T)$ up to 400 K is compared with experiment in Fig. 1.

Although the agreement is good considering the simplicity and absence of adjustments, there are discrepancies beyond experimental error. Our SBCM gives a T^4 variation as $T \to 0$ (as does a Debye-Waller-factor calculation), whereas the data[10] indicate a T^2 variation. (Some fraction of this aspect of the data may have been forced by the averaging procedure employed,[10] but much of it is certainly real.) The T^2 variation results from Fan terms,[1,4] which in our formulation could be expressed as

$$\omega_{\text{TA}}^1(q) = \omega_{\text{TA}}^0(q)\,(1 - f_{cv}\epsilon_0 n_p \kappa/4q^2 N_A), \tag{6}$$

where κ is a constant related to the parabolicity of the valence bands. These Fan terms apply only for q less than $\frac{1}{10}$ the radius of the Brillouin zone, i.e., over only 10^{-3} of the volume of the zone, where the valence band is parabolic[19] around Γ. A suitable choice of κ could improve our fit nearly to the level of experimental uncertainty.[4]

It is also evident from Fig. 1 that our SBCM treatment of the TA contribution to $\Delta S_{cv}(T)$ begins to seriously underestimate the empirical $\Delta S_{cv}(T)$ above 200 °K. This discrepancy increases to a

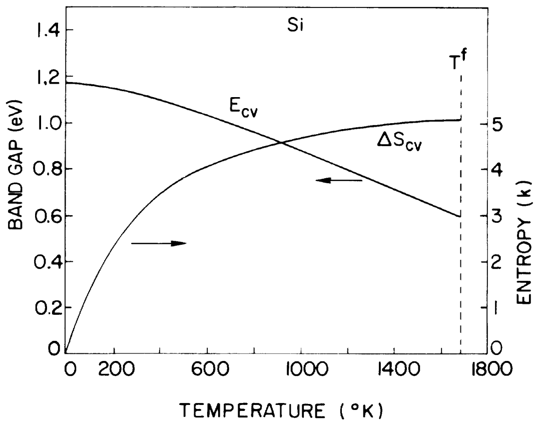

FIG. 2. Temperature variation of the indirect band gap in Si (left scale) and the entropy of this gap, $\Delta S_{cv} = -\partial E_{cv}/\partial T$ (right scale) from zero to the melting temperature T^f.

factor of 2 at T^f. (See Fig. 2.) One might first think that this is because of our neglect of the contributions from (i) thermal expansion, (ii) the TO phonons, or (iii) the LO and LA phonons. However, brief consideration shows that (i) and (ii) have the wrong sign, and all three are much too small to account for the effect. The contribution of thermal expansion is

$$-\left(\frac{\partial E_{cv}}{\partial V}\right)\left(\frac{dV}{dT}\right)<0 \tag{7}$$

because both $\partial E_{cv}/\partial V$ and dV/dT are positive for Si at these temperatures.[20] This contribution saturates around 800 °K at about $-0.2k$. The fact that the optic modes also make a negative contribution can be seen by noting in Figs. 1 and 6 of Ref. 8 that removing Z_b increases the optic frequencies. This is what must occur in any case because the contributions of the bonds are almost entirely off-diagonal (because they are not at atom sites) and consequently do not affect the trace of the dynamic force-constant tensor. Thus, the sum of the eigenvalues $\omega^2(q)$, for fixed q, is not affected. Therefore, $\omega^2_{\mathrm{TO}}(q)$ must increase when $\omega^2_{\mathrm{TA}}(q)$ decreases. The increase in E_{cv} with the excitation of optic phonons was recently shown also by Tosatti using different methods.[21] The magnitude of the TO contribution can be judged by noting that if $-\Delta\omega^2_{\mathrm{TO}}(q)=\Delta\omega^2_{\mathrm{TA}}(q)$, then

$$\Delta\omega_{\mathrm{TO}}(q)/\omega_{\mathrm{TO}}(q)\simeq -[\Delta\omega_{\mathrm{TA}}(q)/\omega_{\mathrm{TA}}(q)] \times[\omega_{\mathrm{TA}}(q)/\omega_{\mathrm{TO}}(q)]^2. \tag{8}$$

Since $\omega_{\mathrm{TO}}/\omega_{\mathrm{TA}}$ is typically 4 and much greater near the zone center, the TO contribution to ΔS_{cv}, Eq. (1), is no more than -7% of the TA contribution or about $-0.4k$ at T^f. By reference to the bond-charge-model calculations,[8,9] one can see that the LA and LO modes are almost totally unaffected by changes in Z_b. Thus, their contribution to ΔS_{cv} is essentially zero.

Therefore, it would appear the large increase in ΔS_{cv} above 200 °K must be due to anharmonicity in the TA modes. Although some anharmonicity should always be expected,[14] the factor-of-2 effect observed would seem to indicate an instability peculiar to the tetrahedral structure. The most likely source of this instability would seem to be the competing β-tin phase, which can be thought of as a major distortion of the structure resulting from a $\langle 111\rangle$ shear.[11] Such a distortion may be described as a superposition of TA modes.

Indeed, in Fig. 1 of Ref. 8 we see that the TA modes would be imaginary, i.e., unstable with $\omega^2_{\mathrm{TA}}(q)<0$ for all q, if $Z_b=0$. There is a finite value of Z_b, Z_b^*, required to stabilize the diamond structure against the competing β-tin phase. We may determine this value Z_b^* by noting that Sn has $\epsilon_0=24$ and is only marginally stable in the diamond structure, transforming to the β-tin phase reversibly above 290 °K.[22] Thus, we have $\epsilon_0^*=24$ and

$$Z_b^*=2/\epsilon_0^*=\tfrac{1}{12}. \tag{9}$$

We propose that the anharmonicity of the TA modes may be adequately described by the approximation

$$\omega^2_{\mathrm{TA}}(q)=F(q)\,\epsilon_0(Z_b-Z_b^*)^2 \tag{10}$$

for large excitations, i.e., high temperatures. We retain Eq. (4) in the low-temperature, low-excitation regime. (It may be noted that the TA-mode frequencies and shear elastic constants of Sn in the α-tin phase are finite, even appreciable, even at the α-β transition temperature.[23]) Therefore, Eq. (5) goes to

$$\omega^1_{\mathrm{TA}}(q)=\omega^0_{\mathrm{TA}}(q)\,[1-f_{cv}n_p\epsilon_0\epsilon_0^*/(\epsilon_0^*-\epsilon_0)4N_A] \tag{11}$$

at high temperatures.

If we take $\epsilon_0=12.0$, the value at room temperature and below, and $f_{cv}=0.85\pm0.10$ as above, Eq. (11) gives the TA contribution to ΔS_{cv} to be $(5.1\pm0.6)k$ at high temperatures. However, ϵ_0 itself increases with temperature due primarily to the reduction of the various band gaps and, in particular, the dielectric average gap E_g.[24,25] There is an increase in ϵ_0 due to the thermal expansion[20,24,25] because E_g varies more rapidly than the plasma frequency.

The empirical extrapolation[25] of ϵ_0 to T^f implies

$$\epsilon_0(\mathrm{Si},T^f)=13.3. \tag{12}$$

With this value, Eq. (11) gives the TA contribution to $\Delta S_{cv}(T^f)$ as $(6.4\pm0.8)k$. Combining this with the previously obtained estimates of the (negative) contribution to ΔS_{cv} from thermal expansion and from the TO modes, we obtain a final estimate

$$\Delta S_{cv}(\mathrm{Si}, T^f) = 6.4 - 0.7 \pm 0.8 = (5.7 \pm 0.8)k. \quad (13)$$

This calculation may be compared with the most recent empirical estimate

$$\Delta S_{cv}(\mathrm{Si}, T^f) = (5.1 \pm 0.2)k \quad (14)$$

obtained by Thurmond.[14] (This empirical estimate is based upon comparison of thermal conductivity measurements[26] of E_{cv} near T^f with optical and Hall-effect measurements[27] to 600 °K.) The earlier empirical estimate by Varshni[13] was $\Delta S_{cv}(\mathrm{Si}, T^f) = 6.0k$. We note that the agreement between calculation and experiment, Eqs. (13) and (14), is within experimental uncertainty.

We conclude by noting that our model explains why all band gaps in one material have the same functional form of temperature dependence [but different amplitudes due to different bond charge reduction factors f_{if} in Eq. (5)]. It seems that the data are not adequate for a critical test.[15] We can split ΔS_{if} into contributions from the valence and conduction bands, respectively. The states at Γ, X, and L in the top valence band are all very bonding in character,[28] in agreement with the observation of Auvergne *et al.*[29] that they appear to move equally with temperature. However, the states in the conduction band differ more widely[18] from one another, L_1 for example being more antibonding than X_1, giving the E_1 gap at L somewhat greater temperature dependence than the E_2 gap at X.[14]

HgTe, HgSe, and $\mathrm{Hg}_{1-x}\mathrm{Cd}_x\mathrm{Te}$ with $x < 0.45$ are interesting cases because the observed gap increases with temperature.[30] These are the exceptions that "prove the rule." The gap corresponding to $E(\Gamma_{2'}) - E(\Gamma_{25'})$ in Si (the direct gap at the zone center in GaAs, CdTe, etc.), becomes $E(\Gamma_6) - E(\Gamma_8)$ in HgTe and is negative due to strong relativistic effects. Thus, the excitation of an *n-p* pair implies the removal of an electron from the antibonding-*s* Γ_6 state, which is below the gap, and the addition of an electron into a bonding-*p* Γ_8 state in the conduction band. This process will stiffen the lattice and, by Eq. (5), will give a positive temperature coefficient. If we consider the alloys $\mathrm{Hg}_{1-x}\mathrm{Cd}_x\mathrm{Te}$ with $x < 0.45$, we note that the positive temperature coefficient persists even for values of $x > 0.15$ for which the fundamental band gap E_{cv} is positive. The reason for this is that the disorder scattering of the alloy crystal potential mixes the symmetry character of the wave functions forming the edges of the valence and conduction bands.[31,32] Whereas these wave functions are pure *p* like, $\Gamma_{25'}$, or Γ_8, and pure *s* like, $\Gamma_{2'}$ or Γ_6, in the pure crystals, they have a strongly mixed character in the alloys as has been demonstrated by the dramatic variation of spin orbit splitting and effective mass with composition in several small band gap alloy systems.[31,32] Thus, the top of the valence band remains predominantly antibonding *s* like for $0.15 < x \lesssim 0.45$, even though the band gap is positive. As the sign of $\partial E_{cv}/\partial T$ is determined by the nature of the initial and final states (rather than by the sign of E_{cv}), this consideration correctly describes the observed "anomalous" temperature variation.

From this we see that the temperature dependence of gaps may indicate the bonding character of bands in new materials. For example, transition between nonbonding *d* states will have a small ΔS.[33]

Finally, the softening of the lattice modes by holes and conduction electrons applies also when these result from vacancies or other traps, thus explaining quantitatively some of the large entropies observed for defects[6] and the isotope shift for optical transitions to/from impurity levels.[7] The softening of the modes is also related to the probability of deexcitation of traps by radiationless transitions.[34]

We are indebted to J. J. Hopfield, T. N. Morgan, and C. D. Thurmond for enlightening conversations.

---

*Portions of this work were completed while the authors were at Bell Laboratories, Murray Hill, N. J.

[1] H. Y. Fan, Phys. Rev. 82, 900 (1951); M. L. Cohen, *ibid.* 128, 131 (1962); G. D. Mahan, J. Phys. Chem. Solids 26, 751 (1965).

[2] C. Keffer, T. M. Hayes, and A. Bienenstock, Phys. Rev. Lett. 21, 1676 (1969); R. V. Kasowski and L. M. Falicov, *ibid.* 22, 1001 (1969); J. P. Walter, R. R. L. Zucca, M. L. Cohen, and Y. R. Shen, *ibid.* 24, 102 (1970); and several more recent papers.

[3] Yu. I. Ravich, Fiz. Tverd. Tela 7, 1821 (1965) [Sov. Phys.–Solid State 7, 1466 (1965)].

[4] V. Heine (unpublished). See also M. Schlüter, G. Martinez, and M. L. Cohen, Phys. Rev. B 12, 650 (1975).

[5] H. Brooks, Adv. Electr. 7, 85 (1955), particularly pp. 121–4.

[6] J. A. Van Vechten, Lattice Defects in Semiconductors, International Physics Conference, Series No. 23, 1974 (Institute of Physics, London, 1975), p. 212; J. A. Van Vechten and C. D. Thurmond (unpublished).

[7] V. Heine and C. H. Henry, Phys. Rev. B 11, 3795 (1975) see also J. A. Van Vechten, Phys. Rev. B 13, (1976).

[8] R. Martin, Phys. Rev. 186, 871 (1969).

[9] W. Weber, Phys. Rev. Lett. 33, 371 (1974).

[10] W. Bludau, A. Onton, and W. Heinke, J. Appl. Phys. 45, 1846 (1974).

[11] R. W. G. Wyckoff, *Crystal Structures* (Interscience, New York, 1963), Vol. 1, p. 28.

[12] C. Kittel, *Thermal Physics* (Wiley, New York, 1969), p. 298.

[13] Y. P. Varshni, Physica (Utr.) 34, 149 (1967).

[14] C. D. Thurmond, J. Electrochem. Soc. 122, 1133 (1975).

[15] R. R. L. Zucca and Y. R. Shen, Phys. Rev. B 1, 2668 (1970).

[16] J. C. Phillips, Phys. Rev. 166, 832 (1968).

[17] G. Dolling, *Inelastic Scattering of Neutrons in Solids and Liquids* (International Atomic Energy Agency, Vienna, 1963), Vol. II, p. 37; H. Palevsky, D. J. Hughes, W. Kley, and E. Tunkelo, Phys. Rev. Lett. 2, 258 (1959).

[18] C. V. de Alvarez and M. L. Cohen, Solid State Commun. 14, 317 (1974); M. L. Cohen (private communication).

[19] M. L. Cohen and V. Heine, Solid State Phys. 24, 37 (1970) and especially *ibid.* 24, 134 (1970).

[20] H. Ibach, Phys. Status Solidi 31, 625 (1969); W. Paul, J. Appl. Phys. 32, 2082 (1961).

[21] E. Tosatti, Phys. Rev. Lett. 33, 1092 (1974).

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

[23] D. L. Price, J. M. Rowe, and R. M. Nicklow, Phys. Rev. B 3, 1268 (1971).

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

[25] P. Y. Yu and M. Cardona, Phys. Rev. B 2, 3193 (1970).

[26] C. J. Glassbrenner and G. A. Slack, Phys. Rev. 134, A1058 (1964).

[27] H. D. Barber, Solid State Electr. 10, 1039 (1967).

[28] J. P. Walter (unpublished).

[29] D. Auvergne, J. Camassel, H. Mathieu, and M. Cardona, Phys. Rev. B 9, 5168 (1974).

[30] C. R. Pidgeon and S. H. Groves, in *Proceedings of the International Conference on II-VI Compounds*, edited by D. G. Thomas (Benjamin, New York, 1967), p. 1080; W. M. Scott, J. Appl. Phys. 40, 4077 (1969); E. F. Skelton, P. L. Radoff, P. Bolsaitis, and A. Verbalis, Phys. Rev. B 5, 3008 (1972).

[31] J. A. Van Vechten, O. Berolo, and J. C. Woolley, Phys. Rev. Lett. 29, 1400 (1972).

[32] O. Berolo, J. C. Woolley, and J. A. Van Vechten, Phys. Rev. B 8, 3794 (1973).

[33] D. E. McCumber and M. D. Sturge, J. Appl. Phys. 34, 1682 (1963); W. D. Grobman and D. E. Eastman, Phys. Rev. Lett. 28, 1038 (1972); R. E. Dietz (private communication).

[34] C. H. Henry and D. V. Lang, in *Proceedings of the Twelfth International Conference on the Physics of Semiconductors* (Teubner, Stuttgart, 1974), p. 411.