PHYSICAL REVIEW VOLUME 166, NUMBER 3 15 FEBRUARY 1968

Covalent Bond in Crystals. I. Elements of a Structural Theory

J. C. PHILLIPS
Department of Physics and James Franck Institute, The University of Chicago, Chicago, Illinois

(Received 3 August 1967)

An *a posteriori* theory is developed for the structural energy of covalent crystals. The microscopic theory is based on ionic pseudopotentials and valence dielectric screening. The theory explains the difference between empirical pseudopotential form factors derived from the optical spectra of semiconductors and the metallic form factors calculated from free-ion term values by Animalu and Heine. A byproduct of the theory, which utilizes Penn's model isotropic semiconductor dielectric function, is a relation between the covalent bonding charge and the macroscopic dielectric constant. In self-consistent form the theory is an example of the "bootstrap" approach, applied here to treat the effect of covalent bonding on the ground-state energy of the valence electron gas. It is argued that the axiomatic character of the covalent theory is to be expected on symmetry grounds, and it is shown that the theory is superior to a nonlinear multiple-scattering theory based on the free-electron dielectric function. The extension of the theory to III–V and II–VI semiconductors is described briefly. The theory may be used to calculate elastic and macroscopic dielectric properties of covalent crystals starting only from ionic pseudopotential form factors.

I. INTRODUCTION

ONE of the most basic and widely used concepts in the quantum theory of matter is the covalent bond. A number of empirical theories have made use of the concept of the covalent bond, but only a few attempts have been made to construct a general microscopic theory.

Because of its periodicity, the covalent bond can be studied with greatest precision in crystals, where it can be immediately subjected to Fourier analysis. In our view, alternative approaches which stress the real-space aspect are restricted at the outset by being subject either to rigid basis states (atomic-orbital method)[1] or to a rigid length of the order of the atomic radius (Thomas-Fermi method).[2] The results obtained from these calculations show poor convergence, and to the extent that the calculations are tractable, they underestimate the degree of localization and the amount of charge contained in the covalent bond.

The most obvious basis states for describing the covalent bond in crystals are thus plane waves. These must be orthogonalized to the atomic core states [Herring's orthogonalized plane-wave (OPW) method][3]; then for a given **k** in the Brillouin zone accurate eigenfunctions require about 60 OPW's. The valence charge density and valence energy can then be obtained in principle by sampling enough different **k**'s in the Brillouin zone.

This method was used[4] with a rather coarse sampling to obtain the Fourier components for diamond for ρ_K with $K=\langle 111\rangle$, $\langle 220\rangle$, $\langle 311\rangle$, $\langle 400\rangle$, and $\langle 222\rangle$ (in units of $2\pi/a$). The results were in good agreement with experiment,[5] but (apart from demonstrating the suitability of plane-wave basis states) so cumbersome an approach shed little light on the qualitative character of the covalent bond.

More satisfactory is the multiple-scattering treatment of Bennemann.[6] He starts from unperturbed plane-wave states, and introduces ionic pseudopotentials. The effect of these on the crystal eigenfunctions is then analyzed using t matrices, modified to include scattering from clusters of one, two, and three atoms. The latter terms include the effects of bonding and bond-bond interactions, respectively. Good results are obtained for the crystal charge density and cohesive energy.

In our view Bennemann's work has not developed further chiefly because too much stress is laid in the exposition[6] on the formal complexities associated with dynamical, nonlocal contributions to the self-energies of the plane-wave (metallic) basis states. (The actual calculations, however, treat these terms in a simple way.)[6] On the other hand, when an energy gap opens in the spectrum it appears that perturbation theory can be considered as only an asymptotic series. Bennemann's calculations suggest that the asymptotic series may give good results, but if we are interested in energy differences, e.g., chemical shifts within a given crystal structure, or relative stability of different structures, or lattice vibration spectra, then the effects of the gap in the spectrum should also be included in the theory.

The aim of the formalism presented here is therefore twofold: (1) To formulate the scattering theory in a simple manner, stressing its structural aspects, and including many-body effects in as simple a way as possible; and (2) to include from the outset the effects of the energy gap in the spectrum, thus refining Bennemann's free-electron perturbation theory by using a covalent spectrum and covalent basis states.

II. ISOTROPIC COVALENT ATOMS

The usual application of t matrices involves plane waves incident upon spherically symmetric scattering

---

[1] P. P. Ewald and H. Hönl, Ann. Physik **25**, 281; **26**, 673 (1936).

[2] H. C. Bolton and J. W. Heaton, Proc. Phys. Soc. (London) **78**, 239 (1961).

[3] C. Herring, Phys. Rev. **67**, 1169 (1940).

[4] L. Kleinman and J. C. Phillips, Phys. Rev. **125**, 819 (1962); I. Goroff and L. Kleinman (unpublished).

[5] S. Göttlicher and E. Wölfel, Z. Elektrochem. **63**, 891 (1959).

[6] K. H. Bennemann, Phys. Rev. **133**, A1045 (1964); **139**, A482 (1965).

centers. The exact wave function $\Psi_{\mathbf{k}}$ is then given in terms of the incident plane wave $\psi_{\mathbf{k}}$ by

$$\Psi_{\mathbf{k}}=\psi_{\mathbf{k}}+\sum_{\mathbf{k}'} g(\mathbf{k},\mathbf{k}')t(\mathbf{k},\mathbf{k}')\psi_{\mathbf{k}'}\,, \tag{2.1}$$

where $t(\mathbf{k},\mathbf{k}+\mathbf{q})$ is obtained as the solution of the integral equation

$$t(\mathbf{k},\mathbf{k}+\mathbf{q})=v(\mathbf{k},\mathbf{k}+\mathbf{q})+\int v(\mathbf{k},\mathbf{k}+\mathbf{q}') \times g(\mathbf{k}+\mathbf{q}',\mathbf{k}+\mathbf{q})t(\mathbf{k}+\mathbf{q}',\mathbf{k}+\mathbf{q})d\mathbf{q}'\,. \tag{2.2}$$

Here g is the electron propagator

$$g(\mathbf{k},\mathbf{k}')=1/[E(\mathbf{k}')-E(\mathbf{k})] \tag{2.3}$$

and the scattering potential v has been assumed to be nonlocal.

In order for (2.1) to be useful, the potential V must be sufficiently weak as not to form bound states below the valence levels. This is accomplished by transformation to the pseudopotential representation.[7] For many elements (e.g., Si and Ge) containing s and p core levels, but for which d valence states are not important, the pseudopotential can be treated as a local function.[8] To facilitate computation and to simplify notation we shall henceforth assume this to be the case. Even for crystals such as diamond, where the absence of p core states means that the pseudopotential is quite different for $l=0$ and $l=1$ states, it is possible[9] to find a local pseudopotential which gives a good description of the crystalline energy levels near the energy gap. These are the levels most important to covalent bonding, so that it is likely that the local pseudopotential approximation can be made to yield semiquantitative results even for cases such as diamond.

Before proceeding to study covalent bonds and bond-bond interactions, we wish to decide on the optimal form for the spherically symmetric atomic form factor $v(q)$. The t-matrix theory is analogous in some respects to Brillouin-Wigner perturbation theory, in the sense that an improvement in the asymptotic expansion of covalency effects can be obtained by calculating each term self-consistently. We propose to do this both with regard to the energy denominators (2.3) and with regard to basis functions. Penn has shown[10] how to calculate the dielectric function $\epsilon_s(q)$ for a model isotropic semiconductor. An average energy gap E_g is introduced into the theory through the relation

$$\epsilon_s(0)=1+(\hbar\omega_p/E_g)^2C(\alpha)\,, \tag{2.4a}$$

$$C=1-\alpha+\tfrac{1}{3}\alpha^2\,, \tag{2.4b}$$

where $\alpha=E_g/4E_F$ and $\omega_p{}^2=4\pi Ne^2/m$ is the plasma frequency. Together with the valence electron density N either $\epsilon_s(0)$ or E_g determine completely the properties of the model isotropic semiconductor.

Because of the presence of the energy gap the model basis functions are no longer plane waves $|\mathbf{k}\rangle$ but instead are the covalent bonding (or antibonding) states

$$|\mathbf{k}\pm\rangle=\alpha|\mathbf{k}\rangle\pm\beta|\mathbf{k}'\rangle\,, \tag{2.5a}$$

where

$$|\mathbf{k}'\rangle=|\mathbf{k}(1-2|\mathbf{k}_F|/|\mathbf{k}|)\rangle \tag{2.5b}$$

is a plane wave with vector parallel to $\mathbf{k}$. The magnitude k_F corresponds to the Fermi wave number of a free-electron gas of density N. The coefficients $\alpha(k)$ and $\beta(k)$ are obtained by solving the model secular equation

$$\begin{vmatrix} \epsilon(k)-E & -\tfrac{1}{2}E_g \\ -\tfrac{1}{2}E_g & \epsilon(|\mathbf{k}-2\mathbf{k}_F|)-E \end{vmatrix}=0 \tag{2.6}$$

for the eigenvectors, with $\epsilon(k)=\hbar^2k^2/2m$. The covalent energies $E(k\pm)$ are the eigenvalues of (2.6). Explicit formulas are given by Penn.[10]

In metals very satisfactory atomic form factors have been obtained by Heine and co-workers[11–13] in the following way. The atomic core is replaced by an l-dependent pseudopotential determined[11] from spectroscopic levels of the hydrogenic ion plus one electron. The pseudopotential form factor of the bare ion, $v_i(q)$, is then derived by taking matrix elements of the l-dependent ionic pseudopotential $\sum_l v_i{}^l$ between plane-wave states $|\mathbf{k}\rangle$ and $|\mathbf{k}+\mathbf{q}\rangle$ normalized to an atomic volume

$$v_i(q)=\textstyle\sum_l\langle\mathbf{k}|v_i{}^l|\mathbf{k}+\mathbf{q}\rangle\,. \tag{2.7}$$

The wave vectors $\mathbf{k}$ and $\mathbf{k}+\mathbf{q}$ are constrained to lie on the energy shell

$$\epsilon(k)=\epsilon(|\mathbf{k}+\mathbf{q}|)=\hbar^2k_F{}^2/2m\,. \tag{2.8}$$

Now the free ion pseudopotential $v_i(q)$ is screened by the Hartree dielectric function $\epsilon_f(q)$ of a free-electron gas to produce the metallic pseudopotential form factor[14]

$$v_m(q)=v_i(q)/\epsilon_f(q)\,. \tag{2.9}$$

In practice the prescription (2.9) is modified slightly to include the effects of exchange and exclusion of the valence charge density from the core region.[13] The values of $v_m(q)$ so obtained agree very well with Fermi surface data.

We believe that a better model for the isotropic covalent atom starts from the covalent pseudopotential form factor defined by

$$v_s(q)=v_a(q)=v_i(q)/\epsilon_s(q)\,, \tag{2.10}$$

---

[7] J. C. Phillips and L. Kleinman, Phys. Rev. **116**, 287 (1959).

[8] M. L. Cohen and T. K. Bergstresser, Phys. Rev. **141**, 789 (1966).

[9] J. C. Phillips, Phys. Rev. **112**, 685 (1958); W. Saslow, T. K. Bergstresser, and M. L. Cohen, Phys. Rev. Letters **16**, 354 (1966).

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

[11] V. Heine and I. Abarenkov, Phil. Mag. **9**, 451 (1964); **12**, 529 (1965).

[12] A. O. E. Animalu, Phil. Mag. **11**, 379 (1965).

[13] A. O. E. Animalu and V. Heine, Phil. Mag. **12**, 1249 (1965).

[14] M. H. Cohen and J. C. Phillips, Phys. Rev. **124**, 1818 (1961).

TABLE I. The contributions to the scattering factor per atom in diamond of the covalent charge density computed in several different ways. In column III are listed the results obtained when the scattering of the free atom Hartree-Fock charge density is subtracted from the experimental values (Ref. 5). Column IV gives the difference between Bennemann's complete T matrix and the atomic t matrices only. Column V shows the difference between experiment and the atomic t matrices.

K	S_b	Expt − (free atom)	$T-t$	Expt $-t$
$\langle 111\rangle$	1	0.14	0.05	0.12
$\langle 220\rangle$	0	0	−0.15	−0.09
$\langle 311\rangle$	−1	−0.07	−0.17	−0.11
$\langle 222\rangle$	−2	−0.15	−0.13	−0.15
$\langle 400\rangle$	−2	−0.11	−0.13	−0.16

where $\epsilon_s(q)$ is the dielectric function calculated by Penn for his model isotropic semiconductor defined in terms of covalent energies and basis states by (2.4)–(2.6). The first feature of (2.10) that should be noted is that because $\epsilon_s(0)$ is finite, $v_a(q)$ does not describe a neutral atom. The reason for this must be understood before we proceed further. The following discussion is guided in part by the statement (made both by Kleinman-Phillips[4] and Bennemann[6]) that the difference between the exact charge density and the linearly screened charge density is described by transfer of charge from the atom to the bond.

III. COVALENT BOND—MACROSCOPIC RELATIONS

In (2.9) and (2.10) only the diagonal part $\epsilon(q)$ of the general dielectric response tensor $\epsilon(q, q+K)$ is used to screen the bare pseudopotential form factor. This procedure,[10] sometimes called the neglect of local field corrections, is suitable for many metallic crystals, but not for covalent crystals. In our case we may anticipate the formation of covalent bonds, and regard this as the principal local field correction. The deficiency in valence charge of the covalent atom is concentrated at the covalent sites located halfway between nearest neighbors. For the specific case of tetrahedral bonding considered here, this means that each atom with four valence electrons contributes the charge

$$Z_b=-|e|/\epsilon_s(0) \tag{3.1}$$

to each of the four bonds. The total charge in each bond is $2Z_b$.

The spatial distribution of the bonding charge can be inferred in diamond from its contribution to the x-ray scattering factors measured experimentally[5] or calculated theoretically.[4,6] Brill[15] has represented the valence charge density as a superposition of spherically symmetric atomic charges plus bonding charges. The scattering factors $F_{\mathbf{K}}$ are then determined from expressions of the form

$$F_{\mathbf{K}}=S_a(\mathbf{K})f_a(K)+S_b(\mathbf{K})f_b(K)\,, \tag{3.2}$$

where $S_{a,b}(\mathbf{K})$ represent the structure factor evaluated at atomic and bonding sites, respectively, and $f_{a,b}(K)$ are the atomic and bonding form factors. Depending on the way in which $F_{\mathbf{K}}$ and $f_a(K)$ are evaluated one obtains different results for $f_b(K)$. These are shown in Table I. It appears reasonable to impose the condition that $f_b(K)$ should be a nonincreasing function of K. In this case the most consistent results are obtained when $F_{\mathbf{K}}$ is taken from experiment[5] and $f_a(K)$ is determined from the free-atom Hartree-Fock charge density.[4] This gives

$$f_b(K)=(0.21\pm0.05)[1+(K/2k_F)^2]^{-1}(e/\text{atom}) \tag{3.3}$$

for $K=\langle 111\rangle$, $\langle 220\rangle$, $\langle 222\rangle$, $\langle 311\rangle$, and $\langle 400\rangle$ (in units of $2\pi/a$). The result (3.3) tells us that the characteristic dimensions of the bonding charge density are very much smaller than those of the atomic charge density. (The characteristic wave number of the latter is $k_F\approx|\langle 111\rangle|$.) We may therefore replace the bonding charge density by a point charge of magnitude $2Z_b$.

Comparison of (3.1) and (3.3) provides the first quantitative check on our model. The value of $\epsilon(0)$ in diamond is 5.8 according to optical measurements.[16] Thus (3.1) gives

$$f_b(0)=1/\epsilon_s(0)=0.17(e/\text{atom})\,. \tag{3.4}$$

Considering that $\epsilon_s(0)$ is a macroscopic quantity, whereas the wave numbers q that are important for microscopic considerations are in the region $q\gtrsim k_F$, we regard the agreement between (3.3) and (3.4) as quite remarkable. This agreement provides some motivation for our *a posteriori* model.

Note added in proof. We have found a simple microscopic derivation of bonding charges that shows that $f_b(K)$ is isotropic, as assumed in (3.3), and that its characteristic wave number is indeed $2k_F$. This derivation appears in a forthcoming monograph to be published by the University of Chicago Press.

IV. COVALENT BOND—MICROSCOPIC EFFECTS

The metallic form factors calculated by Animalu and Heine (AH)[13] are compared in their Table V with experiment for elements in a metallic environment (Al and Pb) and in a covalent environment (Si and Ge). The agreement is quite good for the metallic case, but there is a significant discrepancy of 0.03 Ry in Si and 0.04 Ry in Ge at the value $x=0.55$, where

$$x=q/2k_F\,. \tag{4.1}$$

From the survey of fourteen semiconductors carried out by Cohen and Bergstresser,[8] one can see that this

---

[15] R. Brill, Z. Electrochem. **63**, 1088 (1959). Note that Brill includes *both* carbon atoms in his definition of $f_c(K)$.

[16] H. R. Philipp and E. A. Taft, Phys. Rev. **136**, A1445 (1964).

discrepancy is a genuine one, caused by the covalent environment, and is not an artifact of the semiempirical method of determining values of the form factor. (The empirical values of v_K for $K=\langle 220\rangle$ and $\langle 311\rangle$ on the other hand, are probably not significant to better than 0.01 Ry; at such large K valence effects are small and the empirical and model values agree within 0.01 Ry, as expected.)

The crystalline pseudopotential is

$$v_{\text{tot}}(q)=S_a(q)v_a(q)+S_b(q)v_b(q)\,, \tag{4.2}$$

so that the apparent atomic pseudopotential values are

$$v_p(K)=v_a(K)+[S_b(K)/S_a(K)]v_b(K)\,. \tag{4.3}$$

For $K=\langle 111\rangle$, $S_b(K)/S_a(K)=2^{-1/2}$. In Si the parameters are

$$E_g=4.8\text{ eV}\,,\quad E_F=12.5\text{ eV}\,,\quad \hbar\omega_p=17\text{ eV} \tag{4.4}$$

and using these values in Penn's formulas for $\epsilon_f(q)$ and $\epsilon_s(q)$ we obtain

$$\epsilon_f(0.55)=2.01\,,\quad \epsilon_s(0.55)=1.60\,. \tag{4.5}$$

Using AH's value of $v_m(0.55)=0.18$ Ry for Si, as well as (2.9), (2.10), and (4.5), we obtain

$$v_a(0.55)=-0.225\text{ Ry}\,, \tag{4.6}$$

$$\Delta v_a=v_a-v_m=-0.045\text{ Ry}\,. \tag{4.7}$$

On the other hand we calculate $v_b(K)$ from Poisson's equation to be

$$v_b(0.55)=(8/3\pi)(Z_be^2/a) \tag{4.8}$$

which gives (using $Z_b=\frac{1}{12}$ and $a=10.25a_0$)

$$[S_b(0.55)/S_a(0.55)]v_b(0.55)=0.010\text{ Ry}\,. \tag{4.9}$$

Combining (4.6) and (4.9) we obtain

$$v_p(K)=-0.215(-0.21) \tag{4.10}$$

$$\Delta v_p(K)=-0.035(-0.03)\,. \tag{4.11}$$

The semiempirical values[8] are shown in parentheses. The obtained agreement provides a check on our *a posteriori* model.

The arithmetical details of this calculation are of interest because they show explicitly how the separation of the valence charge density into covalent atomic plus bonding components represents a rather small change from metallic atoms. It is this small change which is decisive in determining the structural properties of the crystal. The strength of the method lies in the fact that the changes are computed entirely in terms of the crystal structure and the universal function $\epsilon_s(x;N,E_g)$ which manifestly reduces to $\epsilon_f(x;N)$ in the limit $E_g=0$.

To close this section we discuss the assumed isotropy of $\epsilon_s(q)$ which is essential to the separation of atomic and bonding charge densities. A realistic model (empirical pseudopotential) has been used[17] to calculate $\epsilon_s(\mathbf{q})$ in Si for $\mathbf{q}$ along $\langle 100\rangle$ and $\langle 111\rangle$ directions. The results show almost no anisotropy, in agreement with Penn's model and the assumption of this paper. In fact, the calculated curves[17] are in excellent quantitative agreement with Penn's results; both exhibit a small *peak* in $\epsilon(q)$ near $q=0$ due to umklapp effects.

V. GROUND-STATE ENERGY

We have indicated in Secs. II and III that both the energy gap E_g and the strength Z_b of the covalent charge can be estimated directly from the macroscopic dielectric constant $\epsilon(0)$. This is certainly a remarkable result, but from a microscopic point of view it places too much emphasis on the macroscopic region $(q\approx 0)$, whereas the region of structural interest is $q\lesssim 2k_F$. In this section we therefore compute the valence energy in order to be able to minimize the ground-state energy of the crystal with respect to N and E_g, thereby determining both these quantities self-consistently.

The energy per atom of the valence electrons is given by

$$Ev=E_1+E_2+E_3+E_4+E_5\,, \tag{5.1}$$

where E_1 is the sum of the single-particle energies

$$E_1=\sum_{k\lesssim k_F}(E_k{}^{+}+C_0) \tag{5.2}$$

and E_2 represents a correction to the energy of the covalent atoms arising because potential energies are counted twice in (5.2). This is

$$E_2=\frac{1}{2}\int_0^\infty \rho_s(r)v_s(r)4\pi r^2dr\,. \tag{5.3}$$

The energy E_3 arises because E_1 and E_2 do not describe a system of neutral atoms. Because Z_b is small and because the bonding charge density is well localized at the bonding sites, we can express E_3 as the sum of two terms. The first is the Madelung energy of a system of point charges $+4Z_b|e|$ located at each atomic site, and $-2Z_b|e|$ at each covalent site. Thus

$$E_{31}=-\tfrac{1}{2}(Z_be)^2\alpha_M/a\,, \tag{5.4a}$$

where α_M is the Madelung constant for this system, a is the lattice constant, and the factor of $\frac{1}{2}$ normalizes the energy per atom, assuming two atoms per unit cell.

The second contribution to the Madelung energy arises from the Coulomb, exchange, and correlation energy between the valence electrons screening the atom and the bonding charges $-2Z_b|e|$. In obtaining E_{31} we have counted this energy twice. Thus we must add the correction term

$$E_{32}=[-\tfrac{1}{4}Z_b(Z-Z_b)e^2(\alpha_M/a)]\bar{f}\,, \tag{5.4b}$$

---

[17] H. Nara, J. Phys. Soc. Japan **20**, 778 (1965).

where $\bar{f}$ is the average value of the Hubbard factor[11] $f(k,q)=[1-q^2/2(q^2+k^2+k_s^2)]$ for this interaction. With $q=k_F$, $k=k_F$, and $k_s=k_F$, we have $\bar{f}=\frac{5}{6}$.

The single-particle energies $(E_k{}^{+}+C_0)$ are given by Penn's expression $(P,2.6)$ for $E_k{}^{+}$, plus the constant C_0. The origin of the latter can be seen by inspection of our model secular Eq. (2.6). The constant C_0 appears on the diagonal and is given by the expectation value in a plane-wave state of the electrostatic energy of the neutral covalent pseudoatom,

$$C_0=-\frac{|e|}{\Omega}\int_0^\infty v_s'(r)4\pi r^2dr\,, \tag{5.5}$$

where Ω denotes the atomic volume and $v_s'(r)$ is the neutral atomic potential

$$v_s'(r)=v_s(r)-(4Z_b|e|/r)\,. \tag{5.6}$$

The term in Z_b arises in (5.6) because we count the Madelung energy in E_3.

Finally, the energy E_4 arises from the overlap of the neutral covalent atomic charges described by the potentials $v_s'(r)$. It can be calculated using methods developed by Ewald, as discussed, e.g., by Cochran.[18] Here again E_4 contains a term arising from not counting valence energies twice. The energy E_5 is defined in Sec. VI.

There is one feature of the expression (5.1) for the ground-state energy which deserves further discussion. From a semiclassical point of view it may be surprising that the bonding charges contribute the potential energy (Madelung) term to E_v, but at the same time there is no explicit kinetic-energy term corresponding to the anisotropic accumulation of valence charge at the bonding sites. The reason for this has been discussed previously.[19] The accumulation of bonding charge arises, as can be seen explicitly in the Fourier synthesis[4] of the valence charge density from OPW's, from constructive interference of the covalent basis functions (2.5), centered on nearest neighbors, at the bonding sites. This interference (called crystal hybridization)[19] cannot be described in a semiclassical framework (Thomas-Fermi method) using a lumped charge density. The interference arises from the crystal structure itself, and the accumulation of bonding charge requires no additional kinetic energy beyond what is already implicit in E_k.

VI. EXCHANGE AND CORRELATION ENERGIES

In the Hartree approximation in (5.3) $\rho_s(r)$ and $v_s(r)$ are related through Poisson's equation. We write

$$v_s(r)=v_i(r)+v_1(r)\,, \tag{6.1}$$

where $v_i(r)$ is the Fourier transform of $v_i(q)$. Then in the Hartree approximation

$$\nabla^2 v_1(r)=-4\pi\rho_s(r)\,. \tag{6.2}$$

We can use (6.2) to compute $\rho_s(q)$, which is

$$\rho_s(q)=\frac{q^2}{4\pi}v_1(q)=\frac{q^2}{4\pi}\left[\frac{1}{\epsilon_s(q)}-1\right]v_i(q)\,. \tag{6.3}$$

The relation (6.3) determines $\rho_s(q)$ in terms of $v_i(q)$, N and E_g.

When the effects of exchange and correlation are included in the screened atomic potential, (6.2) is no longer valid. A number of different procedures have been proposed to modify the dielectric function to include the effects of screened exchange. We prefer the discussion of exchange and correlation in terms of a local potential given by Heine and Abarenkov.[11] When the discussion is extended to calculate the many-body corrections to (5.3), one obtains the result described by Heine and Weaire[20] in an Appendix. The further corrections due to nonlocal many-body effects have been discussed by Animalu.[12] They are small and attain a maximum value of about 3% for $q\approx k_F$.

We therefore conclude that in calculating the ground-state energy E_v it is sufficient to treat exchange and correlation by local dielectric screening techniques, although the crystalline charge density is very far[4,6] from being uniform. If it is desired to include very small corrections for the orthogonality hole in the region of the atomic cores, or the accumulation of charge in the covalent bond, this may be done by the methods used for metals.[13] Otherwise E_5 represents the exchange and correlation energy of a uniform electron gas of density N with interactions screened by Penn's dielectric function.

If we accept (6.3), then (5.3) becomes by Parseval's theorem

$$E_2=\frac{1}{2(2\pi)^3}\int_0^\infty \frac{q^2}{4\pi}[1-\epsilon_s(q)]|v_s(q)|^2 4\pi q^2dq\,. \tag{6.4}$$

VII. COMPARISON WITH T-MATRIX THEORY; GOLDSTONE'S THEOREM

Evaluation of the ground-state energy E_v as a function of N and of E_g, for a given ionic pseudopotential form factor $v_i(q)$, requires detailed numerical calculations that lie beyond the scope of this paper. The numerical tests discussed in Secs. III and IV can, however, be compared with the results of Bennemann's T-matrix calculations. The comparison shows why it is desirable to use the covalent spectrum and the covalent basis states (2.5) from the outset.

Bennemann divides his calculation[6] of the charge density in diamond into three stages. The first stage

---

[18] W. Cochran and R. A. Cowley, in *Handbuch der Physik* (Springer-Verlag, Berlin, 1967), Vol. XXV 2a, p. 59.

[19] J. C. Phillips, in *Proceedings of the International Conference on Semiconductor Physics, 1960* (Czechoslovakian Academy of Sciences, Prague, 1961), p. 41.

[20] V. Heine and D. Weaire, Phys. Rev. 152, 603 (1966).

corresponds to linear metallic screening—our Eq. (2.9). The second stage corresponds to nonlinear metallic screening, in which $v_m(q)$ is replaced by $t_m(q)$. At this stage ρ_K for $K=\langle 111\rangle = 0.55k_F$ is *increased* by the nonlinear theory by 10% over its metallic value, corresponding to a compression of the atomic charge density. We have seen in Sec. IV, however, that $\rho_s(q)$, as given by Eq. (6.3), actually decreases by 20% on passing from the metallic atom to the covalent atom. This conclusion agrees with the intuitive feeling that, because of lower N, smaller coordination number, or no "pressure" from the electron gas in open or "antibonding" directions, the covalent atom should expand relative to the metallic atom. The t-matrix method arrives at the opposite conclusion because the boundary conditions implied by the free-electron dielectric function do not provide a realistic description of the geometry of the covalent atomic cell.

We may summarize the results of nonlinear metallic screening when applied to a covalent structure as follows. The results produced for the covalent bonding charge density itself are reasonable. However, because of the absence of an energy gap in the spectrum of the medium, the good results for the bond are achieved at the price of worsening the atomic charge density. Such a situation can easily arise by perturbation theory when it represents an asymptotic rather than an absolutely convergent series.

It is interesting to consider the presence of an energy gap in our model from the viewpoint of the axiomatic symmetry theorems discussed by Goldstone[21] and by Lange.[22] These were stimulated by the observation that the BCS theory of superconductivity (to which Penn's model of an isotropic semiconductor bears certain resemblances) is a case of broken symmetry (like a Heisenberg ferromagnet), yet there are no macroscopic excitations (like spin waves) whose energy vanishes in the long-wavelength limit. Anderson showed[23] that finite-energy plasma oscillations obliterate the corresponding mode in the superconductor. Lange has shown[22] that a long-range (Coulomb) interaction must be present in the system for this to happen.

In our case the connection between the energy gap and the long-range interactions is contained simply in (2.4a) and (3.1). The absence of broken symmetry in our model suggests that (as a sort of converse to the Goldstone theorem) long-range forces are implicit in the model, and that the covalent atom should *not* be neutral. The Goldstone theorem is amusing because it is most interesting in situations where it fails to apply.

In closing this field-theoretic discussion, we ask whether the Madelung energy E_3 should not be derived (rather than postulated). In fact it is only through E_3 that the crystal structure itself appears explicitly in the present theory. From the point of view of economy, it seems no worse to assume the existence of E_3 than to postulate a given structure factor for the crystal. In this sense the theory is as close to being a "bootstrap" theory as one could expect for a crystal.

The expressions (5.4a) and (5.4b) illustrate the character of our model which divides the valence charge density into two components (analogous to two particle fields). The electrostatic term (5.4a) is corrected by (5.4b) because the two components both consist of electrons. This means first, that the interaction energy must not be counted twice, and second, that there is exchange and correlation between the two fields, as described by $\bar{f}$. Elsewhere in the theory f is implicit, because it has been incorporated into $\epsilon_s(x,N,E_g)$.

In the present theory N and E_g are both regarded as parameters of the medium. If the theory is used in an *a posteriori* manner, both are taken from the observed crystal structure and macroscopic dielectric constant. On the other hand, in self-consistent form both are determined from minimizing E_v. One may contrast our approach, which treats N and E_g on an equal footing, with the approach of conventional Bloch energy-band theory. There N is given its observed value, but E_g is determined by calculating the optical excitation spectrum from a given crystal potential (which is ordinarily not self-consistent). Thus conventional band theory mixes the *a posteriori* and self-consistent approaches in an unsymmetrical way dictated by considerations of computational convenience. The present microscopic model is computationally simple and therefore it is possible to treat both medium parameters symmetrically.

VIII. INTERNAL STABILITY

The positions of the bonding charges in elemental covalent crystals (such as diamond or graphite) can be determined in equilibrium from symmetry considerations alone. As soon as the symmetry is perturbed [e.g., by a lattice wave propagating along a (111) axis, or in a heteropolar semiconductor], we require a general prescription for determining the bonding sites.

An obvious prescription is the classical one that the forces on the bonding charge vanish, so that the sites are positions of mechanical equilibrium. This describes the sites in the symmetrical cases. It cannot be correct, however, as a moment's thought tells us that the sites are positions of unstable equilibrium.

The correct condition is variational and quantum-mechanical in character. Define the overlap function of local potentials $V_\alpha(\mathbf{r}=\mathbf{R}_\alpha)$ as follows:

$$O_\infty(\mathbf{r})=\frac{1}{2}\frac{|\sum_\alpha{}' V_\alpha(\mathbf{r}-R_\alpha)|^2}{\sum_\alpha{}'|V_\alpha(\mathbf{r}-R_\alpha)|^2}. \qquad (8.1)$$

Then the covalent sites are the places where $O_\infty(\mathbf{r})$ attains its maximum value. It is easy, if it is so desired,

---

[21] J. Goldstone, Nuovo Cimento **19**, 154 (1961).

[22] R. V. Lange, Phys. Rev. Letters **14**, 3 (1965); Phys. Rev. **146**, 301 (1966).

[23] P. W. Anderson, Phys. Rev. **112**, 1900 (1958).

to extend the prescription (8.1) to include nonlocal potentials.

If we neglect bond-bond interactions, (8.1) can be simplified by determining the location of each site in terms of the two nearest screened ion potentials. This gives

$$O_2(\mathbf{r}) = \frac{1}{2}\frac{|V_1(\mathbf{r})+V_2(\mathbf{r}-\boldsymbol{\tau})|^2}{|V_1(\mathbf{r})|^2+|V_2(\mathbf{r}-\boldsymbol{\tau})|^2}. \tag{8.2}$$

According to Schwartz's inequality, $O_2(\mathbf{r}) \leq 1$. The maximum value $O_2(\mathbf{r})=1$ is attained when

$$V_1(\mathbf{r}) = V_2(\mathbf{r}-\boldsymbol{\tau}). \tag{8.3}$$

Note that (8.1)–(8.3) involve potentials, not fields, throughout, which reflects the quantum-mechanical character of covalent bonding. Because (8.2) is maximized by (8.3), the locations of the covalent sites derived from O_2 are manifestly gauge invariant. This is not the case for O_∞, which involves long-range forces. Since covalent forces are short-range, we fix the gauge to be used in calculating O_∞ by requiring that in general it shift the covalent sites by as little as possible compared to O_2. This is equivalent to choosing the constant in $V_\alpha(r-R_\alpha)$ so that

$$\lim_{r'\to\infty} V_\alpha(r') = 0. \tag{8.4}$$

IX. APPLICATIONS, EXTENSIONS, AND LIMITATIONS

The foregoing formalism should yield quite satisfactory quantitative results for any property of monatomic covalent crystals which depends on the valence electron gas as a whole. This includes the phonon spectrum (in particular the second- and third-order elastic constants) and the pressure dependence of the dielectric constant. It is interesting to note that the latter has been found[24] to satisfy a universal relation in diamond, Si, and Ge.

The theory may readily be extended to crystals where neutral bonding[25] is predominant, such as the III–V and II–VI semiconductors. Because there is no longer a center of symmetry in the unit cell, it is convenient to divide[8] the ionic potential into symmetric and antisymmetric parts, v_i^s and v_i^a. The purely covalent crystal is generated by v_i^s, which determines N and E_g self-consistently. This gives $\epsilon_s^s(q)$ and from it (e.g., in a III–V crystal) in the approximation of homopolar screening,

$$V^{\mathrm{III}}(q) \approx V_s^{\mathrm{III}} = V_i^{\mathrm{III}}(q)/\epsilon_s^s(q), \tag{9.1}$$

$$V^{\mathrm{V}}(q) \approx V_s^{\mathrm{V}} = V_i^{\mathrm{V}}(q)/\epsilon_s^s(q). \tag{9.2}$$

From (9.1) and (9.2) one can obtain by Fourier transform $V_s^{\mathrm{III}}(r)$ and $V_s^{\mathrm{V}}(r)$. The charge in the covalent bond is still $2Z_b$, and it is still concentrated on the line connecting nearest neighbors at $\mathbf{R}_l$ and $\mathbf{R}_l+\boldsymbol{\tau}$. The equilibrium position along this line is determined approximately by

$$V^{\mathrm{III}}(\mathbf{r}-\mathbf{R}_l) = V^{\mathrm{V}}(\mathbf{r}-\mathbf{R}_l-\boldsymbol{\tau}). \tag{9.3}$$

In cases of large ionicity, the equilibrium position, together with N and E_g, must be determined self-consistently by iteration. Again $V_i^{\mathrm{III}}(q)$ and $V_i^{\mathrm{V}}(q)$ determine all the parameters of the theory.

It is an interesting question whether the theory can be generalized further to treat molecules. We recognize at the outset, of course, that the spacing of the molecular energy levels must be small compared to the energies of interest (such as the binding energy) before a continuum approach is appropriate. Granted that the molecule is sufficiently large, we have already seen in Sec. VII the importance of boundary conditions (or the pressure of the electron gas) in determining atomic size. If each molecular calculation required the determination of a separate dielectric function for that molecule, little would be gained by this approach. Nevertheless, as the discussion of III–V crystals indicates, the theory is so simple as to be quite flexible, and it may be possible at a later stage to generalize the theory quantitatively to certain classes of large molecules. Such a generalization could be described as *quantum electrostatics*. By this we mean that the potential energy of the electron screening cloud is calculated (apart from known correction factors for exchange and correlation) by means of classical electrostatics. The kinetic energy (which has no classical electrostatic analogue in terms of multipole moments) is determined by phase space (quantum) considerations.

In conclusion we add, by way of caution, some remarks on the limitations of the theory. Because of the large anisotropy of the energy gap in real semiconductors, the theory will not give accurate results in situations which involve only a few valence electrons. Such a situation, e.g., would be the temperature dependence of E_g. Other examples will readily occur to the reader; in these cases recourse must be had to complete band calculations.[8]

ACKNOWLEDGMENTS

I am grateful to Professor J. M. Ziman for reviving my interest in the covalent bond. This work was made possible by a Fellowship from the Alfred P. Sloan Foundation, and was supported in part by the Office of Naval Research and a general grant to the University of Chicago by the Advanced Research Projects Agency. I have also benefitted from conversations with members of the theoretical group at Bell Laboratories (Murray Hill).

---

[24] D. F. Gibbs and G. J. Hill, Phil. Mag. 9, 367 (1964).

[25] J. C. Slater and G. F. Koster, Phys. Rev. 94, 1498 (1954).