PHYSICAL REVIEW B VOLUME 14, NUMBER 2 15 JULY 1976

Pseudopotential theory of covalent bonding*

Walter A. Harrison
Applied Physics Department, Stanford University, Stanford, California 94305
(Received 26 February 1976)

The electronic structure of tetrahedral semiconductors is formulated on the basis of a four-orthogonalized-plane-wave (OPW) expansion of the wave functions at the center of each Jones-zone face. The band gap is found to be approximately equal to twice the magnitude of the [111] matrix element of the pseudopotential rather than the [220] matrix element or the Heine-Jones perturbation form. This confirms the dependence of the gap upon even and odd parts of the pseudopotential assumed by Phillips and Van Vechten in formulating their ionicity theory. Fitting of this gap to the optical-absorption peak E_2 gives values in semiquantitative agreement with the corresponding values from the empirical-pseudopotential matrix elements of Cohen and Bergstresser. Neglecting other matrix elements allows a Penn-like calculation of the dielectric constant for polar as well as homopolar semiconductors, but one which in both cases is found to vary as the inverse cube rather than the inverse square of the gap. The form for homopolar semiconductors is only consistent with the Penn formula if, as is approximately true, the gap scales inversely with the square of the bond length; in that case it is also consistent with the bond-orbital formula with the covalent energy V_2 equal to the [111] pseudopotential matrix element. The form for heteropolar semiconductors is inconsistent with the direct extension of the Penn formula, but is consistent with the bond-orbital model.The change in total energy under lattice shear is computed, again retaining only the [111] matrix elements. The elastic constant is found to be proportional to the derivative of the pseudopotential matrix element with respect to wave number. Use of corresponding values from pseudopotential theory gives good estimates of the elastic constants for diamond, silicon, and germanium, and suggests the correct trend for polar semiconductors. These terms which are found adequate for understanding the dielectric constant and lattice rigidity lead to a saddle point rather than a maximum in charge density at the bond center, suggesting that the bond charge should not be thought of as the origin of structural stability of tetrahedral structures. Extending the theory beyond the four-OPW model does lead to charge accumulation in the bonds and retains the structural stability, principally through the kinetic energy of the electrons confined to the Jones zone. The fundamental difference between the pseudopotential theory of covalent solids and of metals is not the inclusion of higher-order terms in the pseudopotential, but the different treatment of zero- and first-order terms analogous to degenerate perturbation theory.

I. INTRODUCTION

The conception of the electronic structure of diamond-structure crystals based upon corrections to free-electron behavior has a long history. A central concept is the Jones zone, given already by Mott and Jones[1] in 1936. It is a zone in wave-number space bounded by Bragg reflection planes and chosen such that: (a) It contains precisely the volume required to accommodate the valence electrons. (b) The structure factor be large so that a large band gap at the face is to be anticipated. (c) It be very nearly spherical so that the entire free-electron Fermi surface may disappear into the zone face, leaving a nonconducting electronic structure. The corresponding zone for diamond, made up of (220) Bragg planes, satisfies these criteria perfectly: it has precisely the correct volume; the structure factor is unity corresponding to all atoms scattering in phase, and as a regular dodecahedron, it lies very close to a sphere. Thus one conceives of a free-electron-like electronic structure dominated by this Jones-zone gap with the influence of other Bragg planes being quite secondary.

An important application of this concept was made by Penn[2] who used it as the basis for an approximate calculation of the optical dielectric constant. He replaced the Jones zone by a sphere of equal volume, took a band gap E_g at the zone face, and made a spherical average of the bands within it; in this way he obtained a dielectric constant,

$$\epsilon_0 = 1 + (\hbar\omega_p/E_g)^2[1 - E_g/4E_F + \tfrac{1}{3}(E_g/4E_F)^2]\,. \tag{1}$$

Here the electron density has been written in terms of the plasma frequency, $\omega_p^2 = 4\pi Ne^2/m$, and E_F is the free-electron Fermi energy. Cardona[3] has subsequently found that a factor $\frac{2}{3}$ should appear in front of the $(\hbar\omega_p/E_g)^2$, but this difference, as well as the correction terms in E_g/E_F which Penn notes were small, will not concern us here. An essential point of the model was obtaining the dielectric constant in terms of the single parameter E_g.

This form of the dielectric constant provided the basis for an ionicity theory of tetrahedral semiconductors by Phillips.[4,5] He assumed that in a polar semiconductor the gap arose from an even and an odd part of the pseudopotential (measured from the bond center) such that the band gap was given by

$$E_g = |E_h + iC| = (E_h^2 + C^2)^{1/2} \tag{2}$$

with E_h arising from the even part of the pseudopotential and iC arising from the odd part, to first order in the pseudopotential.[3] Second, he assumed that in the isoelectronic series of semiconductors (such as Ge, GaAs, ZnSe, and CuBr) only the band gap in Eq. (1) changes. With these two assumptions and the assumption that E_h does not change in the isoelectronic series he could use the measured dielectric constants for the series to deduce E_h and each of the C values for the series. He used these to define an ionicity C^2/E_g^2 for each compound and sought systematic trends in properties as a function of this ionicity.

As plausible as this concept of the electronic structure seems, it runs into serious difficulties upon closer inspection. The matrix elements which determine the band gap depend upon a pseudopotential form factor as well as upon the structure factor. Mott and Jones of course had no way of estimating the former, but once the pseudopotentials were determined for the semiconductors they were found to be very near zero for the (220) Bragg reflection plane.[6] This would at first seem to invalidate the concept from the start. However, a way out was noted by Heine and Jones.[7] They indicated that the states on opposite Jones-zone faces ($k_{110}, k_{\bar{1}\bar{1}0}$) are not only coupled by the pseudopotential matrix W_{220}, but also indirectly through other states. In particular, if one treats coupling with the states k_{001} and $k_{00\bar{1}}$ by perturbation theory, the matrix element producing the gap, which might be written $\frac{1}{2}E_g$, becomes

$$\tfrac{1}{2}E_g = W_{220} + (W_{111}W_{11\bar{1}} + W_{11\bar{1}}W_{111})/K_0\,. \tag{3}$$

Here K_0 is a free-electron kinetic energy $\hbar^2(2\pi/a)^2/2m$, with a the unit cube edge. Heine and Jones suggest that the second-order term may be the dominant one, and the direct term W_{220} unimportant. This is a very appealing suggestion: it implies that the important terms in the covalent solid are just the higher-order terms which are generally neglected in the pseudopotential theory of simple metals. It is of course also just the higher-order terms which produce asymmetries in the charge distribution (deviations from the superposition of spherical densities) which one intuitively associates with bond formation.

There remain, however, serious questions concerning the extension of the concept to polar solids. First, the form taken for the band gap [Eq. (2)] is not directly consistent with pseudopotential theory, Eq. (3). The fact that the structure factor at the (220) face is unity means that the atoms add in phase and only the even part of the pseudopotential enters; there is no contribution to W_{220} from the odd part of the pseudopotential associated with C. If, on the other hand, we follow the suggestion of Heine and Jones and neglect W_{220}, we indeed find that the matrix element products, $W_{111}W_{11\bar{1}}$, for example, are proportional to the sum of the squares of the even and odd parts. That, however, gives a gap of a form $E_g \sim (E_h^2 + C^2)/K_0$, which differs in an important way from Eq. (2). Cardona[3] explored this point with some care and found no way of reconciling the use of Eq. (2).

Recently the first assumption of Phillips has also been drawn into serious question.[8-10] Any of a number of ways of learning how the matrix elements which enter the numerator in the dielectric-constant calculation vary with polarity indicates that they are not at all constant. In fact, they all imply that $\epsilon_0 - 1$ should vary inversely with the cube of E_g in an isoelectronic series rather than with the inverse square.

Thus we are led to doubt the two basic assumptions made in basing an ionicity theory upon a pseudopotential description of the electronic structure. At the same time we find the conceptual view of the electronic structure most appealing. The present study is an attempt at reformulating the description of the electronic structure and the dielectric properties, without making the two suspect assumptions.

II. ELECTRONIC STRUCTURE

We will proceed as with a step-by-step calculation of the electronic structure using the pseudopotential formalism, making approximations in accordance with the insight discussed in Sec. I. At various stages it will be very helpful to substitute values for the pseudopotential matrix elements which enter, and for these we will use empirical pseudopotentials determined by Cohen and Bergstresser[11] by fitting the band structure as revealed by the experimental optical properties. In the end we wish to determine the matrix elements from experiment, as Phillips did in developing his ionicity theory[4,5] and as we did in formulating the linear-combination-of-atomic-orbitals (LCAO) description of covalent bonding,[12] but it will be useful to keep the explicit connection with traditional pseudopotentials. In doing this we should recognize that there is a considerable uncertainty arising ultimately from the familiar arbitrariness among rigorous pseudopotentials.[13] In more familiar terms, one can make up for the neglect of one set of terms by modifying another set. We are much more severely truncating our calculation than did Cohen and Bergstresser so *our matrix elements should differ from theirs*. Nonetheless, it has frequently happened that pseudopotentials determined in different ways are similar and that will turn out to be the case here.

We might imagine expanding the electronic states in orthogonalized plane waves (OPW's), or the pseudo-wave-function in plane waves. A Hamiltonian matrix is constructed and could be diagonalized to obtain the energy bands. However, we focus only on the salient part of the matrix, just as Penn focused on the two coupled states at opposite Jones-zone faces. We consider these two states k_{110} and $k_{\bar{1}\bar{1}0}$ and the two states k_{001} and $k_{00\bar{1}}$, which are coupled to them by the very large matrix elements W_{111}. (For reference the pseudopotential form factors for silicon are[11] $w_{111}=-2.87$ eV, $w_{220}=0.55$ eV, $w_{311}=1.09$ eV, others zero.) We may readily construct the Hamiltonian matrix, based on the states k_{110}, $k_{\bar{1}\bar{1}0}$, k_{001}, and $k_{00\bar{1}}$, numbered 1–4, respectively. In constructing the matrix elements we take our origin of coordinates in one atom. We write the magnitude of the W_{111} matrix element W_1 and its phase factor φ. (For homopolar semiconductors W_1 is the form factor divided by $\sqrt{2}$ and $\varphi=\exp\frac{1}{4}i\pi$.) We write the matrix element W_{220}, which is real, as W_2. (For homopolar semiconductors it equals the form factor.) We drop W_{200} which couples the lower states. (It vanishes for homopolar semiconductors.) The Hamiltonian matrix becomes

$$H=\begin{pmatrix} 2K_0 & W_2 & W_1\varphi^* & W_1\varphi \\ W_2 & 2K_0 & W_1\varphi^* & W_1\varphi \\ W_1\varphi & W_1\varphi & K_0 & 0 \\ W_1\varphi^* & W_1\varphi^* & 0 & K_0 \end{pmatrix}. \qquad (4)$$

This corresponds to a four-OPW calculation of the bands at X in the Brillouin zone. The eigenvectors and eigenvalues are obtained analytically and can readily be verified. The four eigenvalues are $2K_0$, $-W_2$, K_0, and

$$\tfrac{1}{2}\{W_2+3K_0\pm[(W_2+K_0)^2+16W_1^2]^{1/2}\}\,.$$

Note that they do not depend on the phase factor φ. The values corresponding to the top of the valence band and the bottom of the conduction band (see Fig. 1) are, respectively,

$$2K_0-W_2 \quad\text{and}\quad \tfrac{1}{2}\{W_2+3K_0+[(W_2+K_0)^2+16W_1^2]^{1/2}\}$$

corresponding to a gap of

$$E_g=\tfrac{1}{2}\{3W_2-K_0+[(W_2+K_0)^2+16W_1^2]^{1/2}\}. \qquad (5)$$

Values of the empirical pseudopotential and the evaluation of this expression are made in Table I.

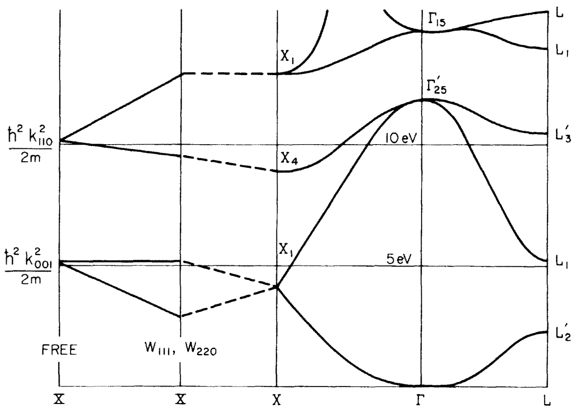

FIG. 1. Electronic structure of silicon. On the left are shown the two doubly degenerate free-electron levels, split by the inclusion of W_{111} and W_{220} in the solution of Eq. (4). To the right are the true energy bands of silicon as given by F. Herman, R. L. Kortum, C. D. Kuglin [Int. J. Quantum Chem. Suppl. 1, 533 (1966)]. Identification is made with dashed lines between the four levels calculated here and the values at X. Note that there are twelve states of k_{110}-type, but only six of k_{100}-type. That is why the 4×4 submatrix only lowers one of the k_{100} states, while the full calculation lowers both.

Note in particular that $16W_1^2$ is generally considerably larger than K_0^2 [or $(K_0+W_2)^2$]. This is in remarkable contrast to the typical situation in simple metals where the relevant kinetic energies are generally much larger than the pseudopotential; that is the basis of the usual pseudopotential perturbation theory used in simple metals. Here the pseudopotential is the large term and the

TABLE I. Pseudopotential parameters and band gaps (eV) for the homopolar semiconductors.

	C	Si	Ge	Sn
W_1 [a]	−7.80	−2.03	−2.23	−2.44
W_2 [b]	4.58	+0.55	0.15	0
K_0	11.88	5.10	4.73	3.59
E_g [Eq. (5)]	18.57	3.22	2.94	3.40
$X_1 - X_4$ [c]	⋯	2.7	2.8	2.2
$2W_1$ [a]	15.60	4.1	4.5	4.9
E_2	12.2 [d]	4.5 [e]	4.4 [e]	3.7 [e]
γ_p	0.52	0.52	0.64	0.86

[a] $2^{-1/2}$ times the (111) pseudopotential form factor, given in Ref. 11.

[b] Reference 11.

[c] F. Herman, R. C. Kortum, C. O. Kuglin, J. P. Van Dyke, and S. Skillman, Methods Comput. Phys. 8, 193 (1968).

[d] R. A. Roberts and W. A. Walker, Phys. Rev. 161, 730 (1967).

[e] R. E. Lindquist and A. W. Ewald, Phys. Rev. 135, A191 (1964).

kinetic energies are the small corrections. This is also consistent with the LCAO picture of the electronic structure.[12] The pseudopotential corresponds to the bonding-antibonding splitting in the solid while the kinetic energy (or bandwidth) corresponds to the *s*-*p* splitting in the atom. (We will make the connection quantitative later.) Thus the fact that the band gap occurs between the bonding and antibonding state, rather than between *s* and *p* bands, also reflects the dominance of the pseudopotential. We note also that the contribution of W_2 is very much smaller than that of W_1, as suggested by Heine and Jones.[7]

The agreement of the calculated E_g with values of the splitting $X_1 - X_4$ from a band calculation is not bad. The discrepancies are hardly surprising for reasons discussed at the beginning of this section, and for a four-OPW calculation; even full-scale band calculations differ from each other by amounts comparable to this. Also listed in Table I is the energy of the peak in the optical reflectivity, E_2, a number frequently associated with Jones-zone gaps. These have similar magnitudes but are systematically larger than the splitting at X. The variations among these numbers suggest that for semiquantitative considerations we might neglect the small corrections to E_g due to W_2 and K_0, in which case E_g becomes simply $2W_1$, and then obtain W_1 by fitting experimental E_2 values.

It is indeed remarkable that $2W_1$ is just the expected splitting at the (111) Bragg plane, the splitting $L_1 - L_2'$, seen in Fig. 1. Indeed the magnitudes of the splittings are similar in that figure and this is true also for diamond, germanium, and tin; however, again the variation from source to source for values of the splitting is large enough to make any detailed comparison questionable.

The same analysis may be applied to the polar semiconductors. This is particularly simple when W_2 and K_0 are neglected. The pseudopotential form factor in this case is replaced by the square root of the average of the squared form factors for the components, since the two atoms in the cell scatter 90° out of phase. This is equal to the square root of the average and difference form factors squared and summed so that it in fact is equivalent to the form [Eq. (2)] assumed by Phillips. The pseudopotential which enters is not that for the Jones zone, but that for a lower gap which from the point of view we described in Sec. I would have been of only secondary importance. There is also a change in the phase factor φ, but that does not affect any of our results.

We may calculate the corresponding pseudopotential matrix elements again using empirical pseudopotentials. The values of W_1 for the series Ge, GaAs, and ZnSe are given in Table II. The average pseudopotential also varies somewhat in this series, so an ionicity theory defined in terms of these empirical pseudopotentials would be quite different depending upon whether one based it upon the average or upon the homopolar value. Again, for our empirical fit we will use $W_1 = \frac{1}{2}E_2$, also given in Table II.

III. DIELECTRIC CONSTANT

For the purposes of this discussion the dielectric constant is written most conveniently in terms of the matrix elements of the gradient[8,9]

$$\epsilon_0 = 1 + 2(\hbar\omega_p)^2\hbar^2\langle 1|\nabla|2\rangle^2/mE_g^3 \qquad (6)$$

neglecting, as we did with the Penn formula, small corrections of order E_g/E_F. The matrix element of the gradient should be taken between two states at the Jones-zone face and averaged over the angle. In order to estimate this matrix element we must construct the states, as well as the energies, from the solution of Eq. (4). Initially we again retain the full matrix. Written as normalized four vectors, the upper valence-band state is simply $(2^{-1/2}, -2^{-1/2}, 0, 0)$ and the lower conduction band state is $(a, a, b\varphi, b\varphi^*)$ with a and b real and a given by $2a^2 = [s^2 + 1 - s(s^2 + 1)^{1/2}]^{-1}$; the expression for b is of the same form with the minus sign before the square root changed to a plus. s is the quantity $(K_0 + W_2)/(4W_1)$ which we have regarded as small. Only a enters the matrix element and for the parameters for silicon in Table I it takes the value 0.88. Thus for this discussion it is reasonable to neglect s or absorb it in a scale factor, consistent with the use of $E_g = 2W_1$. Using plane-wave pseudo-wave-functions in the evaluation of the matrix element of the gradient would lead to i times the component of k_{110} along the direction of the gradient. We may at this point insert a scale factor γ_p in the matrix

TABLE II. Pseudopotential parameters and E_2 for an isoelectronic series, in eV.

	Ge	GaAs	ZnSe
$V_2 = W_1$ (even) [a]	−2.23	−2.40	−1.98
$V_3 = W_1$ (odd) [a]	0	2.55	2.73
$W_1 = \sqrt{V_2^2 + V_3^2}$	2.23	3.50	3.37
$E_2/2$	2.2 [b]	2.5 [b]	3.2 [c]

[a] Reference 11.

[b] R. E. Lindquist and A. W. Ewald, Phys. Rev. 135, A191 (1964).

[c] F. H. Pollack, in *Proceedings of the International Conference on II-VI Semiconducting Compounds, Providence, 1967*, edited by D. G. Thomas (Benjamin, New York, 1968), p. 552.

elements to allow for corrections such as the effect of s. We substitute the resulting matrix element in Eq. (6); the angular average gives a factor of $\frac{1}{3}$, and noting that $\hbar^2 k_{110}^2/2m$ is $2K_0$, we obtain

$$\epsilon_0 = 1 + 8\gamma_p^2(\hbar\,\omega_p)^2 K_0/3E_g^3. \tag{7}$$

In the simplest model γ_p could be taken as being equal to one. Neither the matrix elements nor the gaps depend upon the phase φ, so the result is appropriate for polar as well as homopolar semiconductors. However, substitution of $E_g = E_2$ from Table I gives a value for silicon too high by a factor of 4 for silicon if γ_p is taken as unity.

This situation is not significantly worse than in the LCAO theory of the dielectric constant.[12] There the simplest model for the matrix elements and use of E_2 for the gap gave discrepancies up to a factor of 3 in the dielectric constant so that a scaling factor γ for the matrix elements was introduced in order to bring the dielectric constant into accord with experiment[10,12]; that value (1.06, 1.22, 1.43, and 1.74 for C, Si, Ge, and Sn, respectively) also turned out to depend significantly upon their placement in row in the periodic table, but remained constant in any isoelectronic series. This scaling of matrix elements was also required in Phillips's ionicity theory where empirical corrections associated with d states were required. Corrections are presumably made for a number of the approximations which enter each treatment by the introduction of this scaling parameter and its fitting to experiment. The factor γ_p may be evaluated by taking the E_g equal to E_2 (from Table I) in Eq. (7) and fitting the experimental values of the dielectric constant. The results are given in Table I; they are taken constant in any isoelectronic series.

This completes a description of the assumptions and the parameters of the model. We may now state it explicitly and compare it with earlier treatments.

IV. FOUR-ORTHOGONALIZED-PLANE-WAVE (OPW) MODEL

The electronic states are based upon free electrons with a gap introduced at the Jones zone. The gap arises from the lifting of the conduction-band edge [Eq. (5) with W_2 and K_0 neglected] rather than the more common splitting up and down, a point which will be important when we consider the total energy. The magnitude of the gap is twice the [111] matrix element of the pseudopotential. In a polar system, with pseudopotential form factors for the two constituents of w_1 and w_2 we may write the pseudopotential

$$W_1\varphi = V_2 + iV_3\,, \tag{8}$$

with

$$V_2 = (w_1 + w_2)/(2\sqrt{2})\,,$$
$$V_3 = (w_1 - w_2)/(2\sqrt{2})\,, \tag{9}$$

and the gap is given by

$$E_g = 2(V_2^2 + V_3^2)^{1/2}. \tag{10}$$

[Strictly speaking the phase of each term in Eq. (9) should be rotated 45° for our choice of origin but it doesn't affect any results.] The notation was of course used to fit an identification with LCAO theory which we make later.

In the model, the gap in the homopolar semiconductors is taken to be equal to the optical absorption peak E_2, giving values of V_2 for each row in the periodic table, since V_3 vanishes for the homopolar semiconductors. Following Phillips, the odd pseudopotential is obtained from the dielectric constant assuming that the even part is unchanged. However, the appropriate formula is taken to be Eq. (7) rather than Eq. (1). In any isoelectronic series the scale factors γ_p cancel out and explicit experimental values of the even and odd pseudopotentials are obtained which are in rough agreement with the empirical pseudopotential matrix elements of Cohen and Bergstresser (compare E_2 and $2W_1$ in Tables I and II). Values of γ_p are obtained for each row from the experimental dielectric constant and geometric means of both γ_p and V_2 are used in skew compounds.

A. Comparison with ionicity theory

Phillips[4,5] used the dielectric constant and considerations based on a sum rule to obtain gaps E_g, so our values for the homopolar semiconductors inevitably differ from his; however, not so greatly. His values of E_g for C, Si, Ge, and Sn are 13.5, 4.77, 4.31, and 3.06 eV, respectively, to be compared with E_2 in Table I. However, his principal interest was in the isoelectronic series and there the difference is important. He defined his ionicity as $f_i = C^2/E_g^2$ and obtained it upon the assumption that ϵ_0^{-1} varied as E_g^{-2}. If we instead use Eq. (8), we obtain values f_i' related to his by

$$f_i' = 1 - (1 - f_i)^{2/3}. \tag{11}$$

f_i' is roughly linear in f_i over the range of values which occur in the tetrahedral structure, as was noted earlier,[14] but about $\frac{2}{3}$ as large.

We could of course use the new values of ionicity to interpolate properties but it is much more fruitful, since there is now a clear relation with the underlying electronic structure, to use that model electronic structure to predict the properties themselves.

B. Comparison with the bond-orbital model

In spite of the very much different conceptual basis, the four-OPW model leads to results very close to those of the LCAO theory. Part of this similarity is simply by choice; we chose similar notation and used the same experiments to define parameters. However, part of the similarity is fundamental and, we believe, important.

The dielectric constant in the bond-orbital model is given by

$$\epsilon_0 = 1 + \frac{\pi\gamma^2 N e^2 d^2 V_2^2}{3(V_2^2 + V_3^2)^{3/2}} = 1 + \frac{\pi^2\gamma^2(\hbar\,\omega_p)^2 V_2^2}{4K_0 E_g^3}\,. \qquad (12)$$

The last form has been written as close as possible to the pseudopotential form [Eq. (7)] but remains fundamentally different. However, V_2 was taken as half the E_2 energy (as we have here) and Harrison and Ciraci[12] found empirically that E_2 varies approximately as d^{-2} from material to material. This is only an approximate relation, as may be seen from their curve or from a comparison of E_2 and K_0 (which varies exactly as d^{-2}) in Table I here. In fact, the average E_2 equals the average K_0 within a few percent; with that approximate equality Eqs. (12) and (7) become equivalent with $\gamma_p \sim 0.5\gamma$. As we noted earlier[10] it is only the approximate proportionality of E_2 and K_0 which makes the nearly-free-electron and LCAO models simultaneously tenable.

As we have chosen to obtain parameters from experiment the two theories lead to identical V_2's, but we now have alternative interpretations of V_2 as the matrix element between two hybrids in a bond or the [111] pseudopotential matrix element. Similarly since the V_3 are obtained from the dielectric constants in an isoelectronic series, and γ, γ_p, V_2, $\hbar\,\omega_p$, and K_0 remain constant in such a series (for this we do not need the proportionality of V_2 and K_0), the V_3 values are the same.

The identification may be made more plausible by imagining the pseudopotential to be a sum of δ functions on the two atoms, of the form $(-V_2 \pm V_3)\sqrt{2}\,\Omega_0\delta(\vec{r} - \vec{r}_i)$, with Ω_0 the atomic volume. (The root two is artificial, as is the use of a δ function.) This corresponds to form factors,

$$\Omega_0^{-1} \int w(\vec{r} - \vec{r}_i)e^{-i\vec{q}\cdot(\vec{r}-\vec{r}_i)}d\tau\,,$$

of $\sqrt{2}(-V_2 \pm V_3)$, consistent with Eq. (10). As V_2 may be thought of as the energy gained in putting an electron on the site, $2V_3$ is the energy difference between two sites, which was its definition in the bond-orbital model.

Of course, since V_2 and V_3 are numerically the same in both models, α_p, defined by $V_3/(V_2^2 + V_3^2)^{1/2}$, is the same in both. We do not wish to emphasize this aspect since the theory can go far beyond simple scaling of properties.

V. TOTAL ENERGY

We next take a preliminary look at the theory of bonding properties in the four-OPW model. We consider first the total energy. An important contribution, and the one central to the discussion here, is the sum of one-electron energies. This term is in fact thought to dominate the angular rigidity of covalent lattices.[15] This is relevant to the stability of the tetrahedral structure as well as the elastic shear constants. We will return to the other contributions in Sec. VIII.

The estimate of the change in total energy in terms of the four-OPW calculation is not completely straightforward. In particular, some care must be taken if one is to get the factors of 2 correct. We return to the 4 × 4 matrix of Eq. (4), based on the states k_{110}, $k_{\bar{1}\bar{1}0}$, k_{001}, and $k_{00\bar{1}}$. In the end we neglected W_2 and K_0, so we may drop them here. We may then diagonalize the off-diagonal 2 × 2 blocks, leaving only two nonzero matrix elements, for example, $H_{13} = 2W_1\varphi^*$, $H_{31} = 2W_1\varphi$. This raises the unoccupied conduction-band state by $2W_1$ and lowers one of the lower valence states by the same amount (see Fig. 1, though there W_2 and K_0 were retained). Similarly the treatment of the upper states $k_{1\bar{1}0}$ and $k_{\bar{1}10}$ will lower the other k_{001} state, so they are degenerate (again, see the bands in Fig. 1). Thus two of the four valence bands are lowered by $2W_1$ each at the Jones-zone face. The dominance of the matrix element over the energy denominator and the high density of Jones-zone planes (when translated into the reduced Brillouin zone) both suggest that the X point be considered representative, corresponding to a lowering in energy of W_1 per electron. It is interesting and satisfying that this gain in energy, V_2 per electron for homopolar semiconductors, is the same as in LCAO theory. It is also remarkable that it comes from the lowering of deep states, not states at the Jones-zone faces.

Now let us see how this energy changes under a shear distortion. To be specific, we imagine expanding the crystal along the z axis by a factor $(1+\epsilon)$, and contracting it in the two lateral dimensions by $(1+\epsilon)^{-1/2}$. Then the wave-number lattice is contracted in the z direction by a factor $(1+\epsilon)^{-1}$ and expanded in the lateral directions by $(1+\epsilon)^{1/2}$. This increases the length of each [111] wave number by a factor of $(1+\frac{1}{2}\epsilon^2)$ to order ϵ^2, and we may estimate the resulting change in W_1 as

$$\frac{1}{2}\frac{k\,\partial W_1}{\partial k}\epsilon^2=\frac{k\,\partial w_1}{\partial k}\frac{\epsilon^2}{2^{3/2}}$$

for a homopolar solid. A new parameter has entered, $\partial w_1/\partial k$, which may be estimated from known pseudopotentials. Those of Animalu and Heine[16] give $k\partial w_1/\partial k = 0.56E_F$ at the appropriate $q/k_F = 1.10$ for silicon, germanium, and tin within a percent. (The universality of w_1/E_F as a function of q/k_F which this suggests is exactly the counterpart of scaling of V_2 with d^{-2} noted earlier.) This yields an energy per electron of $0.20E_F\epsilon^2$. We may relate this to the elastic constants by multiplying by the electron density and equating it to the energy density in terms of the elastic constants for the same distortion, $\frac{3}{4}(c_{11}-c_{12})\epsilon^2$. This gives estimates immediately of $c_{11}-c_{12}$ for diamond, silicon, germanium, and tin of 8.7, 1.1, 0.87, and 0.44, respectively (in units of 10^{12} ergs/cm^3). These are in remarkable agreement with the experimental values for the first three of 9.5, 1.0, 0.80 (Sn has not been measured). This predicts (as of course the LCAO theory does) a variation from material to material with d^{-5}, which is well satisfied by experiment. In this pseudopotential theory it is not necessary to introduce a scaling parameter as it was in the LCAO theory.

In order to treat polar semiconductors we must make some assumption as to how the odd part of the pseudopotential, V_3, varies with distortion. The assumption that it is independent of shear which was used in the LCAO theory gives a polar value different from the homopolar value by a factor $V_2/(V_2^2+V_3^2)^{1/2}=\alpha_c$, or values for the isoelectronic series, Ge, GaAs, and ZnSe of 0.87, 0.75, and 0.60, compared to the experimental 0.80, 0.65, and 0.32. The trend is right though it is not quantitatively as accurate as the LCAO theory which gave[15] a proportionality to $\alpha_c - 0.5$.

There is always a temptation to stop when the prediction agrees this well with experiment, but in the pseudopotential theory it seems particularly clear that there are many other terms in the energy which have not been considered, and it is not clear that they cancel each other. We will return to a more complete description in Sec. VII.

VI. CHARGE DENSITY

Next we may consider the charge density in terms of the four-OPW model. The result was quite unanticipated but informative. A good way to carry out the calculation is to note that the sum of the charge densities from the four states entering Eq. (4) is uniform. When the lower three are occupied and the upper left empty the charge density becomes nonuniform and the ratio may be taken as a correction factor for the total density due to the corresponding Jones-zone face. We may then multiply by the factors for each plane. A simpler, and less accurate, way is to take the charge density to be a superposition of that due to occupied states at each X point in the Brillouin zone. We have done it both ways and the results are qualitatively the same.

First, the omitted wave functions are proportional to $\cos\vec{k}_{110}\cdot\vec{r}$ (neglecting the effects of W_2 and K_0), with a charge density proportional to $1+\cos\vec{k}_{220}\cdot\vec{r}$. Thus the wave-number lattice for the expansion has [220] primitive lattice vectors corresponding to the lattice in real space as body-centered cubic. Each atom sits on a bcc site, but the other half of the sites are empty. It is other Fourier components which we have not included (principally [111] components) which must pull the charge density onto the atomic sites.

Second, since it is a cosine-like wave which is eliminated, the affect is to remove charge from the atomic sites and put it in the interstices. (This is in turn because the interaction through levels of *lower* energy is repulsive.) The net effect is a depressed charge density near the atoms, a pileup in the octahedral interstices (of the bcc lattice), and a saddle-point at the bond center. The charge density in this model is quite metal-like and does not show the familiar concentration of charge in the bonds.

This, of course, does not mean that the charge is *not* piled up in the bonds, but it indicates that the terms which cause that concentration are quite distinct from the terms which cause the structural stability. The distinction is best made as we turn finally to a more complete description of the electronic structure in terms of pseudopotential theory in Sec. VII.

VII. BEYOND THE FOUR-ORTHOGONALIZED-PLANE-WAVE MODEL

We have succeeded in formulating a theory of the dielectric properties and the structural stability of tetrahedral solids through a model which focuses on a single band gap arising from a single pseudopotential matrix element. The success came from taking the pseudopotential as the dominant term rather than the more usual assumption that the kinetic energy is dominant. This did not correspond, as might have been anticipated, to taking the pseudopotential to higher order, but to a first-order shift W_1 in the energy per electron rather than the second-order shift which arises in the perturbation expansion in the theory of metals. The difference is more analogous to the use of degenerate perturbation theory in contrast to nondegenerate theory.

In basing the theory on the single gap, many terms have been neglected, as we noted in the discussion of the angular rigidity and of the charge distribution. This neglect has been justified on the basis of the success of the model, but it is useful also to consider the neglected terms; this will help also bring the model into perspective. A most convenient way of doing this is to start with the pseudopotential theory of metals. We earlier, in fact, made a quantitative study of silicon treating it as a simple metal.[17] That calculation included all Fourier components of the pseudopotential and all electrostatic and self-energy terms, but in the context of the perturbation expansion to second order in the pseudopotential. The present study indicates that we should correct this for a more careful treatment of the [111] Fourier component of the pseudopotential. The other terms in the simple metal theory are just those neglected in the four-OPW model.

In the simple-metal theory, the zero-order state of the system is a free-electron gas, with all wave numbers occupied within the Fermi sphere, of radius k_F. There is a shift $\langle k|w|k\rangle$ of each level of first order in the pseudopotential and second-order terms and appropriate self-energy and electrostatic terms.[6] We now seek the consequences of a more careful treatment of the Jones-zone gap.

The zero-order kinetic energy should of course be obtained as a summation over the Jones zone rather than a sphere, but the spirit of the Jones-zone approach is consistent with the replacement of the zone by a sphere. The difference comes only when we shear the lattice. Then the distortion given in Sec. V deforms the zone and the corresponding sphere becomes an ellipsoid. The corresponding change in the zero-order energy is absent in the metal where the Bragg planes move through the Fermi sphere like a sieve, but is very important in the covalent solid. The average kinetic energy $\langle \hbar^2 k^2/2m\rangle$ in the ellipsoid can be calculated exactly and is given by $(\hbar^2 k_F{}^2/10m)[2 + 2\epsilon + (1+\epsilon)^{-2}]$, and the change in energy is obtained immediately as $3\hbar^2 k_F{}^2\epsilon^2/10m = 1.47K_0\epsilon^2$. Note that the value is independent of the magnitude of the pseudopotential matrix element as long as it is sufficiently large to contain the electrons within the zone. This term contributes 3.2×10^{12} ergs/cm^3 to $c_{11} - c_{12}$, a value three times the experimental value.

The first-order term may also change because of the repopulation of states, but that change vanishes if a local pseudopotential is used, and would be only a small correction if the energy dependence of the pseudopotential were included.[18] Some estimate of the four-OPW energy, W_1, is included in the second-order simple-metal calculation, though it is proportional to W_1^2/K_0 rather than W_1; thus we cannot add that contribution directly. It is a rather small part of the [111] contributions to the energy in any case. In the elastic constant calculation for the four-OPW model the shift in energy per electron was $\frac{1}{2}(k\partial W_1/\partial k)\epsilon^2 = 0.20E_F\epsilon^2$. The contribution of the [111] matrix elements in the simple-metal theory is $8(\frac{1}{2}q\epsilon^2)S^*S\,\partial F/\partial q$ per atom, where q is the [111] wave number and F is tabulated in Ref. 6; this contribution is $0.78E_F\epsilon^2$ per electron, four times the four-OPW value. Thus the total pseudopotential contribution is approximately given by the simple-metal calculation.

Unfortunately, the earlier simple-metal treatment[17] did not include a calculation of the shear constants, but a subsequent calculation by Martin[19] did and gave negative values corresponding to an unstable lattice. Although he used a different pseudopotential (a local model potential), we may assume that ours also would have led to large negative contributions which combine with the zero-order kinetic energy term (not included by Martin) to give the stable structure. The calculation has not been carried out, but comparison of the kinetic term and the experimental value indicates that a cancellation of $\frac{2}{3}$ of the kinetic term by the pseudopotential terms is required. This negative value arises from contributions to the band-structure energy from wave numbers greater than the [111] wave number; for many of these $q\,dF/dq$ is negative.

It is remarkable that in this context the stability of the tetrahedral structure arises from the kinetic energy which, in some sense, blows the Jones zone up like a balloon. The net effect of the changes in pseudopotential under distortion is to reduce the corresponding rigidity. This description of the entirety of contributions makes the rigidity understandable, in spite of the instability of the metallic model, and suggests that the stability arises from terms of lower order than second, not from the effects of higher-order terms. The description does not explain why the cancellation is so complete; that is better understood from the identification of terms with the LCAO theory where the reduction in bond energy under distortion is the plausible source of stability.

We may also consider the charge distribution in terms of the more complete description of the electronic structure. In zero order each state is a plane wave (or OPW), and occupying states in the Jones zone rather than in a spherical Fermi surface would leave a uniform distribution (or a uniform distribution with slightly nonspherical orthogonalization holes for OPW's). In the metallic approximation we did carefully calculate the charge density to first order in the pseudopotential,[17] giving the result shown in Fig. 2, which looks very much like the observed distribution.

It is remarkable that this calculation gives the maximum in charge density at the bond center which one intuitively associates with bond charges. This is in spite of the fact that the charge density obtained in this metallic approximation can rigorously be written as a sum of spherically symmetric densities. This calculation also includes [220] components such as were obtained in the four-OPW calculation. Though they are of slightly different magnitude, they are consistent also with a superposition of spherical densities. We may also note that a very large number of Fourier components were required (500 were used) to obtain a density which did show the charge concentration. This further reduces the connection between the terms which dominate the charge redistribution and those which dominate the structural stability.

We are led to make a sharp distinction between the bond charge which one associates with a picture such as Fig. 2 and the deviations from spherical symmetry which give rise to the observation of forbidden diffraction spots. These can arise because of the filling of a Jones zone rather than a sphere, combined with first-order pseudopotential terms or orthogonalization effects, but it may be that they arise primarily from higher-order terms which are not included in any of the theory given here.

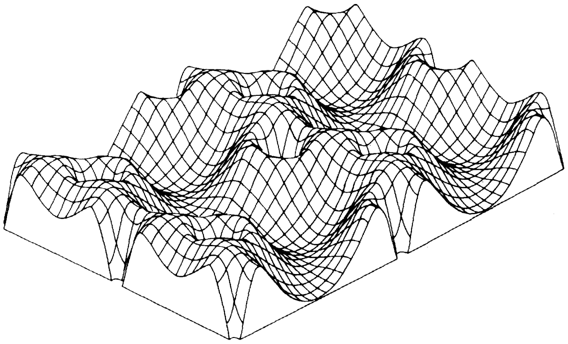

FIG. 2. Relative electron density on a (110) plane in silicon, from Ref. 17, containing only terms of first-order in the pseudopotential.

VIII. CONCLUSIONS

The four-OPW model which is developed here follows closely the Jones-zone concept, the historical development of which was given in Sec. I. There are several features, however, which are new and significant. First, because the [111] matrix element of the pseudopotential is taken to dominate not only the [220] component, but also the kinetic energy, the Jones-zone gap takes the Phillips form $2|W_{111}| = 2(V_2^2 + V_3^2)^{1/2}$ with V_2 and V_3 the even and odd part of the pseudopotential. Substitution of known pseudopotentials then gives predictions of the gap in reasonably good accord with the observed peak in the optical reflectivity. Stated this way, this finding supports the severe truncation of the Hamiltonian matrix which we then extended to other properties. Our procedure was in fact to use the observed peak position to obtain the pseudopotentials which were used in the subsequent studies.

The dielectric constant arises, as in the Penn model, from coupling between states above and below the Jones-zone gap. Because we estimated the matrix elements of the gradient directly rather than by the interpolation scheme Penn used for the homopolar solids, our form differs from his but has the same trends if and only if the band gap scales with d^{-2}, a plausible and well-documented dependence. Our result is also applicable to polar semiconductors, which Penn's was not; in this case as well as in the homopolar case, the results are consistent with those of the bond-orbital model; again only with the d^{-2} scaling. This formula also provides a direct identification of the [111] pseudopotential matrix element as also the matrix element between the two hybrid atomic orbitals forming the bond.

In the calculation of the total energy, the matrix elements which opened the gap depress the energy of states near the middle of the valence band (not those at the top of the valence band) giving a lowering in energy equal to $|W_{111}|$ per electron, exactly consistent again with the identification with the bond-orbital model made above. The decrease in that matrix element is easily estimated for a homopolar semiconductor under shear and leads to good quantitative predictions of the shear constant $c_{11} - c_{12}$ for diamond, silicon, and germanium; no scaling parameter needs to be introduced. The assumption that the odd part of the pseudopotential does not change under distortion (as assumed in LCAO theory) leads correctly to a decrease in the constant with polarity, though it underestimates the decrease. Thus structural stability is predicted by the model and, interestingly enough, without the high-wave-number terms which are responsible for the concentration of charge density in the bonds.

We have addressed here only one aspect of the structural stability question, the stability under shear distortion. There are other distortions which may have equal relevance, an important one being the motion of the second atom in each cell with respect to the first. A convenient procedure for treating this has been described by Schiferl.[20] One can follow the change in pseudopotential matrix elements as one continuously moves from a zinc-blende to a sodium-chloride structure. Note that the kinetic-energy term does not change in the process. The present approach would suggest focusing in particular on the [111] matrix element but this term favors the zinc-blende structure over the sodium-chloride structure at all polarities. It in fact favors internal displacements in which the two atoms in the cell coalesce. Clearly a complete discussion of the stability cannot be made without the inclusion of radial interactions and the other contributions discussed in Sec. VII.

The identification of the LCAO and the pseudopotential approaches could be carried much further. LCAO fits have been made to pseudopotential bands and interatomic $s-s, s-p$, etc., matrix elements obtained.[21] It also must be possible to identify these matrix elements with pseudopotential counterparts.

This formulation has provided a correspondence between the free-electron and LCAO bases for an understanding of the bonding, even to the point of relating the fundamental parameters of the two theories and seeing why both approaches can give reasonable estimates. It seems possible that the LCAO basis may remain the most versatile and simple for treating bonding properties, but it is valuable to have this complementary point of view. We noted, for example, that the pseudopotential theory directly gives a prediction of the decrease in the matrix element V_2 upon distortion of the lattice, while this required a scaling parameter in the LCAO theory. On the other hand, the LCAO theory leads directly to the suggestion that V_3 should be independent of shear (there should be no term linear in ϵ), while that is not obvious from the pseudopotential theory. Such complementary insights greatly enrich the understanding of covalent bonding.

---

*Supported by the National Science Foundation through Grant No. DMR 73-02351.

[1] N. F. Mott and H. Jones, *The Theory of the Properties of Metals and Alloys* (Dover, New York, 1958).

[2] D. R. Penn, Phys. Rev. 128, 2093 (1962).

[3] M. Cardona, Proc. Int. Sch. Phys. "Enrico Fermi", 531, 1971 (unpublished).

[4] J. C. Phillips, Rev. Mod. Phys. 42, 317 (1970).

[5] J. C. Phillips, *Bonds and Bands in Semiconductors* (Academic, New York, 1973).

[6] For an early discussion of this difficulty see W. A. Harrison , *Pseudopotentials in the Theory of Metals* (Benjamin, New York, 1965), p. 210.

[7] V. Heine and R. O. Jones, J. Phys. C 2, 719 (1969).

[8] J. C. Phillips, Phys. Rev. Lett. 34, 1196 (1975).

[9] W. A. Harrison, Phys. Rev. Lett. 34, 1198 (1975).

[10] W. A. Harrison and S. T. Pantelides, Phys. Rev. B (to be published).

[11] M. L. Cohen and T. K. Bergstresser, Phys. Rev. 141, 789 (1966). Tabulated also by M. L. Cohen and V. Heine, in *Solid State Physics*, edited by H. Ehrenreich, F. Seitz, and D. Turnbull (Academic, New York, 1970), Vol. 24, p. 37.

[12] See, for example, W. A. Harrison and S. Ciraci, Phys. Rev. 10, 1516 (1974).

[13] See Ref. 6, p. 11ff.

[14] W. A. Harrison, Phys. Rev. B 8, 4487 (1973), Fig. 4. In this treatment $f_i' = \alpha_p^2$.

[15] See, for example, the LCAO theory of W. A. Harrison and J. Ch. Phillips [Phys. Rev. Lett. 33, 410 (1974)].

[16] A. O. E. Animalu and V. Heine, tabulated in Ref. 6, p. 309.

[17] W. A. Harrison, Physica (Utr.) 31, 1692 (1965), and Ref. 6.

[18] The energy dependence of the pseudopotential $\partial W/\partial E$ is approximately equal to the orthogonalization hole, less than 10% for Si. See R. W. Shaw and W. A. Harrison [Phys. Rev. 163, 604 (1967), Eq. (6.10)].

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

[20] D. Schiferl, Phys. Rev. B 10, 3316 (1974).

[21] Two recent such treatments are by D. J. Chadi and M. L. Cohen [Phys. Status Solidi B 68, 405 (1975)] and by S. T. Pantelides and W. A. Harrison [Phys. Rev. B 11, 3006 (1975)].