REVIEWS OF MODERN PHYSICS VOLUME 42, NUMBER 3 JULY 1970

Ionicity of the Chemical Bond in Crystals

J. C. PHILLIPS
Bell Telephone Laboratories, Incorporated
Murray Hill, New Jersey 07974

The nature of the chemical bond in crystals is discussed. The general theories of L. Pauling based on thermochemical data and of C. A. Coulson based on valence bond concepts are compared with a recent spectroscopic theory. Particular emphasis is placed on binary crystals of formula A^NB^{8-N} which includes most tetrahedrally coordinated semiconductors as well as crystals of the rocksalt (NaCl) family. A wide range of physical properties is discussed, including crystal structure, energy bands, elastic constants, ionization energies, and impurity states. The role of quantum-mechanical sum rules and spectral moments in constructing simplified models of bond and band behavior is explored. Stress is laid throughout on methods for incorporating quantum-mechanical effects into properties of chemical bonds through algebraic relations rather than through variational solutions of the wave equation.

CONTENTS

1. WHY STUDY THE CHEMICAL BOND IN CRYSTALS?

One of the many apocryphal stories about the great men of science concerns the illustrious mathematical physicist P. A. M. Dirac. His theory of the relativistic electron was abstract and mysterious to a generation of physicists nurtured on the concepts of classical mechanics. An intrinsic feature of his theory was the symmetry between electrons in positive energy states and holes (or positrons) in negative energy states. When the positron was discovered, Dirac's reaction epitomized the attitude of the pure physicist. "The rest," he is reported by his colleagues at Cambridge to have said, "is chemistry."

Dirac's prediction has proved correct. The neutron was discovered to be the catalyst of nuclear reactions, and there may well be as many "elementary" particles as there are hydrocarbons, although the world is possibly not large enough to accommodate the cyclotron needed to find all of them. In any event, if by chemistry one means substantive concepts concerning the structure of matter, then this is as much the domain of modern physics as it is of modern chemistry.

Unfortunately, many physicists who cannot emulate Dirac's mathematical accomplishments have adopted an attitude of lofty indifference to substantive theory. Yet today the need for such theories is greater than ever. Marvellous experimental techniques have been developed to synthesize new materials and to measure their properties on an atomic scale. The experimental mathematicians with their giant computers produce reams upon reams of numbers, but then interpret them using makeshift concepts introduced two or three or more decades ago, when an intelligent guess at what was needed was all that could be achieved.

The thesis of this article is that many new concepts can be generated by studying the chemical bond in crystals. This thesis is in direct conflict with the philosophy of classical physical chemistry, which starts with individual atoms and constructs more complex systems by bringing atoms together one at a time to form larger and larger molecules. According to this viewpoint, the simplest system of aggregate atoms is the diatomic molecule, and the most complex is the macroscopic crystal.

This atomistic view is, in fact, not that of organic chemistry. The organic chemist begins with building blocks, such as the benzene ring, which are appropriate to the large molecules of interest. Similarly, crystals may actually be simpler than molecules in some respects. It is possible to find crystals which contain only one type of bond and no unbonded electrons. For example, sp^3 bonds between B and N are best studied in cubic BN, and sp^2 bonds in layer BN. The closest molecular approximation to the former is the ethanelike molecule H_3BNH_3, and to the latter the benzenelike molecule

$B_3N_3H_6$. In covalent systems the itinerant character of the valence electrons means that the study of one kind of bond is necessarily more complicated when other kinds of bonds are present.

There is another good reason for studying the chemical bond in crystals. Molecular orbital theory, as a method of calculation, is unavoidably dependent on the use of a definite kind of representation, that based on atomic orbitals. Thus it is customary to define bond order, bond populations, etc., in terms of the coefficients of these atomic orbitals. But such quantities are not observables, and according to the operational philosophy of the underlying quantum mechanics, it is not desirable to introduce such "hidden variables" into the theory. I have dealt with the problem of representation-independent concepts at considerable length elsewhere (Phillips, 1970), and so will not give an extended discussion here. Suffice it to say that by using dispersion theory it is possible to discuss the chemical bond in crystals employing quantum mechanics, but without reference to any particular representation.

As an example of the kind of problem to which we address ourselves, consider the question of ionicity. A typical predominantly ionic bond is Na^+Cl^-. From studies of both crystals and molecules it is known that there is a qualitative relationship between the dipole moment of such a bond and its ionicity. But if one wants to take full advantage of such a concept, one must be able to treat not only strongly ionic situations involving closed-shell atoms, as in the example of Na^+Cl^-, but also cases of weak and intermediate ionicity involving open-shell configurations. Here the correlation between ionicity (as defined, for example, through differences in electronegativity or differences in ionization potential) and dipole moment practically disappears. This leads one to question the utility of ionicity as a tool for structural analysis of predominantly covalent materials. This question, one should note, is basic to the treatment of heteroatoms in theories of the Hückel, valence bond, or molecular orbital type. It will be answered here for crystals, and it will turn out that the answer is representation independent.

Another question of great structural significance is the validity of the one-electron approximation. For light diatomic molecules some information is available on this question from elaborate computer calculations, but it is not easy to see how to apply this knowledge to larger molecules and crystals. Because our theory is representation independent, by implication it provides a basis for analyzing this question. Most of our discussion will utilize one-electron language for convenience, because this is the language that will be most familiar to most readers. Most of our results, however, do not depend on this approximation. Those results which do can be modified to include the effects of electron correlation and exchange in a systematic manner that is *not* representation dependent.

2. PAULING'S THEORY OF ELECTRONEGATIVITY AND IONICITY

> *"What we call quantum theory [is] the entering wedge of scientific bolshevism." G. N. Lewis*, Valence *(1916)*

The concept of the relative electronegativity of the elements is an old one which arose in connection with oxidation-reduction potentials in the eighteenth century. Crystallographers early noted that binary compounds of atoms A and B with large differences in electronegativity tended to form rocksalt structures, while smaller differences in electronegativity favored more open covalent structures. When the difference in electronegativity is large, the heat of formation of the AB bond is large, providing a thermochemical use for the concept of electronegativity. The greater the difference in electronegativity, the more ionic the bond is said to be, and the greater is its heat of formation.

With the advent of quantum mechanics it appeared that chemistry would finally be put on a quantitative footing, and the qualitative ideas with structural, electrochemical, and thermochemical origins would be precisely defined. Now, more than forty years after the establishment of the wave equation and the explanation of the properties of atoms and of diatomic molecules, we seem to be farther than ever from agreeing on the meaning of these terms.

Why is this the case? In practice, because the wave equation is so difficult to solve, workers have often become so enmeshed in the details of their work that they lose sight of the broad empirical ideas that formed the basis of classical theory. For this reason the two broad approaches to electronegativity, ionicity, and bond energies—the thermochemical one and the molecular orbital one—are no closer to each other now than they were three decades ago. At the risk of creating further disagreement, a third approach to the problem, based on spectroscopic considerations, is presented here. The third approach makes possible comparison of the first two approaches. It shows that, at least conceptually, all approaches have a great deal in common in spite of their use of quite different operational methods. The great precision achieved spectroscopically permits evaluation of the relative merits of each approach.

The thermochemical approach, pure, simple, and unsullied by quantum-mechanical paraphenalia, is advocated by Pauling (1960). He defines electronegativity as "the power of an atom in a molecule to attract electrons to itself." This definition, which reflects the historical origins of the concept of charge flow in electrochemistry, focuses attention on the actual charge centered on each atom. However even if the complete charge distribution of the molecule or crystal were known precisely, one would still need a prescription for decomposing the total distribution into a superposition of distributions centered on component atoms.

To circumvent this difficulty, Pauling turns from charge distributions to bond energies, which are known from heats of formation. The cohesive energy of diamond, for instance, is 160 kcal/mol, and since each C molecule corresponds to two bonds, the C–C single bond energy is about 80 kcal/mol. (A hydrocarbon value of 85 kcal/mol is also quoted at times. Here we are primarily concerned with crystals, and so quote the diamond value.) In this case the difference in electronegativity is zero by definition.

When two elements, e.g., A and B, differ in electronegativity (denoted by X_A and X_B, respectively), one finds that in general the heat of formation D_{AB} of the AB bond satisfies the relation

$$D_{AB} > (D_{AA} + D_{BB})/2, \tag{2.1}$$

where D_{AA} and D_{BB} represent the bond energies of the elements A and B, respectively. Of course the AB bond energy generally refers to a structure in which the coordination numbers of A and B are different from what they were in the pure A or pure B structures. Thus, in the language of atomic orbitals, the states of hybridization of A and B are different in the AB compound than they were in the pure A and pure B compounds, which raises the question of the orbital dependence of electronegativity.

The ionization energies of free atoms are known to be quite different for different multiplets, and when a suitable average over multiplets is taken corresponding to each state of hybridization, one finds substantial variations in the ionization energies. This led Moffitt (1950) to suggest that differences in hybridization influence the properties of predominantly covalent bonds to a greater extent than do differences in electronegativity. Moffitt substantiated this claim by studying C–H, N–H, and O–H bond lengths. However, the proton potential is much more singular than are other atomic valence potentials, because in other atoms the valence electrons do not penetrate the core region because of the exclusion principle. At the same time the electron–electron interactions which are responsible for multiplet formation in polyvalent atoms are less important in large molecules and crystals, where the spacing of energy levels is much smaller than a typical bond energy. One is therefore justified in retaining the concepts of electronegativity and ionicity. Differences in hybridization states can be taken into account by following the usual Hückel approach (Streitweiser, 1961), and allowing the parameters of the theory to vary with bond length in a smooth manner, thereby allowing quantitatively for most hybridization effects. Ideally one should study homologous systems in which the hybridization states vary as little as possible, in order to isolate trends associated entirely with electronegativity and ionicity. That is the procedure followed here in studying $A^N B^{8-N}$ crystals.

Pauling resolves this problem by using the concept of resonating bonds. In simple situations, e.g., $A^N B^{8-N}$ crystals, the number of resonating bonds M per atom is equal to its classical valence, i.e., $M = N$. In diamond, $M = 4$; in cubic BN, $M = 3$; and in BeO, $M = 2$. When the number of bonds is different from the coordination number of the atom, one has "resonating" bonds. Thus in diamond each C atom shares four single bonds with its tetrahedrally coordinated neighbors, but in BeO only half of the valence electrons are in bonding states where they resonate from one neighbor to another. The other half of the valence electrons are localized about the oxygen anions. For more complicated structures, the rules for determining M become quite elaborate (Pauling, 1960).

Whenever (2.1) holds, Pauling defines the "extra-ionic" energy Δ_{AB} by the relation

$$\Delta_{AB} = D_{AB} - (D_{AA} + D_{BB})/2, \tag{2.2}$$

and relates this to the electronegativities X_A and X_B and the number of resonating bonds/mole by

$$\Delta_{AB} = -23M(X_A - X_B)^2. \tag{2.3}$$

The heat of formation can be obtained from (2.3) after allowing for the extra stability of O_2 and N_2 bonds. Let n_N and n_O be the number of nitrogen and oxygen atoms per AB molecule. Then the heat of formation ΔH_{AB} is given by

$$\Delta H_{AB} = -\Delta_{AB} + 55.4n_N + 26.0n_O. \tag{2.4}$$

If one accepts Pauling's prescriptions for M, then from a collection of heats of formation one can prepare a table of values of X_A. It has been found (Pritchard and Skinner, 1955) that the most consistent results are obtained by studying polyatomic molecules of the type ST_n (where T is an univalent atom) or SR_n (where R is an univalent radical). In this way one is more likely to be treating single bonds. The most recent compilation of this type appears to be that of Allred (1961), who used 16 homoatomic and 38 heteroatomic bond energies to determine his values of X_A for 69 elements from 162 computations.

It is interesting to check Pauling's basic formula (2.3) by using the values of X_A derived from polyatomic ST_n or SR_n molecules to calculate the mean value of M in tetrahedrally coordinated crystals $A^N B^{8-N}$.

For $N = 2$, one finds $\bar{M} = 2.2$ with an rms deviation of 1.2 (computed using ΔH itself as a weighting factor in performing averages, so as to minimize errors from cases like CdTe, where ΔH is itself small). For the case $N=3$, one finds $\bar{M}=4.2\pm1.5$. Thus the approximation $\bar{M}=N$ is not bad for $N=2$, but assuming there are as many bonds per molecule for $N=3$ as for $N=4$ gives better results, i.e., assuming that all bonds are present and equal the coordination number gives good results for the more covalent crystals. Perhaps the resonating-bond prescription for determining M is better suited to molecules than it is to crystals, where the valence electrons are spread throughout a large volume and interact with each other to a lesser extent.

In spite of the ambiguities attached to M (and hence to the entire table of electronegativities), Pauling goes on to use these values to make rough estimates of the fraction of ionic character $f_i(\text{AB})$ of an AB bond. He assumes that $f_i(\text{AB})$ is a function of $(X_\text{A}-X_\text{B})$ only, and is independent of other quantities (e.g., the bond length r_AB). Clearly f_i must be an even function of $X_\text{A}-X_\text{B}$, and as it measures the fraction of ionic character it should lie in the interval between zero and one. The conditions

$$f_i(\text{A, B})=f_i(\text{B, A}), \qquad (2.5)$$

$$0\leq f_i(\text{AB})\leq 1, \qquad (2.6)$$

are satisfied by all the definitions of ionicity that we discuss, including Pauling's definition:

$$f_i(\text{AB})=1-\exp[-(X_\text{A}-X_\text{B})^2/4]. \qquad (2.7)$$

There is a rough correlation between f_i as defined by (2.7) and dipole moments of diatomic molecules (Pauling, 1960).

3. VALENCE BOND THEORY

Many different attempts have been made to formulate the concepts of electronegativity and ionicity in the language of molecular orbital (MO) theory. For our present purposes the work of Coulson, Rèdei, and Stocker (1962) is most relevant. These authors, hereafter referred to as CRS, have studied the electronic properties of tetrahedrally coordinated crystals of the A^NB^{8-N} family using a simple MO description. Their prescription for treating electronegativity differences is the one advocated by Mulliken (1949). According to this approach, apart from an additive constant, the electronegativity of element A, denoted by X_A, should be proportional to the quantity α_A defined by

$$\alpha_\text{A}=(I_\text{A}+E_\text{A})/2, \qquad (3.1)$$

where I_A and E_A are the ionization energy and electron affinity, respectively, of the neutral atom A. The disadvantage of (3.1) is that it refers to the isolated atom, whereas what we want is something that describes the atom in its bonded state in the molecule or crystal. This can be achieved if the net charge Q on atom A is known from some other experiment, or Q can be determined self-consistently (Moffitt, 1949). Let us write

$$\alpha_\text{A}(Q)=\alpha_\text{A}+\alpha_\text{A}'Q. \qquad (3.2)$$

To determine Q and to determine the electronic distribution in the crystal A^NB^{8-N}, let φ_A and φ_B denote hybridized sp^3 orbitals suitable for tetrahedral coordination. In the $(\varphi_\text{A}, \varphi_\text{B})$ representation, the matrix elements of the Hamiltonian are

$$H_{11}=\alpha_\text{A}(Q), \qquad (3.3)$$

$$H_{22}=\alpha_\text{B}(-Q), \qquad (3.4)$$

$$H_{12}=\beta. \qquad (3.5)$$

By charge neutrality $Q_\text{A}=Q=-Q_\text{B}$, and according to Mulliken α_A' and α_B' can be determined from ionization energies of the neutral and singly positive atoms, and the electron affinity, which fix the slopes of energy from A^- to A^0 to A^+.

We must now determine the overlap parameters β. The first basic assumption made here (which is justified later in comparison with pseudopotential and spectroscopic models) is that β is a function only of the average principal quantum number of the valence electrons. Thus β is the same for diamond, BN, and BeO. One needs for interpolation purposes only the values of β in the crystals formed from Group IV elements, diamond, Si, Ge, and grey Sn. We refer to these collectively as β (IV).

From a thermochemical view, the proper choice of β (IV) is obvious: These energies are just the energy gained per electron from formation of the crystal, and since there are two electrons per bond,

$$\beta(\text{BeO})=\beta(\text{BN})=\beta(\text{C})=-40\text{ kcal/mol}$$
$$=-1.76\text{ eV}. \qquad (3.6)$$

Another representative value is $\beta(\text{Si})=-1.05$ eV. The model wave function (bonding state) is

$$\psi=P^{-1/2}(\varphi_\text{A}+\lambda\varphi_\text{B}), \qquad (3.7)$$

$$P=1+\lambda^2, \qquad (3.8)$$

and the secular equations are simply

$$\alpha_\text{A}-E+\lambda\beta=0, \qquad (3.9)$$

$$\beta+\lambda(\alpha_\text{B}-E)=0. \qquad (3.10)$$

The fraction of valence charge on atom A is P^{-1}, while that on atom B it is λ^2/P. From this one can determine Q and solve (3.9) and (3.10) self-consistently. This leads to an interesting quartic equation for λ, which is

$$\beta\lambda^4+J\lambda^3+K\lambda-\beta=0, \qquad (3.11)$$

where J and K are given in the general case by

$$J=\alpha_0+N\alpha', \tag{3.12}$$

$$K=\alpha_0-(8-N)\alpha', \tag{3.13}$$

with α_0 and α' given by

$$\alpha_0=\alpha_A-\alpha_B, \tag{3.14}$$

$$\alpha'=\alpha_A'+\alpha_B'. \tag{3.15}$$

For BN, one finds $\alpha_0 \approx 16.5$ eV; for BeO, $\alpha_0 \approx 26$ eV. Thus β is very much smaller than α_0 or J and K. The complex quartic equation (3.11) reduces within an accuracy of a few percent to

$$\lambda^2=[(8-N)\alpha'-\alpha_0]/(N\alpha'+\alpha_0). \tag{3.16}$$

In general α_A' is of the same magnitude as α_A. Thus from (3.14) and (3.15) there will be cancellation in α_0 between α_A and α_B, but addition in α'; and α' is several times α_0. This means that (3.16) can be reduced further,

$$\lambda^2 \approx [(8-N)/N]\{1-[N^{-1}+(8-N)^{-1}](\alpha_0/\alpha')\} \tag{3.17}$$

to first order in (α_0/α'). The net charge Q is given by

$$Q=Q_A=N-[8/(1+\lambda^2)], \tag{3.18}$$

where N represents the core charge of atom A. As one would expect, the leading term in (3.17) is canceled by the second term in (3.18), leaving quite simply

$$Q_A=-\alpha_0/\alpha'. \tag{3.19}$$

One notes that upon reversing A and B on the right-hand side of (3.19), the numerator reverses sign, while the denominator is unchanged. This verifies the charge neutrality requirement $Q_B=-Q_A$.

With the values of β proposed by CRS the terms first order in β are small. As we shall see, larger values of β may actually be more appropriate. In this case, working to first order in both α_0/α' and β/α', one finds corrections to (3.17) and (3.19). The first of these is

$$\lambda^2 \approx \frac{8-N}{N}\left[1-[N^{-1}+(8-N)^{-1}]\frac{\alpha_0}{\alpha'}\right]+\frac{\beta}{\alpha'}\frac{16(N-4)}{N^{5/2}(8-N)^{1/2}}, \tag{3.20}$$

and the second is

$$Q_A=-[\alpha_0+g(N)\beta]/\alpha', \tag{3.21}$$

$$g(N)=2(N-4)/[N(8-N)]^{1/2}. \tag{3.22}$$

With the sign convention $\beta<0$, one can verify that both terms in the bracket are positive for $N<4$, and both satisfy $Q_B=-Q_A$.

The CRS equations can be used to define an ionicity scale. In the crystal cell containing one $A^N B^{8-N}$ pair, the valence electrons are centered on the A atom P^{-1} of the time, on the B atom $(1-P^{-1})$ of the time. This suggests the definition

$$\begin{aligned} f_i(AB) &= 1-2P^{-1} \\ &= (\lambda^2-1)/(\lambda^2+1). \end{aligned} \tag{3.23}$$

To zeroth order in α_0/α' and β/α', (3.23) reduces to

$$f_i(A^N B^{8-N})=1-\tfrac{1}{4}N. \tag{3.24}$$

The limit $\alpha' \rightarrow \infty$ thus corresponds to the picture of "neutral bonding," in which each atom A has about it just enough electrons N to neutralize the core charge. The reader may note, to avoid confusion, that (2.6) and (3.23) imply $\lambda>1$. By redefining λ in the trial function (3.7), one could have had $\lambda<1$. Then the sign of (3.23) should be reversed. The actual values of λ quoted by CRS are less than one.

4. POLARIZABILITIES IN THE ONE-ELECTRON APPROXIMATION

The electronic polarizability tensor α_{ij} of a crystal or molecule can be obtained by perturbation theory. Let Ψ denote the complete many-electron wave function, and $\mathcal{H}_0$ the unperturbed Hamiltonian with fixed nuclear coordinates. The perturbation is

$$\mathcal{H}'=\boldsymbol{\mu}\cdot\mathbf{F}, \tag{4.1}$$

where $\mathbf{F}$ is the electric field, and the total dipole operator for N electrons is

$$\boldsymbol{\mu}=e\sum_{i=1}^{N}\mathbf{r}_i. \tag{4.2}$$

The Schrödinger equation

$$(\mathcal{H}_0+\mathcal{H}')\Psi=E\Psi \tag{4.3}$$

is solved by perturbation theory to order F^2. If the system under study has no permanent dipole moment,

$$E=E_0+F_iF_jE_{ij}'', \tag{4.4}$$

and the second-order term can be transformed to principal axes. Then one finds

$$\alpha_{ii}=-2E_{ii}''. \tag{4.5}$$

While the foregoing prescription is quite general and is applicable to both small molecules and crystals, it is not the one actually used for crystals. Dielectric properties of crystals are generally treated by the random-phase approximation (RPA). The application of the RPA to calculating the dielectric response has been shown to be equivalent to the Hartree approximation (Ehrenreich and Cohen, 1959).

For small molecules the Hartree approximation generally gives poor results for E_0, so its use in calculating E'' is suspect. However, it has been shown (Bell and Long, 1950) that even for the hydrogen molecule H_2 the Hartree approximation gives better results than any other one-parameter wave function. Thus one does not go far wrong in making the one-electron approximation.

The situation is even more favorable in crystals than in molecules. In crystals the electrons are free to move in three dimensions, rather than in two (as in planar molecules) or one (as in long-chain molecules). The Coulomb interaction is more singular as $r \rightarrow 0$ in effectively one-dimensional situations than it is in three-dimensional ones. Moreover valence electron densities are higher in crystals than in molecules, with the result that the screening of electron–electron interactions is more complete. Indeed breakdown of the one-electron approximation in crystals is indicated by the presence of exciton lines (bound states of electron–hole pairs) in the spectrum. Elliott (1957) has shown that the oscillator strength of such states is proportional to ϵ_0^{-3}, where ϵ_0 is the low-frequency dielectric constant. For this reason exciton oscillator strengths are negligible in covalent crystals, and probably do not represent more than 10% of the total oscillator strength even in crystals of the NaCl type.

The complex dielectric function $\epsilon_1(\omega)+i\epsilon_2(\omega)$ has two parts, the absorptive $\epsilon_2(\omega)$ and the dispersive $\epsilon_1(\omega)$. These satisfy the Kramers–Kronig relation (Landau and Lifshitz, 1960)

$$\epsilon_1(\omega)=1+\pi^{-1}\int_{-\infty}^{\infty}\frac{\epsilon_2(\omega')}{\omega'-\omega}\,d\omega', \tag{4.6}$$

and are related to the complex index of refraction $n-ik$ by

$$\epsilon_1=n^2-k^2, \tag{4.7}$$

$$\epsilon_2=2nk. \tag{4.8}$$

Thus it is sufficient to specify $\epsilon_2(\omega)$. In the RPA this is given by (Phillips, 1966)

$$\epsilon_2(\omega)=\frac{e^2h^2}{m}\sum_{j,j'}{}_d\Omega^{-1}\int\frac{f_{jj'}(\mathbf{k})}{E_{jj'}\,|\,\nabla_{\mathrm{k}}E_{jj'}\,|}\,dS_{\mathrm{k}}, \tag{4.9}$$

where Ω^{-1} is the volume of the unit cell, j and j' label occupied and unoccupied valence bands, **k** is the crystal momentum, and the integral is taken over the surface S_{k} in momentum space such that

$$\hbar\omega=E_j(\mathbf{k})-E_{j'}(\mathbf{k}). \tag{4.10}$$

The factor $f_{jj'}(\mathbf{k})$ is an interband oscillator strength defined in the dipole approximation by

$$f_{jj'}(\mathbf{k})=(2/3m)\{|\,\langle\mathbf{k}_j\,|\,\mathbf{p}\,|\,\mathbf{k}_{j'}\rangle\,|^2/[E_{j'}(\mathbf{k})-E_j(\mathbf{k})]\}. \tag{4.11}$$

In addition to the Kramers–Kronig relation (4.6), there is a second integral relation (Nozières and Pines, 1959) which is analogous to the f-sum rule for atoms. This may be written

$$\pi^{-1}\int_0^{\infty}\omega\epsilon_2(\omega)\,d\omega=\tfrac{1}{2}\omega_p^2, \tag{4.12}$$

where ω_p is the plasma frequency given by

$$\omega_p^2=4\pi Ne^2/m, \tag{4.13}$$

and N is the number of electrons per unit volume. In crystals not containing d or f electron atomic subshells, the absorption associated with valence electrons has been nearly exhausted at a cutoff frequency ω_c which is smaller than the frequency needed to excite core electrons. One may then replace (4.12) by

$$\pi^{-1}\int_0^{\omega_c}\omega\epsilon_2(\omega)\,d\omega=\tfrac{1}{2}\omega_{p'}^2, \tag{4.14}$$

where $\omega_{p'}$ is the plasma frequency defined by (4.13), but N measures the valence electron density only. More generally, one may define (Philipp and Ehrenreich, 1963)

$$N_{\mathrm{eff}}(\omega)=\frac{2}{\pi\omega_{p'}^2}\int_0^{\omega}\omega'\epsilon_2(\omega')\,d\omega'{}_d \tag{4.15}$$

as the fraction of valence electron oscillator strength exhausted up to ω.

The detailed implications of (4.9)–(4.14) have been analyzed extensively (Philipp and Ehrenreich, 1963; Phillips, 1966) for A^NB^{8-N} crystals. Many of these details can be explained using pseudopotentials V_A and V_B which in turn can be related to the spectra of free atoms. For the present, however, we propose to reduce the complicated bands in **k** space to simple bonding and antibonding orbitals $\varphi_A+\lambda\varphi_B$ and $\lambda\varphi_A-\varphi_B$ of the type discussed in the preceding chapter. We do this, however, not by appealing to the spectra of free atoms, but by examining the spectra of the crystals themselves. This removes all questions concerning charge effects and enables us to focus directly on observable quantities such as $\epsilon_1(\omega)$ and $\epsilon_2(\omega)$.

5. DISPERSION THEORY OF A^NB^{8-N} CRYSTALS

For a diatomic molecule, the form of the model Hamiltonian (3.3)–(3.5) can be made plausible by reference to atomic orbitals φ_A and φ_B. Here we wish to use a similar approach to justify using this Hamiltonian—or one closely analogous to it—for binary A^NB^{8-N} crystals. These crystals include the diamond, zincblende, wurtzite, and rocksalt types. All except the wurtzite have two atoms per unit cell, while the wurtzite has four atoms per unit cell. The c/a ratios in all wurtzite crystals are close to the ideal value 1.633, which corresponds to tetrahedral coordination.

If we describe any physical property as the superposition of properties of AB_4 and A_4B tetrahedral units, i.e., neglect interactions beyond those with nearest neighbors, then one can show (Robinson, 1968) that all tensorial properties of wurtzite crystals can be expressed in terms of those of zincblende crystals. Thus we are justified in constructing models for two atoms per unit cell.

We choose our origin of coordinates in a symmetrical way, halfway between the two atoms in the unit cell. The crystal potential is periodic, and hence can be represented as a Fourier series

$$V(\mathbf{r}) = \sum_{G} V_{\mathrm{G}} \exp(i\mathbf{G}\cdot\mathbf{r}), \tag{5.1}$$

where $\mathbf{G}$ is a reciprocal lattice vector. In order to avoid problems connected with cases where the cores of A and B are not isoelectronic, it is convenient to assume that each V_{G} is obtained from the pseudopotentials $V_{\mathrm{A}}(\mathbf{r})$ and $V_{\mathrm{B}}(\mathbf{r})$ of each atom as follows:

$$V_{\mathrm{G}} = V_{\mathrm{A}}(\mathbf{G}) \exp(i\mathbf{G}\cdot\mathbf{R}_{\mathrm{A}}) + V_{\mathrm{B}}(\mathbf{G}) \exp(i\mathbf{G}^{-}\mathbf{R}_{\mathrm{B}}), \tag{5.2}$$

$$V_{\mathrm{A}}(\mathbf{G}) = \Omega^{-1}\int V_{\mathrm{A}}(\mathbf{r}) \exp(-i\mathbf{G}\cdot\mathbf{r})\, d\tau. \tag{5.3}$$

One can now separate V into its even $(V_{\mathrm{A}}+V_{\mathrm{B}})$ and odd $(V_{\mathrm{A}}-V_{\mathrm{B}})$ parts, and write (5.2) as

$$\begin{aligned} V0 = {} & [V_{\mathrm{A}}(\mathbf{G}) + V_{\mathrm{B}}(\mathbf{G})] \cos \mathbf{G}\cdot\boldsymbol{\tau} \\ & + i[V_{\mathrm{A}}(\mathbf{G}) - V_{\mathrm{B}}(\mathbf{G})] \sin \mathbf{G}\cdot\boldsymbol{\tau}, \end{aligned} \tag{5.4}$$

where $\boldsymbol{\tau}$ is defined by

$$2\boldsymbol{\tau} = \mathbf{R}_{\mathrm{A}} - \mathbf{R}_{\mathrm{B}}.$$

In the limit of $V_{\mathrm{A}} = 0 = V_{\mathrm{B}}$, the bonding state $\varphi_{\mathrm{A}} + \lambda\varphi_{\mathrm{B}}$ and the antibonding state $\lambda\varphi_{\mathrm{A}} - \varphi_{\mathrm{B}}$ will have the same energy, which we will take to be zero. This gives

$$H_{11} = H_{22} = 0, \tag{5.5}$$

while the form taken by (5.4) between real basis functions is just (Phillips, 1968)

$$H_{12} = H_{21}^{*} = \tfrac{1}{2}(E_h + iC), \tag{5.6}$$

where, to first order in V, E_h arises from the real part of (5.4), and iC arises from the imaginary part. One can bring (5.5) and (5.6) into the same form as (3.3)–(3.5) by partial diagonalization to give

$$H_{11}' = -C/2, \tag{5.7}$$

$$H_{22}' = C/2, \tag{5.8}$$

$$H_{12}' = H_{21} = E_h/2. \tag{5.9}$$

Thus C corresponds to $\alpha_{\mathrm{A}} - \alpha_{\mathrm{B}}$, and E_h corresponds to 2β. Interchange of A and B reverses the sign of C, but leaves E_h unchanged.

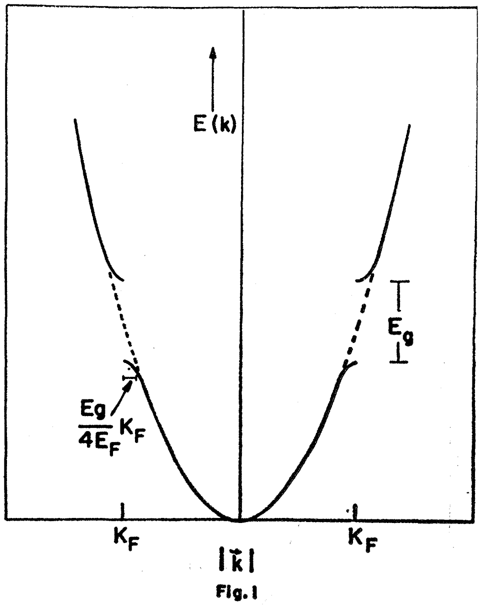


FIG. 1. The energy bands of isotropic model semiconductor which reduce in the limit $E_g \to 0$ to the energy bands of a free electron gas.

One advantage of writing the Hamiltonian in the form (5.5)–(5.6) is that the energy gap E_g between the bonding and the antibonding states is obviously given by

$$E_g^2 = E_h^2 + C^2. \tag{5.10}$$

Thus E_h is equal to E_g in the case of a purely covalent Group IV crystal such as diamond or Si. The ionic or charge-transfer contribution to E_g is represented by C. The symmetry between E_h and C which is present in (5.6) and (5.10) will be retained so far as possible in the following discussion.

To determine E_h and C spectroscopically, let us consider the diamond-type crystals where $C = 0$. A simple model for diamond-type semiconductors has been proposed by Penn (1962; Van Vechten, 1969).

The energy bands in Penn's model are shown in Fig. 1. The complicated energy bands of the real crystal, which are usually shown as multivalued functions of $\mathbf{k}$ in the reduced Brillouin zone (Phillips, 1966), are represented instead by isotropic bands $E(k)$ derived from the model Hamiltonian

$$H_{11} = \hbar^2 k^2/2m, \tag{5.11}$$

$$H_{22} = \hbar^2(2k_f - k)^2/2m, \tag{5.12}$$

$$H_{12} = E_g/2, \tag{5.13}$$

where k_f is the Fermi wave number of a free electron gas with density equal to that of the valence electrons. The motivation for introducing kinetic energy into the Hamiltonian in (5.11) and (5.12) is that the energy

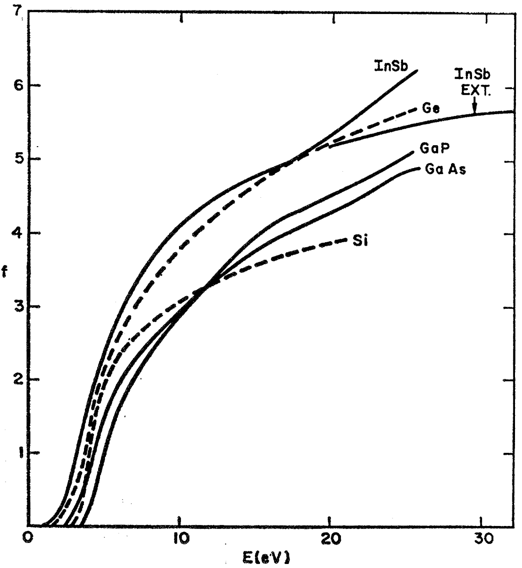


FIG. 2. A plot of $n_{\text{eff}}(\omega)=4N_{\text{eff}}(\omega)$ as defined by Eq. (4.15). Note that for Si (no d electrons) it appears that $n_{\text{eff}}(\omega)$ will level off near 4 at large ω, whereas for the remaining crystals (which contain d electrons in some or all of the atomic cores) $n_{\text{eff}}(\omega)>4$ for large ω. The extrapolated curve for InSb is leveling off at $4D$, where D is the parameter defined in the text.

bands of diamond-type semiconductors do resemble those of a nearly free electron gas. As can be seen from Fig. 1, in Penn's model the splitting of the valence and conduction bands is important only in the shell

$$k_f-k\lesssim(E_g/4E_f)k_f\lesssim k-k_f \tag{5.14}$$

about $k=k_f$. Here bandwidth effects are presented by

$$E_f=\hbar^2k_f^2/2m, \tag{5.15}$$

and typical values of $E_g/4E_f$ are 0.1.

The energy gap E_g in Penn's model is to be evaluated from $\epsilon_1(0)$, the low-frequency limit of the real part of the electronic dielectric constant. Assuming that there are no important contributions to $\epsilon_1(0)$ from polarization of d or f subshells in the atomic cores, his model gives, to lowest order in E_g/E_f, the relation

$$\epsilon_1(0)=1+(\hbar\omega_p/E_g)^2(1-E_g/4E_f), \tag{5.16}$$

from which it can be seen that bandwidth effects are small. One can apply (5.16) to diamond and Si to obtain

$$E_h(\text{diamond})=13.6\text{ eV}, \tag{5.17}$$

$$E_h(\text{Si})=4.8\text{ eV}. \tag{5.18}$$

These values differ drastically from the values of 2β in the valence bond theory [Eq. (3.6)]. The reason for the differences will be discussed later.

We now wish to extend the analysis to obtain energy gaps in the heavier elements Ge and (grey) Sn. In Rows 3 and 4 of the Periodic Table, each atom contains either a $3d$ subshell or a $4d$ subshell. The binding energies of these subshells are only a few rydbergs or less, compared to nearly 10 Ry or more for the last s–p core subshells. As a result there is appreciable mixing of valence and conduction band levels with d levels. This mixing is quite difficult to include in molecular orbital theory without sacrificing the simplicity of the wave function (3.7), and as it varies with crystal momentum $\mathbf{k}$, it is doubtful if a simple bonding–antibonding two-level model is appropriate in any event.

Most of these difficulties are resolved by the use of the dispersion approach. There are two effects associated with core behavior which must be considered. The first of these is the direct contribution of $d(\text{core})\rightarrow p$ conduction band transitions to $\epsilon_2(\omega)$. These are reflected by a rise in $N_{\text{eff}}(\omega)$ as defined by (4.14), and produce a "knee" in $N_{\text{eff}}(\omega)$ at the threshold (Philipp and Ehrenreich, 1963). An interesting feature of the presence of d states is that extrapolation of $N_{\text{eff}}(\omega)$ below this knee gives a limit greater than 1 in InSb, although the same extrapolation for Si gives about 1, as it should. This second effect is shown in Fig. 2. As explained by Van Vechten (1963), because of the mixing of valence states with d states, the oscillator strength of valence-conduction band transitions is increased above expectations even at frequencies below the threshold for real $d(\text{core})\rightarrow p$ conduction band transitions.

Another way of describing this effect is to note the qualitative change in $\epsilon_2(\omega)$ which occurs between diamond and Si on the one hand, and Ge and Sn on the

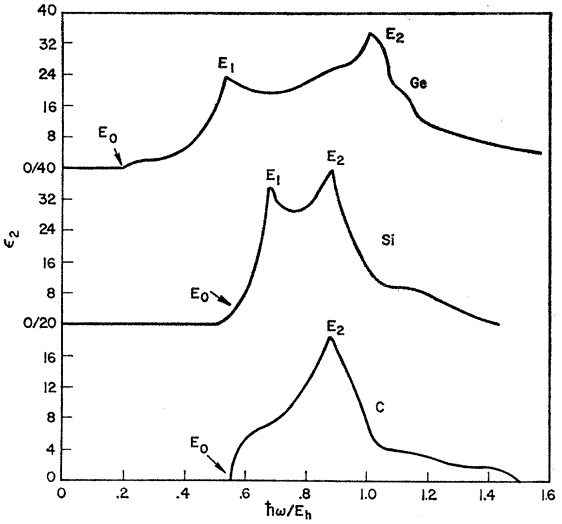


FIG. 3. The broad features of $\epsilon_2(\omega)$ for Group IV crystals plotted in reduced units. The purpose of the figure is to show that the one-gap approximation is relatively good for diamond, but that a low-energy peak near $\hbar\omega/E_h=0.5$ or 0.6 develops in Si, becomes stronger in Ge, and becomes still stronger in grey Sn (not shown).

other. This change is illustrated in Fig. 3, which shows that although the one-peak approximation is good for diamond, a lower energy peak develops in Si which becomes quite prominent in Ge and still more so in grey Sn (not shown).

Detailed analysis (Phillips, 1966) of the energy bands of three crystals shows that the energy levels in the valence bands of these crystals are all quite similar, but that there is substantial lowering of s-like conduction or antibonding levels from Si to Ge, and further lowering in grey Sn. The mechanism responsible for the lowering from Si to Ge is partial penetration of the $3d$ core subshell by $4s$ electrons to see the larger core attractive potential. The further lowering on going from Ge to Sn is a relativistic effect, as shown in Fig. 4 based on atomic calculations (Herman and Skillman, 1963).

Neither of these core effects contributes significantly to the bonding–antibonding energy gap E_g. It would be a mistake to include them in the relation (5.16), because we wish E_g to describe valence bond strengths, not peripheral core properties. The simplest way to circumvent this problem is to assume (Phillips, 1968b) that E_h scales with the cubic lattice constant a according to some power s,

$$d \log E_h / d \log a = s, \tag{5.19}$$

and to determine s from the values of E_h in diamond and Si. This gives (Van Vechten, 1969; Penn, 1962) $s=-2.5$, and from (5.19) and the Si value one can compute E_h for any tetrahedrally coordinated $A^N B^{8-N}$ crystal.

In particular one can use the extrapolated values of E_h to modify (5.15). Each value of ω_p^2 is multiplied by a correction factor $D>1$ for compounds containing Row-3 or Row-4 atoms. This corresponds to $N_{\mathrm{eff}}(\omega)>1$, as shown in Fig. 3. For the details of one way of determining D, see Van Vechten (1969). His method is not unique, but it is consistent with the physical mechanisms that have been discussed, and it has yielded excellent results for the chemical trends within the family of $A^N B^{8-N}$ crystals. Moreover, it has been checked against experimental data on the hydrostatic pressure dependence (Van Vechten, 1969) of $\epsilon_1(0)$ and against the optical spectra of alloys of these crystals among themselves (Mott and Jones, 1958) (e.g., $\mathrm{GaAs}_x\mathrm{P}_{1-x}$). In both cases the theoretical values are quite sensitive to the way in which the correction factor D for oscillator strengths is introduced, and in both cases the agreement with experiment is excellent. Prior to the introduction of this spectroscopic bond model, there had been no theory of either of these effects, apart from the classical Clausius–Mosotti model, which gives the wrong sign (Van Vechten, 1969) for the pressure dependence $d \log \epsilon_1(0)/d \log P$.

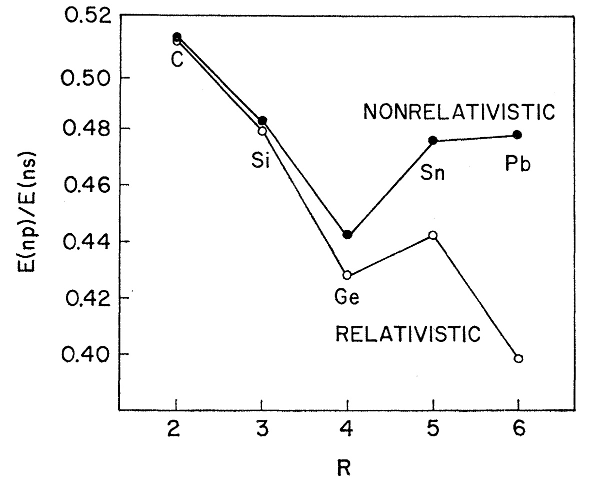


FIG. 4. Increasing value of $(E_s-E_p)/E_s$, or decreasing value of E_p/E_s, as one goes from carbon to lead. Atomic term values taken from Herman and Skillman (1963).

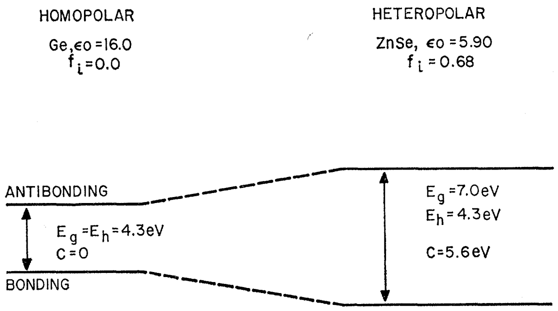


FIG. 5. A sketch of the effect of increasing valence difference $\Delta Z=8-2N$ in isoelectronic $A^N B^{8-N}$ atom pairs on the average energy gap E_g between bonding and antibonding states. The relevant parameters for Ge and ZnSe are indicated.

Knowing $E_h(a)$ and D, one can use (5.16) to evaluate C. Typical values of this ionic energy are shown in Fig. 5. Other values are $C(\mathrm{BN})=7.8$ eV, $C(\mathrm{BeO})=14.1$ eV, $C(\mathrm{GaAs})=2.9$ eV, $C(\mathrm{ZnSe})=5.6$ eV. Because the bond lengths in BeO and BN are nearly the same, and those of GaAs and ZnSe are nearly the same, we see immediately that C is proportional to the valence difference, ΔZ, which is $8-2N$ in $A^N B^{8-N}$ compounds. It is to be stressed that these values are derived not from atomic spectra but from the crystalline spectra themselves, as contained in $\epsilon_1(0)$.

With spectroscopic values of $C(A^N B^{8-N})$ available for nearly 40 tetrahedrally coordinated crystals, one can seek an analytic formula to describe the observed values. The following formula works well:

$$C(\mathrm{AB}) = b[(Z_A/r_A) - (Z_B/r_B)]e^{-k_s R}. \tag{5.20}$$

Here Z_A and Z_B are the valence numbers N and $8-N$ of the cation and anion, respectively. The atomic radii r_A and r_B are defined as half the bond length of the Group IV element belonging to the same rows of the

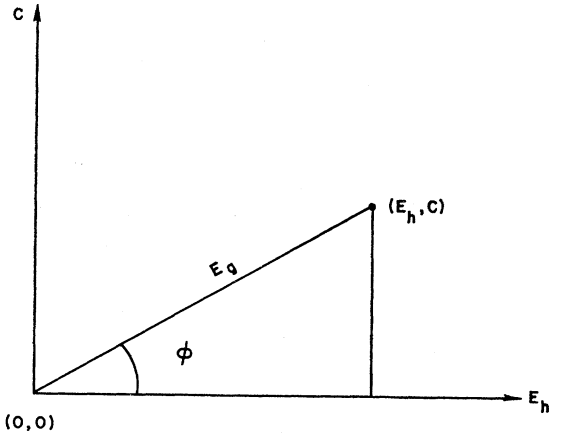


FIG. 6. A sketch showing how (E_g, φ) are polar coordinates corresponding to the covalent and ionic energies represented as Cartesian coordinates (E_h, C). The polar angle φ is called the *ionicity phase angle.*

Periodic Table as atoms A and B, respectively. Thus the atomic radius of B is half the nearest-neighbor distance in diamond, for example. In (5.20) the electronic charge $e=1$, so that b is a dimensionless number. It is of order unity because of the presence of the Thomas–Fermi exponential screening factor (Mott and Jones, 1958) expressed in terms of

$$R=(r_A+r_B)/2, \tag{5.21}$$

and the Thomas–Fermi screening wave number k_s defined by

$$k_s^2=4k_f/\pi a_0. \tag{5.22}$$

Here k_f is the Fermi momentum of a free electron gas of density equal to that of the valence electrons, and a_0 is the Bohr radius. In the absence of the Thomas–Fermi screening factor $C(AB)$ would equal (Phillips, 1968b) X_A-X_B as tabulated by Pauling (1960). The screening factor varies by about 40% from diamond to grey Sn. One can easily recognize it in the ratio $C(BN)/C(InSb)$, or in $C(BeO)/C(CdTe)$.

Because so many values of $C(AB)$ are known, it is instructive to make a histogram plot (Van Vechten, 1969; Penn, 1962) of the prefactor b in (5.20). Almost all the experimental values for tetrahedrally coordinated crystals cluster in the interval

$$1.4<b<1.6. \tag{5.23}$$

In most cases therefore $b=1.5$ to within an accuracy of 10% or better, the remaining scatter is probably caused by variations in core properties which are not wholly represented by r_A and r_B. These variations are quite small, and in any event are reduced in significance by using empirical values for C rather than the theoretically based formula (5.20).

Because of its similarity to X_A-X_B, one may refer to $C(AB)$ as a spectroscopically defined electronegativity difference. There is an important difference between the two, however, which is emphasized by (5.20). Because of the exponential screening factor which multiplies both terms and depends on r_A and r_B through (5.21), $C(AB)$ cannot be represented as a difference of two terms, one of which depends only on A, and one only on B. This is because screening depends in a nonseparable way on the bond length r_A+r_B.

We are now prepared to give a spectroscopic definition of ionicity. Both (5.6) and (5.10) suggest that we regard E_h and C as Cartesian coordinates, and transform to more symmetrical polar coordinates. Then E_g is the polar radius, and we call φ the *ionicity phase angle,*

$$\tan\varphi=C/E_h, \tag{5.24}$$

where these coordinates are shown in Fig. 6. We want f_i to satisfy (2.5) and (2.6), and also to be zero when $C=0$. This suggests the definition

$$f_i=\sin^2\varphi=C^2/E_g^2. \tag{5.25}$$

The definition (5.25) is not unique; one could use $f_i=g(\sin^2\varphi)$, with $g(0)=0$, $g(1)=1$, and $dg/dx>0$ for $0\le x\le 1$. Certainly (5.25) is the simplest choice, and it corresponds to the idea that a *fraction* of the bond is ionic. The covalent fraction f_c is simply

$$f_c=\cos^2\varphi=E_h^2/E_g^2, \tag{5.26}$$

and the symmetry of f_i and f_c is appealing. We shall demonstrate presently that (5.25) is the proper choice with $g(x)=x$.

6. GLOSSARY OF THERMOCHEMICAL, MOLECULAR ORBITAL, PSEUDOPOTENTIAL AND SPECTROSCOPIC TERMINOLOGIES

It is convenient at this point, before entering into detailed comparisons of theories of A^NB^{8-N} crystalline bonds, to summarize the different approaches to this

TABLE I. Notation describing theories of covalent and ionic contributions to A^NB^{8-N} crystalline bonds.

Kind of bond / Theory	Covalent	Ionic
Thermochemical (Pauling)	$(D_{AA}+D_{BB})/2$	$\lvert X_A-X_B\rvert \propto \Delta_{AB}^{1/2}$
Valence bond—MO (Coulson)	β_{IV}	$\alpha_A-\alpha_B$
Pseudopotential	V_A+V_B	V_A-V_B
Dispersion theory	$E_h(r_{AB})$	$C(AB)$

TABLE II. Experimental origins of parameters of theories.

Kind of bond Theory	Covalent	Ionic
Thermochemical (Pauling)	Heats of formation of elements	Extra-ionic or heteropolar energy
Valence bond–MO	Heats of formation of homopolar IV crystals	Atomic spectra
Pseudopotential	Three parameters fitted to interband energies	Three parameters fitted to interband energies
Dispersion theory	Dielectric constants of homopolar IV crystals	Dielectric constants of heteropolar crystal

question. In Table I the four theories are compared in terms of notation. In Table II the experimental sources used to determine the parameters of the theories are indicated.

There are some interesting aspects to Table II. First, we note that Pauling employs the heats of formation of the elements A and B. If $N\neq 4$, then these elements will generally have different structures and different coordination numbers than the A^NB^{8-N} crystals. From Pauling's viewpoint this is no obstacle, since as long as all the bonds involved are normal single bonds involving a single pair of electrons, the fact that the coordination numbers are greater than the respective valences is explained in terms of resonating bonds.

The remaining three theories take a quantum-mechanical approach in which the potential energy is separated into its even and odd parts, as indicated most clearly for the pseudopotential in Table I. However, Coulson's theory determines these two terms in quite different ways. This leads to serious errors, as we shall see presently.

The pseudopotential treats both energies on an even footing and evaluates them spectroscopically. From a chemical point of view, however, the pseudopotential approach is ambiguous because $V_A(\mathbf{G})\pm V_B(\mathbf{G})$ must be evaluated at three different inequivalent nonzero reciprocal lattice vectors $\mathbf{G}$. This is accomplished by adjusting the shape of the spectrum to fit three or more peaks in the optical spectra $\epsilon_2(\omega)$. (See Fig. 3, for example.) Such an adjustment is necessarily subjective and corresponds to weighting certain portions of the spectrum much more heavily than others. (Mathematically speaking, the weighting function is singular, consisting of δ functions at the peak energies which have been selected for fitting.) The Kramers–Kronig transform (4.6), on the other hand, uses a much more reasonable weighting function. The use of $\epsilon_1(0)$ corresponds to putting $\omega=0$ in the weighting function $(\omega'-\omega)^{-1}$ in the K–K transform. This has the effect of treating both right- and left-hand polarized light on the same footing. The choice $\omega=0$ corresponds to zero exchange splitting, and zero is the only value of ω consistent with nonmagnetic bonds.

We now turn to definitions of ionicity. These are summarized in Table III. In the Pauling theory the ionicity of the chemical bond is determined only by extra-ionic energies, covalent energies not being used to define f_i. In the Coulson theory, f_i is determined to zeroth order by the condition (3.24) of charge neutrality, and to first order by electronegativity differences in charged and uncharged states, the degree of charging being determined self-consistently. The values of β used are so small that covalent effects are an order of magnitude smaller than ionic ones. Finally the dispersion theory determines E_h and C symmetrically, and these enter the definition of ionicity on an equal footing. This symmetry plays a key role in the theory.

TABLE III. Nature of definitions of ionicity.

Kind of bond Theory	Covalent	Ionic
Thermochemical	Not included	Depends on Δ_{AB} only
Valence bond–MO	Included in second order only	$(\alpha_A-\alpha_B)/(\alpha_A'+\alpha_B')$
Dispersion theory	Depends on *both* C and E_h	

Table IV. Comparison of even terms in $A^N B^{8-N}$ crystal energies for $N=4$. The second column is based on cohesive or resonance energies; the third column on spectroscopic dispersion theory. All energies are in electron volts.

Crystal	2β	E_h	$2\beta/E_h$
Diamond	3.7	13.6	0.27
Si	2.0	4.8	0.42
Ge	1.5	4.3	0.35
Sn	1.5	3.1	0.48

7. COMPARISON OF MOLECULAR ORBITAL AND SPECTROSCOPIC ENERGIES

To understand further the differences between the CRS valence bond and molecular orbital model and spectroscopic theory, it is helpful to consider the numerical values for 2β and $\alpha_A - \alpha_B$ compared to those for E_h and C_1, respectively. These are listed for a few representative examples in Tables IV and V.

From Table IV, we see that $2\beta \ll E_h$, i.e., the resonance or thermochemical estimate of covalent energies yields a much smaller value than the spectroscopic estimate. This is not surprising to those familiar with the literature of the Hückel theory of hydrocarbons. According to Streitwieser (1961), resonance estimates of β_{CC} for π bonds in benzenoid hydrocarbons have often been placed at 0.7–0.8 eV, compared to spectroscopic values of about 2.5 eV. Thus one would *expect* the CRS resonance value to be 0.3–0.4 times the spectroscopic value, as it is here for sp^3 crystalline bonds.

Turning now to the odd part of the bond energy shown in Table V, we must distinguish two sets of MO values. The first set is represented by $\alpha_A(0) - \alpha_B(0)$; it is a measure of the difference in ionization energies of the free atoms. As recognized by CRS this is a gross overestimate of the ionic energy in the bonded state, because charge will flow from one atom to the other to partially neutralize the electronegativity difference. The corrected values are represented by $\alpha_A(Q) - \alpha_B(-Q)$, and as one can see from Table V, these are less than 10% of the initial values. The comparable spectroscopic values are also shown in Table V, and they fall between the atomic and self-consistent MO values.

From Table V one can see that CRS have grossly overestimated the degree of neutralization of the ionic potential by electron redistribution. The reason for this is that if β were zero, there would be no covalent energy to oppose the charge flow, and one would find, apart from rounding errors, $\alpha_A(Q) - \alpha_B(-Q) = 0$. As we have seen in discussing Table IV, the actual covalent energies 2β used by CRS are about three times smaller than the spectroscopic values. The similarity of the model Hamiltonians, together with the self-consistency condition used by CRS to determine Q, leads to the approximate result

$$[\alpha_A(Q) - \alpha_B(-Q)]/2\beta \approx C_{AB}/E_h, \qquad (7.1)$$

which means that the values in the third column of Table V for first-row compounds can be approximately obtained by multiplying the values from the fourth column by the value of $2\beta/E_h$ shown in Table IV for diamond; similarly the third-row compounds use the Ge ratio.

The foregoing comparison shows that the numerical values of the CRS theory would be at least approximately consistent with those of the spectroscopic theory if β were determined spectroscopically rather than thermochemically. (Determining α spectroscopically and β thermochemically is, metaphorically speaking, like mixing oil and water.) However, even if this were done, there would still be the substantial difference between $[\alpha_A(0) - \alpha_B(0)]$ and C_{AB} which would have to be removed by the self-consistency procedure. This difference, which is about a factor of 2 for the first-row compounds, and a factor of 3 for the third-row ones, is a very large correction to build into a theory. It almost necessarily leads to washing out of the very chemical trends which the theory was constructed to explain in the first place. It follows that the dispersion approach is likely to be more accurate because it determines these parameters correctly at the outset. However, in some contexts comparison with atomic energies might still be useful.

8. COMPARISON OF THERMOCHEMICAL AND SPECTROSCOPIC PARAMETERS

We saw in Sec. 6 that the ionic energy C_{AB} is analogous to Pauling's electronegativity difference $X_A - X_B$. A simple formula for the electronegativities X_A in Pauling's table (Pauling, 1960) which holds quite well for

Table V. Comparison of odd terms in $A^N B^{8-N}$ crystal energies for selected examples. The second column gives the difference in A and B energies without charge transfer, according to molecular orbital theory. The third column gives the difference after allowing for charge transfer. The fourth column lists the comparable spectroscopic values. All energies are in electron volts.

Crystal	$\alpha_A(0) - \alpha_B(0)$	$\alpha_A(Q) - \alpha_B(-Q)$	C_{AB}
Diamond	0	0	0
BN	16.5	1.5	7.8
BeO	26.3	3.3	14.1
Ge	0	0	0
GaAs	11.6	0.5	2.9
ZnSe	16.0	1.2	5.6

elements from Groups IIb–VII of the Periodic Table is

$$X_A - 0.5 \propto Z_A/r_A, \qquad (8.1)$$

where r_A is half the nearest-neighbor distance in the diamond-type crystal corresponding to the row of the Periodic Table to which element A belongs. The electronegativity defined by (8.1) refers to hybridized *s–p* bonds, and does not describe electronegativities of transition or lanthanide series elements, or Ia, Ib, and IIa elements where *d* and *f* states are important.

The spectroscopic values for C_{AB} are similar to that for $X_A - X_B$ derived from (8.1), but include the density dependence of the Thomas–Fermi screening wave number k_s. Specifically the formula for C_{AB} which gives the best fit is (5.20) which is

$$C_{AB} = 1.5[(Z_A/r_A) - (Z_B/r_B)] \exp(-k_s R). \qquad (8.2)$$

For compounds all of whose atoms belong to a given row of the Periodic Table (e.g., Ge, GaAs, ZnSe) the valence electron density is sensibly constant, and $C_{AB} \propto [(Z_A/r_A) - (Z_B/r_B)] \propto (X_A - X_B)$. This suggests that we define (Phillips, 1968b) a dielectric electronegativity by

$$Y_A = 4.0(Z_A/r_A) \exp(-k_s R_A) + 0.5, \qquad (8.3)$$

which can be compared directly to Pauling's values providing that r_A is given in atomic units. Because both are proportional to Z_A, in Table VI we list only the values for the Group IV elements, from which the remaining values can be obtained immediately using (8.1) and (8.3). The similarity of the thermochemical and dielectric values is encouraging.

When the atoms A and B no longer belong to the same row of the Periodic Table, the decomposition (8.3) of Formula (8.2) is no longer possible. Moreover, in some of these cases, the experimental value of C_{AB} corresponds to a numerical factor in (8.2) which differs somewhat from 1.5. Therefore it is of interest to compare the tabulated values (Van Vechten, 1969; Penn, 1962) of C_{AB} with $X_A - X_B$. This is done in Fig. 7. There is considerable scatter, which is thought to arise chiefly from the absence of the Thomas–Fermi exponential screening factor from (8.1).

TABLE VI. Comparison of thermochemical and spectroscopically defined electronegativities for elements from Group IV of the Periodic Table (see Phillips, 1968b).

Element	Pauling	Dielectric
C	2.5	2.50
Si	1.8	1.41
Ge	1.8	1.35
Sn	···	1.15

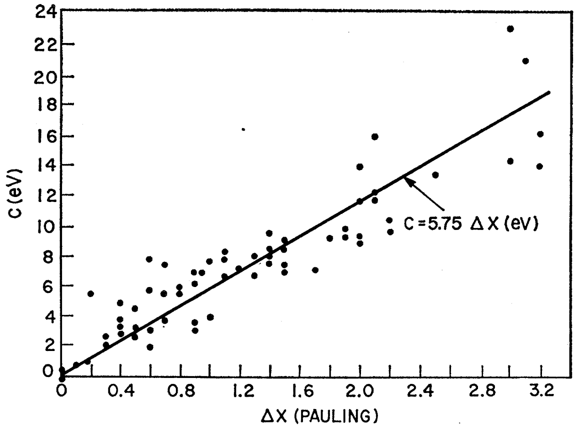


FIG. 7. Comparison of thermochemical electronegativity differences ΔX_{AB} [as defined by Pauling (1960)] with spectroscopically defined ionic energies C_{AB}, for $A^N B^{8-N}$ crystals.

Is the electronegativity difference by itself a meaningful quantity, or is it more useful to treat C and E_h simultaneously, in terms of the polar coordinates E_g and φ? We shall now examine a number of structural features of $A^N B^{8-N}$ crystals to see whether electronegativity (i.e., C) or ionicity (i.e., $\sin^2 \varphi$) plays the more important role in determining the observed chemical trends.

9. CRYSTAL STRUCTURES OF $A^N B^{8-N}$ COMPOUNDS

In the entire field of crystal physics there is one question which is without doubt *the* fundamental one. This is the question of why a given compound is formed in a definite crystal structure. This question may be broadened to include the larger one of the correlation of crystal structure and constitutive properties. Empirically it is well established that magnetism, superconductivity, ferroelectricity, and semiconductivity tend to occur in definite structures and to be absent in other structures. Many qualitative explanations for this behavior are known, but in practice the explanations are unconvincing because of the lack of quantitative predictions. Of course, before one considers the many different constitutive properties, the basic question of crystal structure should be understood.

Here we take up this problem from the point of view of chemical bonding in the $A^N B^{8-N}$ crystals which are either fourfold or sixfold coordinated. This is the simplest possible case, and in the language of directed valence bonds, the question has a simple qualitative answer. In the fourfold coordinated crystals, the *s–p* valence electrons are fully hybridized and form $s^1 + p^3 = (sp^3)$ (Streitweiser, 1961) directed valence orbitals. In the sixfold coordinated crystals, the atoms are much closer to the ionic or closed-shell limit, and interatomic interactions not of an electrostatic type are much weaker. As a result, the s and p^3 bonds do *not* hybridize, and the dominant valence orbitals are the p_x, p_y, and p_z valence orbitals. (In the language of energy-band theory one would say that the bandwidths of the s and p bands become, in the ionic limit, small compared to the anion $s-p$ energy difference.) Thus in the ionic limit the nearest neighbors are symmetrically distributed along the Cartesian coordinate axes, and the atom is sixfold coordinated. (The unhybridized p orbitals do not distinguish between $\pm x$, etc.)

This is a good explanation of the facts so far as it goes. However, if one inspects the energy bands of many Group II–VI semiconductors (Cohen and Bergstresser, 1966; Walter and Cohen, 1969) one finds that the s and p bands are already quite well separated even though the atoms are still fourfold coordinated. The residual s–p hybridization must be great enough to produce the fourfold coordinated structure, but the presence of this residual hybridization is difficult to measure or to identify either in the energy levels or in the wave functions.

The straightforward approach to the problem of predicting crystal structures is the following: Starting from the wave equation one calculates electronic energy levels and wave functions in the one-electron approximation; from these one calculates a crystal charge density and produces a new crystal potential; the process is iterated until it converges. Many-body exchange and correlation interactions are added through some approximate method. Then the total energy is calculated, and the lattice constant is varied to minimize the ground state energy. Finally the entire procedure is repeated for alternate lattice structures, and the one with the lowest energy is the predicted structure.

This procedure has never been carried through to conclusion for any insulator. The closest approximation to it is the work on more ionic crystals which assumes that each ion has a closed-shell or rare-gas electronic configuration. In this case one is reasonably justified in discarding the quantum-mechanical kinetic, exchange, and correlation energies, and calculating the total energy following Born and Mayer by using classical electrostatics and a phenomenological core–core repulsive interaction. This classical model has been the subject of many review articles (e.g., Tosi, 1965). The calculations have led to the conclusion that theoretical predictions of the relative stability of fourfold and sixfold coordinated structures have been generally unsuccessful from a quantitative point of view, even when confined to Group I–VII crystals.

A more productive approach to this problem was taken by Mooser and Pearson (1959; Pearson, 1960) on purely phenomenological grounds. They observe that two factors must influence the tendency towards formation of directed bonds. The first is the obvious one of charge transfer between anion and cation, which they measure in terms of Pauling's electronegativity difference $X_A - X_B$. The second is the extent of dehybridization (or, as they term it, metallization) of directed s–p bonds through admixture of $d, f \cdots$ states. The metallization is weakest for elements from the C row (principal quantum number of valence electrons $n=2$) because of the large energy required to promote an electron from a $2s$ or $2p$ state to a $3d$ or $4f$ state. The degree of metallization increases as n increases, and is responsible for the decreasing stability of diamond-type crystals against transition to a metallic structure. Thus for $n=5$ one finds Sn in both metallic and diamond-type structures, and for $n=6$ one finds Pb only in metallic structures.

To exhibit the dependence of crystal structure on these two factors, one makes a Mooser–Pearson plot with abscissa $\bar{n}=(n_A+n_B)/2$. As a general rule, crystals lying in the upper right quadrant will then prefer the ionic structure, those in the lower left will prefer the covalent one. The Mooser–Pearson plot for $A^N B^{8-N}$ crystals is shown in Fig. 8. Also included in the plot are some tetrahedrally coordinated crystals of the type $A^{N-M}B^{N+M}C_2^{8-N}$, which can be regarded as a superlattice of the $A^N B^{8-N}$ zincblende structure.

As one can see from the figure, the domains of fourfold and sixfold coordination are almost disjoint. Actually there is a small overlap, with seven open circles (fourfold coordinated crystals) lying in the sixfold coordinated region, and one solid circle (sixfold coordinated crystal) lying in the fourfold coordinated region. This is a total of eight errors for over 100 crystals, and as we shall see, it is better than par for the course.

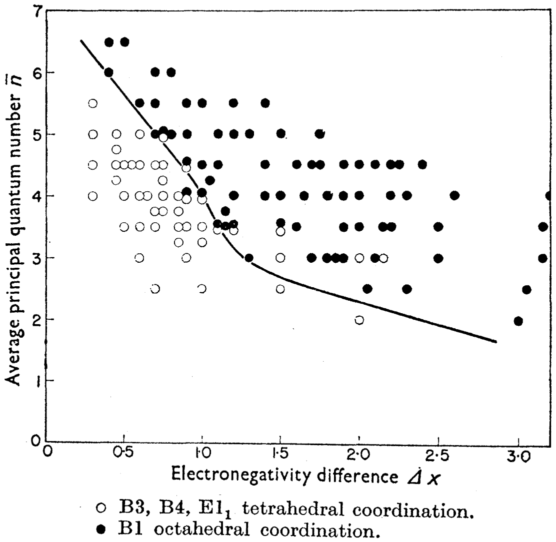


FIG. 8. The separation of fourfold and sixfold coordinated structures, according to Mooser and Pearson (1959). The abscissa is Pauling's electronegativity difference; the ordinate is the average principal quantum number of the valence electrons.

How difficult is the problem of predicting the coordination numbers of the family of A^NB^{8-N} crystals? All the nonmetallic crystals with $N=3$ or $N=4$ have fourfold coordination. A tableau showing the coordination numbers for $N=1$ and 2 is given in Fig. 9. If one simply uses the classification of the elements implicit in the Periodic Table itself, one would expect sixfold coordination when A is a Group Ia or IIa element, and fourfold coordination when A is a Group Ib or IIb element. This classification makes nine errors, and any theoretical model should do better. Actually the Periodic Table contains so much chemical information that a theoretical model which makes only nine errors is really doing quite well, even if it does not improve the situation appreciably.

By this standard the Mooser–Pearson plot is moderately successful, although it is clear that the curve separating the two coordination domains is equivalent to several free parameters. The real purpose of the Mooser–Pearson paper, however, was not to discuss the crystal structures of insulators but to discuss the structures of intermetallic compounds. Although this topic will not be discussed in this review, we remark here that it appears to us that the approach adopted by Mooser and Pearson has great potential for future research, especially on the structures of transition-metal compounds. We have in mind, of course, that the Mooser–Pearson analysis can be refined by methods analogous to those discussed here for A^NB^{8-N} compounds.

By now the reader will have recognized that ΔX is analogous to C, and that $\bar{n}$ plays the role of E_h. The correspondence for each pair of variables is not one to one, however:

(1) We have already seen in Fig. 7, and in the discussion of C and ΔX, that the values of ΔX given by Pauling are not an accurate measure of the ionic energy C.

(2) The situation is worse for E_h and $\bar{n}$. We have $E_h \propto a^{-2.5}$, while a is not a smooth function of $\bar{n}$. In fact, $a(n)$ changes by 50% from $n=2$ to $n=3$, by only 4% from $n=3$ to $n=4$, and by 15% from $n=4$ to $n=5$. This means that $\bar{n}$ for skew compounds behaves quite differently from a. Since lattice constants or bond lengths are observables while principal quantum numbers are not, there is every reason to believe that a is more fundamental than $\bar{n}$. For example, $\bar{n}(\mathrm{MgTe}) = \bar{n}(\mathrm{GaAs}) = \bar{n}(\mathrm{CdS})$, but the bond lengths are 2.76, 2.44, and 2.53 Å, respectively. This represents a range of more than 10%, or a range in E_h of more than 30%, without even including any atoms for which $n=2$.

	Ia				Ib	
	Li	Na	K	Rb	Cu	Ag
F	6	6	6	6	4	6
Cl	6	6	6	6	4	6
Br	6	6	6	6	4	6
I	6	6	6	6	4	4

	IIa			IIb				
	Ca	Sr	Ba	Be	Mg	Zn	Cd	Hg
O	6	6	6	4	6	4	6	6
S	6	6	6	4	6-4	4	4	6
Se	6	6	6	4	6-4	4	4	4
Te	6	6	6	4	4	4	4	4

FIG. 9. Coordination numbers (4 or 6 or in cases of metastable structures, both) of A^NB^{8-N} crystals with $N=1$ or 2. These structures are the ones that are stable or metastable at STP.

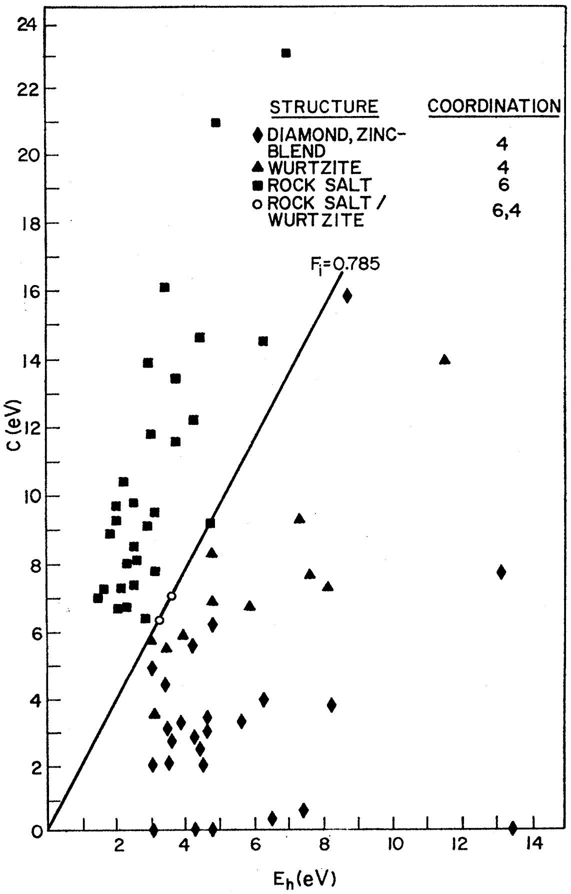


FIG. 10. The separation of fourfold and sixfold coordinates structures, using the spectroscopically defined covalent and ionic energy gaps E_h and C, respectively.

This suggests that if we were to use E_h and C as coordinates, rather than ΔX and $\bar{n}$, a better separation of crystal structures might be obtained. This is done in Fig. 10. The result is spectacular. The ***straight*** line passing through the origin $E_h=0=C$ exactly separates the fourfold and sixfold coordinated structures. There are *no* errors. The straight line corresponds to $\varphi=\varphi_c$ or to

$$F_i = \sin^2 \varphi_c = 0.785 \pm 0.010. \tag{9.1}$$

For all values of $f_i(\mathrm{AB}) < F_i$, the crystals are fourfold coordinated. For all values of $f_i(\mathrm{AB}) > F_i$, the crystals are sixfold coordinated.

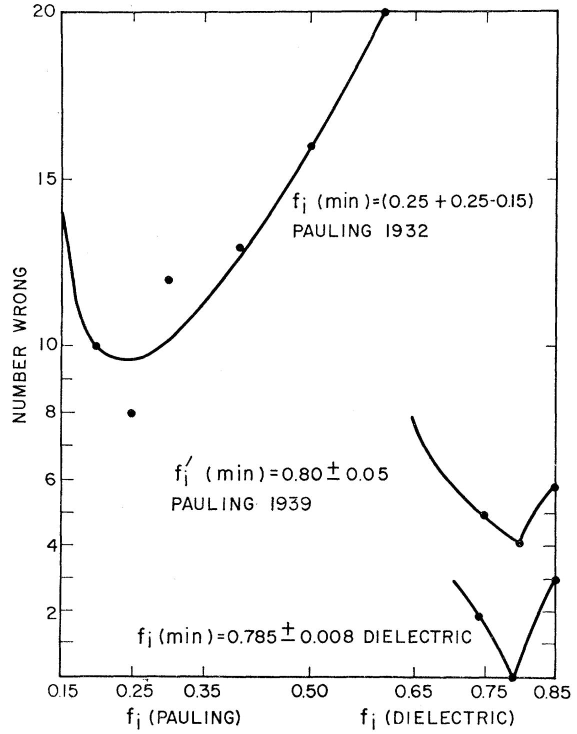


Fig. 11. The statistical test of the ability of an ionicity scale to predict crystalline coordination numbers is applied to Pauling's molecular thermochemical scale and to the spectroscopic crystalline scale based on dispersion theory.

The result (9.1) shows that it is the ionicity f_i of the bond, not its electronegativity difference ΔX, that is the critical factor in determining the crystal structure. Mooser and Pearson already anticipated this result by using the *two* variables ΔX and $\bar{n}$. However, because of their phenomenological approach several inaccuracies are introduced, and the results are less convincing. In this case a connection has been made between phenomenology and quantum mechanics, and surprisingly enough this connection actually improves matters by several orders of magnitude.

Recognizing that ionicity is the critical factor, we can utilize the family of $A^N B^{8-N}$ crystals to test statistically any given definition (Phillips, 1969). If a general prescription for ionicity is proposed, we use that prescription to list f_i(AB) for the seventy crystals not containing transition-metal atoms. We then suppose that there is a critical ionicity F_i such that for $f_i(\mathrm{AB}) < F_i$ the crystal is fourfold coordinated, while for $f_i(\mathrm{AB}) > F_i$ the crystal is sixfold coordinated. In general the ionicity scale will not be wholly accurate, so for each value of F_i we calculate $N(F_i)$, the number of mistakes made in predicting coordination numbers this way. We then vary F_i to minimize $N(F_i)$. Call the minimum point F_i^0 and the minimum value $N_0 = N(F_i^0)$. The accuracy of the ionicity scale is determined by the excursion in F_i about the minimum point F_i^0 required to double $N(F_i)$, i.e.,

$$N(F_i^0 - \Delta F_i^-) = 2N_0, \qquad (9.2)$$

$$N(F_i^0 + \Delta F_i^+) = 2N_0. \qquad (9.3)$$

The number of ionicity scales to which this statistical test can be applied is rather small. For example, the valence bond-molecular orbital calculations of CRS are deliberately restricted to (sp^3)-fourfold coordinated crystals, because a (p^3)-hybridization scheme is better suited to sixfold coordination. Apparently only Pauling's ionicity definition, Eq. (2.7), is sufficiently general to be tested in this way. The result is shown in Fig. 11. Pauling's scale produces about the same number of errors as the Mooser–Pearson plot, although the latter is actually somewhat more accurate. (Some transition-metal compounds are included in Fig. 8.) Note that constant ionicity as defined by Pauling implies $\Delta X =$ const, and corresponds to a vertical boundary line in Fig. 8. Clearly the Mooser–Pearson curved line represents an improvement. However, much better results are obtained using E_h and C as coordinates, because as indicated in Figs. 10 and 11, no errors are made in this way.

Our understanding of any structural property which may depend or E_h, on C, or on some other variable, can be greatly enhanced by finding out what the functional dependence is, without necessarily carrying through an *ab initio* calculation. (The latter might be of little interest even if it were possible, which is very seldom the case.) Thus for structure itself in the $A^N B^{8-N}$ family, it is the ionicity phase angle φ which is important. Other physical properties can be examined in the same way.

Note added in proof: Professor Pauling has drawn my attention to the fact that the values of his ionicity f_i discussed in Sec. 2 are those relevant to *single bonds only*, and that a different definition f_i' is appropriate in crystals. My error arose through reliance on the third edition of his book (1960), which contains only the definition f_i. The definition f_i' is given on pp. 72ff. of the first edition of his book (1939). Professor Pauling's viewpoint, together with my reply, may appear in a 1970 issue of *Physics Today*.

The definition of f_i' given by Pauling proceeds as follows:

Let the coordination number of an $A^N B^{8-N}$ crystal be M. The new definition is

$$f_h' = 1 - f_i' = (N/M) f_h = (N/M)(1 - f_i)$$
$$= (N/M) \exp[-\tfrac{1}{4}(X_A - X_B)^2]. \qquad (9.4)$$

The physical interpretation of (9.4) is that the total covalency $N f_h$ is being shared among M resonating bonds ($M = 4$ or 6). The median deviation between f_i' defined by (9.4), and the spectroscopic ionicity defined

TABLE A. Parameters of the 68 compounds considered. In the first column the structure(s) in which the compound is found is indicated using the abbreviation: D for diamond, Zb for zincblende, W for wurtzite, and R for the rocksalt of NaCl structure. For compounds followed by a †, the values are predicted. For purposes of comparison, the last three columns contain the f_i values predicted by the theories due to Pauling and Coulson.

Crystal	Type	f_i Phillips	f_i Pauling (1932)	f_i' Pauling (1939)	f_i Coulson *et al.*
C	D	0	0	0	0
BN	Zb	0.256	0.22	0.42	0.35
BeO	W	0.602	0.63	0.81	0.64
LiF	R	0.915	0.89	0.98	⋯
BP†	Zb	0.006	0.00	0.25	0.33
SiC	Zb, W	0.177	0.11	0.11	0.06
BeS†	Zb	0.312	0.22	0.61	0.60
AlN	W	0.449	0.43	0.56	0.36
LiCl	R	0.903	0.63	0.94	⋯
MgO	R	0.841	0.73	0.88	⋯
NaF	R	0.946	0.91	0.98	⋯
Si	D	0	0	0	0
BAs†	Zb	0.002	0.00	0.25	⋯
AlP	Zb	0.307	0.01	0.25	0.37
MgS	W, R	0.786	0.34	0.67, 0.78	⋯
BeSe†	Zb	0.299	0.18	0.59	0.60
GaN†	W	0.500	0.39	0.55	0.36
ZnO	W	0.616	0.59	0.80	0.65
LiBr	R	0.899	0.55	0.93	⋯
NaCl	R	0.935	0.67	0.94	⋯
CaO	R	0.913	0.79	0.97	⋯
KF	R	0.955	0.92	0.99	⋯
CuF	Zb	0.766	0.67	0.92	⋯
BeTe†	Zb	0.169	0.09	0.55	0.60
AlAs	Zb	0.274	0.06	0.27	0.37
GaP	Zb	0.374	0.06	0.27	0.37
ZnS	Zb, W	0.623	0.18	0.59	0.61
MgSe	W, R	0.790	0.29	0.65, 0.77	⋯
LiI	R	0.890	0.43	0.91	⋯
CdO	R	0.785	0.55	0.85	⋯
InN†	W	0.578	0.34	0.50	0.36
CuCl	Zb, W	0.746	0.26	0.82	0.85
CaS	R	0.902	0.43	0.81	⋯
NaBr	R	0.934	0.59	0.93	⋯
KCl	R	0.953	0.70	0.95	⋯
SrO	R	0.926	0.79	0.93	⋯
RbF	R	0.960	0.92	0.99	⋯
AgF	R	0.894	0.67	0.95	⋯
Ge	D	0	0	0	0
AlSb	Zb	0.426	0.04	0.26	0.36
GaAs	Zb	0.310	0.04	0.26	0.37
InP	Zb	0.421	0.04	0.26	0.37
MgTe	W	0.554	0.18	0.59	⋯
ZnSe	Zb, W	0.676	0.15	0.57	0.64

TABLE A (*Continued*)

Crystal	Type	f_i Phillips	f_i Pauling (1932)	f_i' Pauling (1939)	f_i Coulson *et al.*
CdS	Zb, W	0.685	0.18	0.59	0.63
CuBr	Zb, W	0.735	0.18	0.80	0.85
AgCl	R	0.856	0.26	0.82	⋯
NaI	R	0.927	0.47	0.91	⋯
CaSe	R	0.900	0.39	0.90	⋯
SrS	R	0.914	0.43	0.91	⋯
KBr	R	0.952	0.63	0.91	⋯
RbCl	R	0.955	0.70	0.95	⋯
GaSb	Zb	0.261	0.02	0.26	0.36
InAs	Zb	0.357	0.02	0.26	0.37
ZnTe	Zb	0.546	0.06	0.53	0.66
CuI	Zb	0.692	0.09	0.78	0.84
CdSe	W	0.699	0.15	0.58	0.61
CaTe	R	0.894	0.26	0.88	⋯
AgBr	R	0.850	0.18	0.86	⋯
KI	R	0.950	0.50	0.92	⋯
SrSe	R	0.917	0.39	0.80	⋯
RbBr	R	0.957	0.63	0.94	⋯
Sn	D	0	0	0	0
InSb	Zb	0.321	0.01	0.25	0.37
CdTe	Zb	0.675	0.04	0.52	0.61
AgI	Zb, W	0.770	0.09	0.80	0.84
SrTe	R	0.903	0.26	0.75	⋯
RbI	R	0.951	0.51	0.92	⋯

by (5.25), is only 0.06, and the f_i' scale does much better than the 1932 f_i scale in predicting crystal structures. See Fig. 11, which has been revised to exhibit this fact. Thus Pauling's new scale (9.4) appears to be quite satisfactory, at least for qualitative purposes.

To analyze these differences in ionicity scales in greater depth, the reader should examine Table A, which shows for 68 $A^N B^{8-N}$ compounds the values of f_i (spectroscopic), f_i [Pauling, 1932, Eq. (2.7)], f_i' [Pauling, 1939, Eq. (9.4)], and f_i (Coulson, 1962). The vast improvement in over-all agreement with f_i (spectroscopic) and f_i (Coulson, 1962) achieved by f_i' (Pauling, 1939) compared to f_i (Pauling, 1932) is obvious at a glance. However, a more quantitative analysis of the correlation of these 272 numbers requires a little organization.

First note that (9.4) can be written

$$f_i' = (N/M) f_i + (1 - N/M). \tag{9.5}$$

Clearly the inequalities

$$0 \leq N/M \leq 1, \qquad 0 \leq f_i \leq 1 \tag{9.6}$$

imply that

$$0 \leq f_i' \leq 1, \tag{9.7}$$

TABLE B. Ranges spanned by various definitions of ionicities f_i (A^NB^{8-N}) as a function of N. Ionicities are listed in detail in Table A.

N	f_i (Phillips, 1969)	f_i (Pauling, 1932)	f_i' (Pauling, 1939)	f_i (Coulson, 1962)
1	0.69–0.96	0.09–0.92	0.78–0.99	0.84–0.85
2	0.17–0.91	0.09–0.79	0.55–0.97	0.60–0.66
3	0.00–0.58	0.00–0.43	0.25–0.56	0.33–0.37

so that the new scale is confined (as it should be) to the unit interval. However, for the new scale, f_i' is generally shifted to higher values because of the constant term in (9.5). As one can see from Fig. 11, this shift is necessary to predict crystal structures better, and indeed part of its success stems from this shift.

Some of the limitations of the new definition (9.5) of f_i' are the following:

(1) The spectroscopic definition treated covalent effects (E_h) and ionic effects (C) symmetrically. Pauling's first definition of f_i is asymmetrical because it depends only on $X_A - X_B$ (see Table 3). The crystal definition is asymmetrical in a different way, because it assumes that it is the covalency f_h' which is "shared" [according to the factor (N/M) in (9.4)] rather than the ionicity f_i'. To some extent these asymmetries balance each other, which partially accounts for the improved agreement.

(2) A large part of the improved agreement stems from a device in the definition (9.5) which one could easily overlook. In order better to distinguish between fourfold and sixfold coordination, the factor M is introduced indicating that covalency is shared between $M=4$ or $M=6$ neighbors. This automatically guarantees that the ionicity per bond will be greater (all other things being equal) for $M=6$ than for $M=4$, because the former makes a larger constant shift in the last term of (9.5). But of course, this is what we were trying to show in the first place. For the crystals MgS, MgSe, and HgS which are found with $M=4$ and 6, the definition (9.5) of f_i' assigns a different ionicity to each, depending on its structure; the two values differing by about 0.12. It is therefore difficult to understand *why* these particular crystals are found in both structures. The spectroscopic theory explains this because all have f_i defined by (5.25) equal to the critical ionicity F_i within 0.01.

(3) Although the deviations of f_i' (Pauling, 1939) and f_i (spectroscopic) are small in many cases, there are some glaring discrepancies. These can be summarized in Table B, which shows the range spanned by all four definitions for crystals with $N=1$, 2, 3. The table once again shows the improvement in f_i' resulting from the upwards shift produced by the constant term in (9.5). It is seen that the center of each group of values of f_i' for constant N agrees well with Coulson's values for f_i, which are almost constant for a given N. This table also shows that the factor (N/M) multiplying f_i in (9.5) has the effect of narrowing the interval spanned by each group. [Note that this narrowing factor must be present in (9.5) in order that (9.6) will ensure the validity of (9.7).] The ranges are narrowed so much that overlap of f_i' ($N=2$) and f_i' ($N=3$) is virtually eliminated, whereas according to the spectroscopic scale, for example, $f_i(\text{BeTe}) \ll f_i(\text{InN})$.

(4) An individual pair of crystals which is particularly striking is BP ($f_i=0$, spectroscopically) and SiC ($f_i=0.18$). According to Pauling's new scale, $f_i'(\text{BP})=0.25$, which is more than twice $f_i'(\text{SiC})=0.11$. The effective charge (see Sec. 12) of BP is too small to measure, whereas that of SiC is quite large (even larger than the spectroscopic scale would predict).

The interested reader can use Table A to multiply these examples. The linear transformation (9.5) corrects many of the shortcomings of Pauling's old f_i scale, especially near $f_i'=0.8$, but it appears to introduce equally serious new shortcomings of its own. All these difficulties appear to stem from the asymmetrical treatment of covalent and ionic effects in the original definition (2.7).

There is another point of overlap between Pauling's definition and the spectroscopic one. In discussing the Ag halides, Pauling suggests that their covalent character should probably be doubled because Ag is frequently divalent. This would give 46% covalent character, or 54% ionic character, instead of the 77% given by f_i'. Obviously there is an ambiguity here, because the halides remain univalent. In the spectroscopic theory, this ambiguity is readily resolved (Van Vechten, 1969) by setting $Z(\text{Ag})$ equal to 2 in Eq. (5.20) for C.

10. COHESIVE ENERGIES AND HEATS OF FORMATION

As we saw in Sec. 2, one can use observed heats of formation of molecules to construct a table of elemental electronegativities based on Pauling's formula (2.3) for the extra-ionic part Δ_{AB} of the bond energy D_{AB}. One can then apply this formula to crystals and estimate heats of formation with an accuracy of 30%–50%. This accuracy is not high, but considering that the input data are obtained from molecules, the results are not bad. More serious, from our present viewpoint, is the absence from the theory of the covalent or homopolar energy, denoted by β in molecular orbital theory, and by E_h in dielectric dispersion theory.

The absence of this factor may be correlated with the fact (Sec. 2) that the resonating-bond concept works well for A^NB^{8-N} crystals with $N=2$, where $C>E_h$, and less well with $N=3$, where $C<E_h$.

Knowing both E_h and C, one feels that it should be possible to develop a general theory of bond energies in A^NB^{8-N} crystals that works well for *both* $N=2$ *and* $N=3$. Because the cohesive energies of A and B as pure crystals may arise not from sp^3 covalent bonding, but from the kind of bonding appropriate to quite different structures, we should express this theory *not* in terms of the heat of formation ΔH_{AB} of the AB crystal from A and B crystals, but rather in terms of the Gibbs free energy ΔG_{AB} at STP required to convert the crystal to atoms. This energy has three parts,

$$\Delta G_{\mathrm{AB}}=\delta G_{\mathrm{AB}}+\Delta G_{\mathrm{A}}+\Delta G_{\mathrm{B}}. \qquad (10.1)$$

Here ΔG_{A} and ΔG_{B} are the free energies required to atomize the elements, and δG_{AB} is the free energy of formation.

In order to say something useful, our theory of the dependence of ΔG_{AB} on E_h and C must be quite accurate. The reason for this is that in most cases

$$|\,\delta G_{\mathrm{AB}}\,| \ll |\,\Delta G_{\mathrm{A}}\,| + |\,\Delta G_{\mathrm{B}}\,|. \qquad (10.2)$$

Thus according to (10.2) most of ΔG_{AB} comes from ΔG_{A} and ΔG_{B}, and the interesting quantity δG_{AB} is an order of magnitude smaller than the quantity ΔG_{AB} which appears naturally in the theory. This is why predicting heats of formation is so difficult.

To determine the dependence of ΔG_{AB} on E_h and C, consider the diamond-type crystals with $A=B$ and $N=4$. Because E_h is a function of lattice constant a, one may investigate the equivalent question of the dependence of ΔG_{AA} on a. If one assumes that

$$\Delta G_{\mathrm{AA}} \propto a^{-n}, \qquad (10.3)$$

then one finds (broadly speaking) that $n=2$ gives a good fit. Because kinetic energy also scales like (10.3) with $n=2$, and because a convenient measure of kinetic energy is the Fermi energy E_F of a free electron gas of density equal to that of the valence electrons, one may measure ΔG_{AA} in units of E_F. When this is done (Phillips, 1969b), the results shown in Table VIII are obtained. Clearly $\Delta G_{\mathrm{AA}}/E_F$ is not constant. This is because there is an important contribution to ΔG_{AA} from *p–d* hybridization. This contribution is virtually absent in diamond, because of the large energy required to promote *2p* states to *3d* states. It reaches a peak in Si, just before the *3d* transition series begins. It is reduced in Ge and Sn, because the *4d* and *5d* valence wave functions, respectively, are orthogonal to *3d* and *4d* core states. This explanation of *p–d* hybridization has also been utilized to explain (Pauling, 1960) trends in the molecular binding energies of diatomic halides and polyatomic hydrides.

More interesting than the trends in ΔG_{AA} are the trends in ΔG_{AB} when the cores of A and B are isoelectronic, i.e., when A and B belong to the same row of the periodic table. Two examples of such sequences are: first row: $N=4$, diamond; $N=3$, BN, and $N=2$, BeO; third row: $N=4$, Ge; $N=3$, GaAs, and $N=2$, ZnSe. According to Pauling, in cases where both atoms belong to the same row, $\Delta X_{\mathrm{AB}}(N=2)$ is just twice ΔX_{AB} for $N=3$. The number of resonating bonds in each case is just N. Therefore the heats of formation $\Delta H_{\mathrm{AB}}(N)$ will have the ratio

$$\Xi=\Delta H_{\mathrm{AB}}(2)/\Delta H_{\mathrm{AB}}(3)=\tfrac{2}{3}(2)^2=2.67. \qquad (10.4)$$

In the first-row series, corrections are required for N and O multiple bonds. When these are made, the effective experimental ratio Ξ corresponding to (10.4) can be plotted against lattice constant in Fig. 12. The value given by Pauling's theory is seen to straddle the experimental values, but there is an interesting trend in $\Xi(a)$ which is omitted by the molecular model.

The cases where the cores of A and B are isoelectronic are of particular interest because in such cases a is virtually independent of N, i.e., only C varies appreciably as a function of N. This also holds true for any sequence A^NB^{8-N}, providing either A or B belongs to one row, and B or A belongs to another. In Fig. 13 we therefore plot $\Delta G_s(\mathrm{AB})$ for all such sequences, with the single exception of compounds with one and only one first-row atom. In this case the size disparity between the cores of the two atoms is so great that there is no simple pattern.

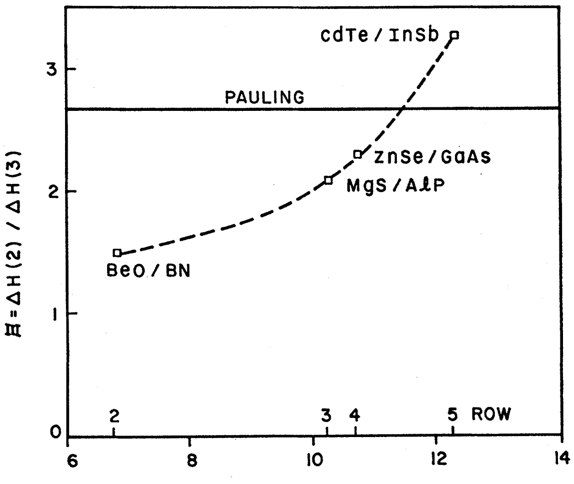


FIG. 12. The ratio $\Xi(a)=\Delta H(2)/\Delta H(3)$ for each series of crystals A^NB^{8-N}, with the cores of A and B isoelectronic, and with lattice constant a for $N=4$. The value for this ratio given by Pauling's thermochemical theory is 8/3.

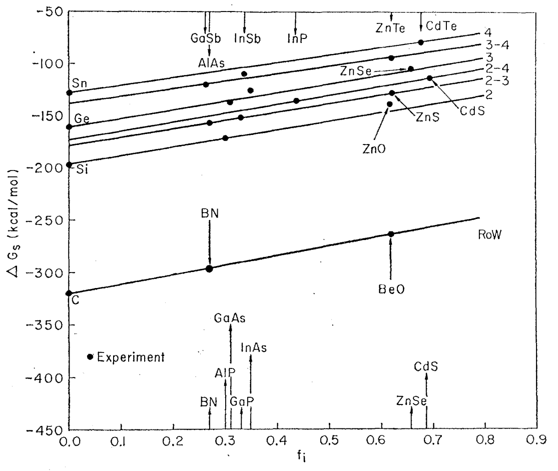


FIG. 13. Plot of the Gibbs free energy of atomization of $A^N B^{8-N}$ crystals for sequences in which one atom belongs to Row R_1 of the Periodic Table, and the second atom belongs to Row R_2. The energies of atomization are seen to be linear functions of ionicity.

The results which are illustrated in Fig. 13 are the following:

(1) When the cores are strictly isoelectronic, i.e., both atoms belong to the row R, the following simple relation (Phillips, 1969a; Phillips and Van Vechten, 1970) holds to within the limits of experimental error:

$$\Delta G_s(\mathrm{R}, f_i) = \Delta G_s(\mathrm{R}, 0)[1 - k(\mathrm{R}) f_i], \tag{10.5}$$

where f_i is the spectroscopically defined ionicity, accurate to 1%.

(2) When one atom comes from Row R_1 and one atom from Row R_2, a relation similar to (10.5) is approximately valid (Phillips, 1969c), with $\mathrm{R} \approx (\mathrm{R}_1 + \mathrm{R}_2)/2$. However, there is a small tendency for $\Delta G_s(\mathrm{AB})$ to be displaced towards the value given by (10.5), with $\mathrm{R} = \mathrm{R}_\mathrm{B}$, i.e., there is a small shift towards the linear relation valid for the row R to which the *anion* belongs. For GaSb and InAs ($R_1 = 3$, $R_2 = 4$) this shift amounts to ±6 kcal/mol, respectively, and it represents a core effect absent from any valence bond model.

(3) The values of $k(\mathrm{R})$ are remarkably constant. The only appreciable variation is the difference between the value found for $\mathrm{R} = 1$, $k(1) = 100$ kcal/mol, and the values for other values of R, $k(\mathrm{R} \geq 2) = 80$ kcal/mol. If one writes

$$k(\mathrm{R}(a)) \propto a^{-n_2}, \tag{10.6}$$

then $n_2 \approx 0.5$.

The scaling results (10.5) and (10.6) suggest the following picture for the binding energy of tetrahedrally coordinated crystals. For the homopolar values $\Delta G_s(0)$, it appears that most of the binding energy has kinetic character and is therefore associated with delocalized (itinerant) valence electrons. This confirms the relevance of the single-particle approximation (4.9) because even though the one-electron functions ψ_i have not been used to calculate ΔG, the translational periodicity of the crystal requires ψ_i to have the Bloch or delocalized form. If similar scaling arguments are applied to small molecules (such as MH_3, where $\mathrm{M} = \mathrm{N}$, P, As or Sb) one finds $n_1 = 1$, i.e., the binding energy scales like a potential energy, as one would expect for localized bonds. Thus scaling provides a simple way of showing that localized bonds are appropriate for discussing cohesive energies in molecules, but that the valence electrons in covalent crystals are itinerant, in much the same way as in metals like Na or Al.

The second scaling result concerns $k(\mathrm{R})$, the slope which describes the reduction in cohesive energy with increasing magnitude of C compared to E_h. The fact that $n_2 \approx 0.5$ suggests that $|\Delta G|$ is reduced because energy is required to screen $V_\mathrm{A} - V_\mathrm{B} = V_\mathrm{ionic}$. Dielectric screening can be described (Mott and Jones, 1958) at short wavelengths by the Thomas-Fermi screening wavelength $\lambda = k_s^{-1}$. This suggests

$$k(l) \propto V_\mathrm{ionic}(a/\lambda), \tag{10.7}$$

since when λ is large, little screening of V_ionic has taken place. Of course, $V_\mathrm{ionic} \propto a^{-1}$, and the Thomas-Fermi formula shows $\lambda \propto a^{1/2}$. Thus (10.7) gives $n_2 = 0.5$, as observed experimentally.

We remarked previously that it is difficult to develop a quantitative theory of δG_AB or of the heat of formation ΔH_AB because these energies are small compared to ΔG_A, ΔG_B, ΔH_A, and ΔH_B. Phillips and Van Vechten (1970) have shown how to construct an accurate phenomenological model which explains ΔH_AB (or ΔG_AB). They find that two effects are important: ionicity, as implied in Pauling's thermochemical theory, and the tendency toward metallization discussed by Mooser and Pearson (1959; Pearson, 1962). The former factor is represented, of course, by the spectroscopic ionicity f_i, while the latter is described by the dehybridization factor D.

Dehybridization describes the decrease in energy of the tetrahedral sp^3 covalent structure compared to a much less directional (although probably still partially covalent) metallic one. This occurs for the heavier elements because of the decrease in atomic s valence energies compared to atomic p valence energies which is illustrated in Fig. 4. In the diamond-type crystals this effect causes the metastability of white Sn and grey Sn. Dehybridization implies a breakdown not only of sp^3 bonding but also of the Penn one-gap model. This breakdown is apparent in the growth of the low-energy optical peak (labeled E_1 in most experimental papers), compared to the main peak (labeled E_2). This trend is apparent in Fig. 3. Furthermore the smallest direct gap (labeled E_0) itself actually goes to zero in grey Sn.

Phillips and Van Vechten found (1970) that ΔH_{AB} in more than 20 zincblende and wurtzite-type tetrahedrally coordinated semiconductors could be fitted very well by the formula

$$\Delta H_{AB} = \Delta H_1 (a_1/a_{AB})^s f_i(AB) D(E_0, E_1, E_2), \quad (10.8)$$

where ΔH_1 and a_1 are the heat of formation and lattice constant of a reference crystal (e.g., GaAs) and the scaling index s is about three. It was found that the best functional form of $D(E_0, E_1, E_2)$ was

$$D(E_0, E_1, E_2) = 1 - b(E_2/\bar{E})^2, \quad (10.9)$$

where $\bar{E}$ could be any average of E_0 and E_1. (They used $2\bar{E} = E_0 + E_1$.) The value of b is given by

$$b = (\bar{E}/E_2)^2_{\text{grey Sn}}, \quad (10.10)$$

which incorporates dehybridization and metallization (grey Sn–white Sn) in a natural way. By adjusting ΔH_1 and s, (10.8) can be made accurate to about 10%. The very small heats of formation of crystals like InSb are explained through the combined effects of f_i and D. One can also use (10.10) to explain the nonformation of tetrahedral structures by TlSb and TlBi, because for these crystals D is negative.

Apart from the dehybridizing factor D, the presence of $f_i(AB)$ in (10.8) implies that $\Xi = \Delta H(2)/\Delta H(3)$ has a different meaning than in the Pauling theory, i.e., the heats of formation of II–VI compounds compared to III–V compounds are explained quite differently. As C^2 or $(\Delta X)^2$ increases, both theories give a saturating behavior. In Pauling's theory this occurs because the discrete variable N, which measures the number of bonds per $A^N B^{8-N}$ molecule, has dropped from 3 to 2. In the spectroscopic theory, C also doubles (if the bond length does not change, e.g., on going from GaAs to ZnSe), but saturation is achieved because of the denominator $(E_h^2 + C^2)^{-1}$ in the definition

$$f_i = C^2/(E_h^2 + C^2). \quad (10.11)$$

One may compare this with Pauling's expression

$$N(X_A - X_B)^2 \quad (10.12)$$

by rewriting (10.11) as

$$f_i = (C^2/E_h^2)[E_h^2/(E_h^2 + C^2)] = (C^2/E_h^2) f_c, \quad (10.13)$$

where f_c is the fraction of covalent character defined by Eq. (5.26). One sees that the factor C^2/E_h^2 corresponds to $(X_A - X_B)^2$, while f_c corresponds to N. The variation of N is integral, while f_c is a continuous spectroscopic variable which has no artificial atomistic character (i.e., counting s and p states in a partial wave expansion as one does to define valence). In general f_c is not proportional to N, i.e., f_c is *not* the same for all $N = 3$ crystals, and the average value of f_c for all $N = 2$ crystals is *not* 2/3 the value of f_c for all $N = 3$ crystals. (The actual ratio is closer to 3/7.) This explains to some extent the point (CRS, 1962) discussed previously in Sec. 2, that Pauling's choice of molecular parameters gives good results for $N = 2$ crystals, but not for $N = 3$.

11. INTERATOMIC FORCES

The usual molecular approach to the interatomic forces which determine molecular vibration frequencies in covalent molecules (e.g., hydrocarbons) is a parametric one based on what is called a valence force field (VFF).[1] It is assumed that the interatomic forces are primarily short-range ones connected with stretching of covalent bonds, or bending of the angle between covalent bonds. This model is particularly well suited to atoms whose valence electrons normally have principal quantum number $n = 1$ or 2. As we saw in the preceding section, this effectively confines the electrons to a subspace of s and p states, because the energy of promotion to $3d$ or $4f$ states is so large. Then the bond angles are not easily changed, and bond-bending characterizes the noncentral forces quite well. Nearest-neighbor bond-stretching forces are simply nearest-neighbor central forces such as one would expect to find in any model.

It is not obvious *a priori* that nearest-neighbor central forces, characterized by the parameter α, and next-nearest-neighbor bond-bending noncentral forces, denoted by β, will be sufficient to describe the lattice vibrations of tetrahedrally coordinated crystals. Indeed for vibrations of very short wavelength, such forces are not adequate. However, it has been shown by Keating (1966) that the two-parameter α–β model gives very good results for the three cubic elastic constants c_{11}, c_{12}, and c_{44} of diamond, Si, and Ge. The extention (Martin, 1970) of Keating's model to zincblende crystals such as GaAs requires the inclusion of long-range Coulomb forces between the cation and anion sublattices. These can be determined independently from experimental values of the splitting of the longitudinal and transverse optic frequencies at $\mathbf{k} = 0$, and so introduce no additional parameters.

The proof that the two-parameter model gives a good account of the three elastic constants is obtained by eliminating α and β from the expressions for c_{11}, c_{12}, and c_{44}. This leads to an identity which is satisfied to within 10% for diamond, Si, and Ge (Keating, 1966) as well as for most zincblende crystals (Martin, 1970).

Because β is a measure of the strength of the noncentral forces while α measures the strength of central forces, the dimensionless ratio β/α measures the

---

[1] An interesting example of the precision attainable with the use of a VFF is provided for zigzag long-chain hydrocarbons by Schactschneider and Snyder (1963).

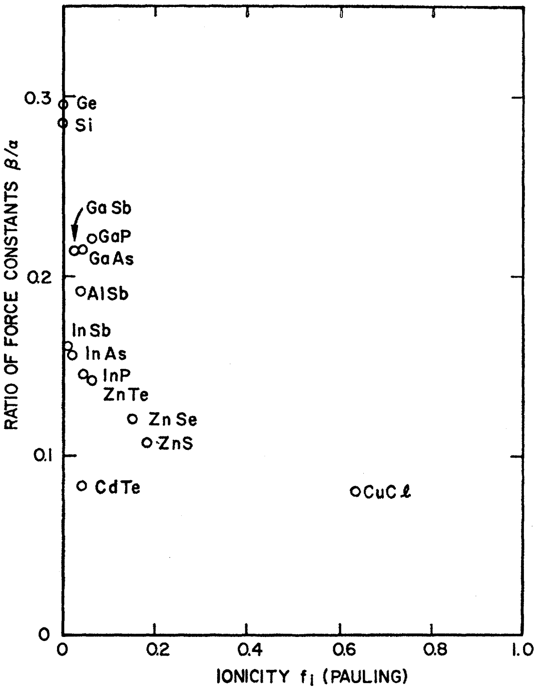


Fig. 14. The directionality ratio β/α as determined from elastic constants plotted against Pauling's ionicity.

importance of the covalent bond in determining the stability of the tetrahedral structure. This is a nontrivial point, because one would not expect the open, low-density diamond or zincblende structure to be stable in the absence of bond-bending forces. Indeed in the limit $\beta\rightarrow 0$ both shear constants [c_{44} for (100) transverse modes, and $c_{11}-c_{12}$ for (111) transverse modes] tend to zero (Keating, 1966). Interestingly enough, the value of this ratio is about the same for Si, Ge, and presumably grey Sn, where it has the value of about 0.3, much smaller than the diamond value which is near 0.7. If we confine ourselves to the family of zincblende crystals composed entirely of atoms *not* from the diamond row, we may find interesting trends in β/α as a function of either E_g or φ.

What trend would one expect to find? Presumably α and β depend on E_g in the same way, because the ratio β/α does not change from Si to Ge. Indeed Keyes showed some time ago that α scales like a^{-4}, where a is the lattice constant. (This is a result that can be obtained by dimensional analysis.) Thus α scales like $E_h{}^n$ with $n=8/5$, since E_h scales like $a^{-2.5}$, according to Eq. (5.19).

The interesting trend is therefore the dependence of β/α on φ. As $f_i=\sin^2\varphi$ tends toward unity, the binding tends towards the fully ionic or closed-shell limit. In this limit it is known from studies of rare-gas solids (which always have close-packed structures) that bond-bending forces are negligible. One would therefore expect that

$$\beta/\alpha=0.3(1-f_i)^m \tag{11.1}$$

should give a good fit to the experimental data. Two questions arise: which of the ionicity scales we have discussed fits (11.1) best, and what is the value of m?

In Fig. 14 we show the values for β/α plotting against f_i as defined thermochemically by Pauling, Eq. (2.7). For most of the crystals composed of atoms from the Si, Ge, and Sn rows, Pauling's definition makes f_i much too small. All the points are crowded to the left-hand side of the figure, and no satisfactory plot is obtained.

The molecular orbital values of f_i on CRS discussed in Sec. 3 are used as coordinates for plotting β/α against f_i in Fig. 15. The CRS values of f_i are nearly constant for all A^NB^{8-N} crystals with the same value of N, provided neither A nor B is a first-row atom. As a result, chemical differences between GaSb and InP, the least and most ionic crystals with $N=3$ of this type, are not brought out. However, the over-all trend of β/α towards zero as f_i tends to unity is indicated approximately. Thus the CRS index of ionicity is superior to Pauling's index so far as directionality (as manifested by β/α) is concerned.

The spectroscopic values of f_i as determined by dispersion theory are used as coordinates for plotting β/α against f_i in Fig. 16. The high accuracy of a completely

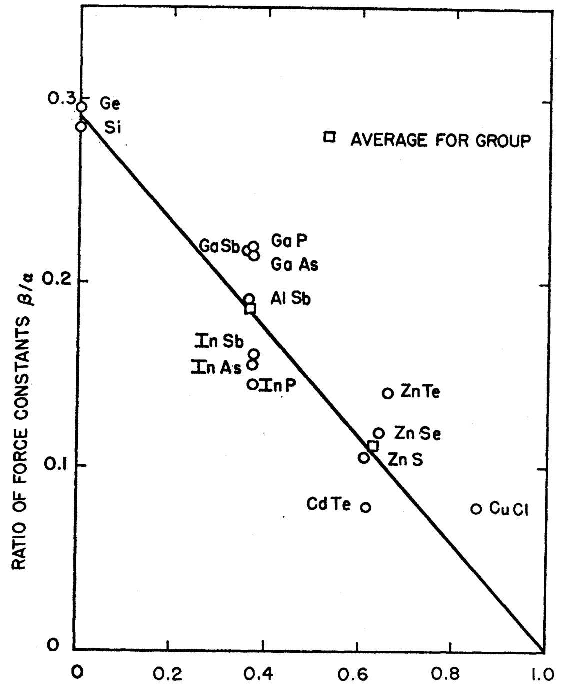


Fig. 15. The directionality ratio β/α plotted against ionicity as derived by Coulson *et al.*, from valence bond and molecular orbital theory.

spectroscopic theory is brought out by this plot. The relation (11.1) is satisfied with $m=1$, just as one would have expected from the linear behavior of cohesive energies as a function of ionicity as described by Eq. (10.5). The most ionic crystal shown in Fig. 16 with $N=3$ is InP, the least ionic with $N=2$ is ZnTe. The two ionicities are almost equal, as are the respective values of β/α. Even CuCl, which has $N=1$ and is just barely stable in the zincblende structure, falls on the line extrapolated to $\beta/\alpha=0$ at $f_i=1$.

The results of the preceding section and this section are of special interest because they test the functional dependence of ionicity on electronegativity. As we indicated in connection with Eq. (10.5), any monotonic increasing function of the ionicity phase angle φ will predict with equal success the coordination numbers of A^NB^{8-N} crystals. The statistical test of ionicity discussed in Sec. 9 demonstrates that the proper combination of ionic and covalent energies is the obvious dimensionless one

$$y=C/E_h, \qquad (11.2)$$

and of course φ itself is a monotonic function of y,

$$\varphi=\tan^{-1} y. \qquad (11.3)$$

However, the particular form of the functional dependence of f_i on φ or on y should be determined in such a way that structural properties such as the cohesive energy and the strength of directional bonding behave in the simplest possible way, i.e., linearly. The fact that this is achieved by (11.3) using y rather than $2y$ or $y/2$ is a consequence of the symmetrical way in which E_h and C enter the spectroscopic definitions, especially Eq. (5.10).

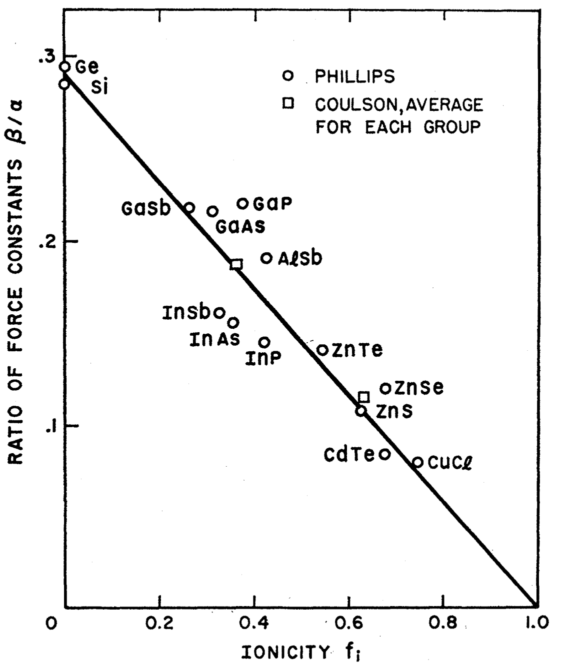


FIG. 16. The directionality ratio β/α plotted against spectroscopic ionicity. Also shown are the averages of the Coulson values for the $N=2$ and $N=3$ groups shown in Fig. 15.

12. IONICITY AND EFFECTIVE CHARGES

In lieu of a theoretical scale of ionicity for crystals, for some time it has been customary to use as an empirical measure of ionicity the dynamic effective charges associated with optic modes of vibration at $\mathbf{k}=0$. The situation is particularly simple for crystals of the zincblende A^NB^{8-N} class. There are three acoustic modes as $k\to 0$ which are described by the elastic constants discussed in Sec. 11. There are also three optic modes, and if we let $\mathbf{k}\to 0$ along a (100) or (111) symmetry axis, two of these are transverse and one is longitudinal. In the absence of long-range forces the transverse and longitudinal optic frequencies ω_{to} and ω_{lo} would necessarily become degenerate at $\mathbf{k}=0$, since for short-range forces the modes are indistinguishable. The Coulomb force, however, gives rise to surface forces which split the lo mode off from the to model. One may derive the general phenomenological relation

$$\omega_{\text{lo}}^2=(\epsilon_s/\epsilon_0)\,\omega_{\text{to}}^2 \qquad (12.1)$$

known as the Lyddane–Sachs–Teller relation.[2] Here ϵ_0 is the low-frequency limit of the electronic polarizability measured in the infrared at frequencies above the lattice vibration absorption bands. On the other hand, ϵ_s is the quasistatic polarizability of the crystal measured at frequencies below these absorption bands, and it includes contributions to the polarizability associated with ion motion.

If one chooses a particular electrostatic model, one can use the splitting $\omega_{\text{lo}}^2-\omega_{\text{to}}^2$ to define an effective charge. Two definitions (Callen, 1949; Burstein *et al.*, 1967) which are useful for semiconductors are derived by considering a sample with the geometry of a slab. If the ions vibrate in the plane of the slab, one finds that the induced dipole moment per ion pair displaced relative to each other by x is:

$$\mathbf{p}_T=e_T^*\mathbf{x}. \qquad (12.2)$$

Here e_T^* is the transverse or Born effective charge (Burstein *et al.*, 1967). On the other hand, when the ions vibrate normal to the plane of the slab (corresponding to a longitudinal mode in an infinite crystal), there is a macroscopic internal field $\mathbf{E}$. Since $\mathbf{D}=0$, outside the crystal one has

$$\mathbf{E}=-4\pi\mathbf{P}_L, \qquad (12.3)$$

---

[2] For a discussion of this relation and its generalization to cases containing more than two atoms per unit cell, see Cochran and Cowley (1962).

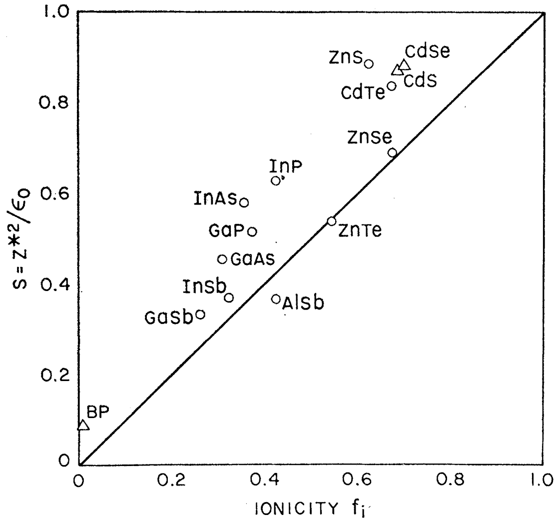


FIG. 17. The effective charge parameter S as a function of the spectroscopic ionicity f_i.

where $\mathbf{P}_L$ is the longitudinal polarization. This is given by

$$\mathbf{p}_L = e_T^*\mathbf{x} + (\epsilon_0 - 1/4\pi)\mathbf{E}. \tag{12.4}$$

One can use (12.3) to rewrite (12.4) as

$$\mathbf{p}_L = e_L^*\mathbf{x} = (e_T^*/\epsilon_0)\mathbf{x}. \tag{12.5}$$

One can use (12.2) and (12.5) to calculate $\omega_{lo}^2 - \omega_{to}^2$. The results are (here M is the reduced mass of the two atoms in a unit cell of volume Ω)

$$\omega_{lo}^2 - \omega_{to}^2 = (4\pi\epsilon_0/M\Omega)(e_L^*)^2 \tag{12.6}$$

and

$$\omega_{lo}^2 - \omega_{to}^2 = (4\pi/M\Omega)[(e_T^*)^2/\epsilon_0]. \tag{12.7}$$

The effective charge e_L^* was introduced by Callen (1949), and e_T^* was introduced by Born. It has been pointed out (Burstein *et al.*, 1967) that e_T^* is much more nearly constant in semiconductors than e_L^*. The difference between the two arises from the way in which allowance is made for polarization of the crystal during the optical vibration.

It has been suggested by Martin (1970) that perhaps one should not attempt to factor the splitting

$$S = (e_T^*)^2/\epsilon_0 = \epsilon_0(e_L^*)^2 \tag{12.8}$$

into effective charges and polarization factors, neither of which is measurable separately. Instead Martin has proposed to study S itself as a function of f_i. The results of such a study are shown in Fig. 17, where the spectroscopic values of f_i are used as abscissas. As $f_i \to 1$, it appears that S tends to slightly more than one. Moreover, S is nearly linear in f_i.

The crystals which are exhibited in Fig. 17 are the zincblende ones not containing first-row atoms. These are the same crystals which produced a simple pattern for the directionality parameter β/α discussed in the preceding section. We noted there that in crystals composed of first-row atoms like diamond and cubic BN, β/α is much larger. This may be an indication that the role played by bond stretching, as measured by α, is different from that of bond bending as measured by β. Apparently bond bending makes a greater contribution to S than does bond stretching because all the tetrahedrally coordinated crystals containing first-row atoms exhibit values of S which are not shown, but which lie well above the line drawn in Fig. 17. (The one that lies farthest above is BN with $f_i = 0.26$, $S = 1.36$.) Similarly, in the sixfold coordinated rocksalt structure cases (also not shown), all the experimental values of S lie well below the line.

The foregoing picture, although incomplete, shows that the splitting $\omega_{lo}^2 - \omega_{to}^2$ is indeed a measure of ionicity through the parameter S for some tetrahedrally coordinated crystals. Because the concept of effective charge varies with the geometry assumed, it is perhaps more useful to study the splitting $\omega_{lo}^2 - \omega_{to}^2$ directly rather than construct a specific model.

The concept of ionicity can also be used to elucidate trends in piezoelectric constants (Phillips and Van Vechten, 1969a). These are even more sensitive to geometrical factors, and so will not be discussed in detail here.

13. SUBSTITUTIONAL IMPURITIES

The replacement of an atom in a tetrahedrally coordinated $A^N B^{8-N}$ semiconductor by an impurity atom of valence Z may create shallow impurity states. Let the valence of the replaced host atom be Z_0, then the valence difference is

$$\Delta Z = Z - Z_0. \tag{13.1}$$

In general, shallow impurity states are obtained for $\Delta Z = \pm 1$, 0. For $\Delta Z = +1$, the impurity is said to be a donor; for $\Delta Z = -1$, it is said to be an acceptor, and for $\Delta Z = 0$ it is said to be isoelectronic. Transition-metal impurities lie outside the scope of the present discussion, although it is possible that the generalization of the Mooser–Pearson theory discussed in Sec. 9 would yield useful insight into their properties as well.

The early work on shallow substitutional impurities concerning the cases $\Delta Z = \pm 1$ was reviewed by Kohn (1957). At that time it was recognized that with the exception of the ground state all the remaining bound states of the impurities tended to have hydrogenic character, i.e., their energies conformed approximately to a Rydberg series

$$E = -R/n^2 \tag{13.2}$$

for $n \geq 2$. Because of the complexity of the energy-band structure of semiconductors near the bottom of the conduction band or top of the valence band, (13.2) is not the whole story (Kohn, 1957). One must allow for the multiplicity of conduction-band edges, which exhibit ellipsoidal symmetry, and for the degeneracy of the valence-band edge, which produces spheroidal surfaces of constant energy. When this is done and use is made of the most recent experimental data, one can show (Faulkner, 1969) that the part of the donor potential seen by p donor electrons with $n \geq 2$ has the dielectrically screened Coulomb form

$$V_{\text{imp}}(r) = -\Delta Z/\epsilon_0 r \qquad (13.3)$$

to an accuracy of better than 1% for donor impurities in Si and Ge.

In view of these results, it seemed natural to assume (Kohn, 1970) the validity of (13.3) right into $r = R_c$, where R_c is a cutoff radius approximately equal to the atomic radius of the impurity. Within this model short-range effects associated with the potential interior to R_c are presumably not large. This expectation was upset by the discovery (Thomas and Hopfield, 1966) that bound states occur for $\Delta Z = 0$, e.g., N replacing P in GaP, or Bi replacing P in GaP. This discovery was surprising for two reasons:

(1) Because the long-range part of the impurity potential (13.3) is zero for $\Delta Z = 0$, there must be an appreciable short-range part which presumably arises from differences in electronegativity (or some related quantity) between the impurity and the host atom it has replaced. Any rough order-of-magnitude estimate will then show that if a similar short-range potential is present when $\Delta Z = \pm 1$, the ground-state energies would depart from the hydrogenic values by much more than is observed to be the case. Stated differently, if an upper limit is imposed on the magnitude of the interior potential by the deviations of the 1s energies from the hydrogenic value for $\Delta Z = \pm 1$, then that interior potential will be too weak to produce binding when $\Delta Z = 0$.

(2) There is a minimum strength for a short-range potential to produce a bound state. (In a square-well model, the strength is measured by $V_0 R^2$, where V_0 is the depth of the well and R is its radius.) If the strength exceeds the minimum value, the binding energy rapidly increases and soon becomes comparable to the energy gap ΔE_{cv} between the bottom of the conduction band and the top of the valence band. Thus it seems very unlikely that V_0 will be close enough to the critical value to produce binding energies which are $\lesssim 0.01 \Delta E_{cv}$. Nevertheless, this is the range of observed binding energies for most isoelectronic impurities. It is clear that some physical mechanism must be operating to reduce binding energies to smaller values than one would have expected from electronegativity arguments alone.

To these problems one should add another, which concerns the hyperfine coupling between the donor electron and the nuclear spin of the impurity. This hyperfine coupling is proportional to the expectation value in the 1s ground state of the singular operator $\delta(\mathbf{r})$, i.e., it is proportional to $|\Psi(0)|^2$. From many studies of atomic wave functions it is known that $|\Psi(0)|^2$ is more sensitive than any other observable to small changes in the potential. For example, one may know the potential well enough to calculate the ground-state energy (which satisfies a variational condition) to a few hundredths of a percent, and still err in calculating $|\Psi(0)|^2$ by a factor of 2 or more. Moreover, the parts of the potential that affect $|\Psi(0)|^2$ strongly, such as the potential in the neighborhood of the atomic core 1s state, have almost no effect on valence energies corresponding, say, to atomic 3s or 3p states.

In spite of these obstacles, Kohn and Luttinger (1955) proposed the following method for calculating $|\Psi(0)|^2$. The wave function $\Psi(r)$ of the donor electron is written (schematically) in the product form

$$\Psi(r) \propto F(R)\varphi_k(r), \qquad (13.4)$$

where $\varphi_k(r)$ is a Bloch function associated with the conduction-band edge. The envelope wave function $F(R)$ satisfies an effective wave equation containing the potential (13.3) for $R \geq R_c$. The potential for $R < R_c$ is left unspecified. In order to compensate for incomplete knowledge of the Hamiltonian, the *observed* value E^0 of the 1s binding energy, not the hydrogenic value $E_H{}^0$, is inserted into the effective wave equation, which is then integrated inwards from $r = \infty$. Of course in the potential (13.3) only wave functions corresponding to hydrogenic eigenvalues would be finite at $r = 0$. Therefore the integration is stopped at $r = R_c$, and it is assumed that

$$|\Psi(0)|^2 = s\,|F(R_c)|^2\,|\varphi_k(0)|^2, \qquad (13.5)$$

where s is a multiplicity factor measuring the number of equivalent band edges.

When (13.5) is used to study the hyperfine coupling constant for P impurities in Si, one finds that the observed value of $|\Psi(0)|^2$ is about 10 times greater than that which would be obtained from (13.5) using the hydrogenic ground state energy $E_H{}^0$. The factor of 10 is, however, explained by (13.5) using the *observed* value of the binding energy E^0 and setting R_c equal to the atomic radius.

This remarkable result appeared most impressive some 15 years ago. As more information has accumulated concerning impurity properties, however, it has become more difficult to understand. For example, it seems strange that when one knows so little about the Hamiltonian that one is unable to calculate the ground-state energy, one should be able to calculate the expectation value of a singular operator which depends on the behavior of the wave function at the center of the region of which one disclaims any knowledge. Because the wave function is diverging as $r \rightarrow 0$, *any* value of $|\Psi(0)|^2$ greater than the hydrogenic value can be obtained by a suitable choice of the cutoff radius R_c. In fact the latter constitutes the one adjustable parameter which explains the one observable.

From the viewpoint of information theory, to have a meaningful model it is necessary that there be fewer adjustable parameters than there are observables explained. This means that in order for the Kohn–Luttinger theory to be meaningful, it is necessary that R_c remain equal to the atomic radius when fitting the hyperfine coupling constants of impurities other than P in Si (e.g., As, Sb in Si, or P, As, Sb in Ge). Even if the fit is not quantitative, one would hope that the observed values of $|\Psi(0)|^2$ would correlate monotonically with deviations of the observed binding energy E^0 from the hydrogenic value $E_H{}^0$. Unfortunately this is not observed to be the case. With Sb impurities in Si, $|F(R_c)|^2$ is found experimentally to be twice as large as for P impurities in Si, even though $E^0 - E_h{}^0$ is smaller for Sb than for P.

Kohn and Luttinger explain the failure of their theory to account for the hyperfine coupling of As and Sb impurities compared to P by stating that the As and Sb perturbations are "violent" compared to the P perturbation, presumably because only the P atomic core is isoelectronic to the Si atomic core. However, by direct calculation based on studies of interactions between rare-gas atoms, one can show (Phillips, 1969a), that when Si is the host crystal, core–core interactions are negligible, except possibly in the case of Bi.

All these problems become much less serious when one realizes that the chemical shifts of ground-state energies are indeed a chemical problem, and that before proceeding to try to understand something so complex as $|\Psi(0)|^2$, one should first try to understand E^0. This is the procedure ordinarily adopted in most quantum-mechanical problems, and if it has not always been applied to substitutional impurities in semiconductors, that may be partly explained in terms of the lack of a quantitative theory of electronegativity and ionicity in crystals.

To begin with, we should understand why ground-state binding energies tend to be small, regardless of whether ΔZ is zero or nonzero. The simplest hypothesis seems to be (Phillips, 1969a) that the lattice relaxes in such a way as to reduce the amount of bound charge $e_c{}^*$ which accumulates in the impurity cell. In Si it appears that the total donor charge $e_c{}^*$ in the atomic cell seldom exceeds $0.1e$, while in Ge, the total donor charge which can accumulate is even less. As the crystal becomes more ionic, e.g., $N=3$ and then $N=2$ in $A^N B^{8-N}$ crystals, possibly more charge (say $0.2e$) can accumulate, but in any case really large binding energies (of order ΔE_{cv}) do not usually occur, because this would lead to an unreasonably large charge accumulation.

The next step, of course, is to explain why large charge accumulations do not occur (Phillips, 1970). The reason for this is that each unit cell in the crystal has four covalent bonds and therefore the valence energy is minimized (maximum binding energy) when there are exactly eight valence electrons in each unit cell. The extra energy associated with binding a donor electron, of course, makes it favorable to attract the donor electron. Because all these semiconductors are highly polarizable, if a certain amount of charge $e_c{}^*$ associated with the donor electron accumulates in the central cell, then an amount of valence charge of order $(1-1/\epsilon_0)e_c{}^*$ will be displaced from the central cell. With $\epsilon_0 \sim 10$, this means that accumulation of donor charge displaces a nearly equal amount of valence charge, which of course costs energy. Because $e_c{}^*$ is of order $|F(0)|^2$, and because $|F(0)|^2 \propto (E^0)^3$ in a hydrogenic approximation, it is clear that when E^0 is small enough, the displacement effect is unimportant. As E^0 increases, a point is reached where the rate of reduction in total energy associated with binding the additional carrier is equal to the rate of increase in total energy associated with the displacement of valence electrons. This stabilizes the total energy and determines the amount of donor charge which is allowed to accumulate.

The foregoing discussion assumes that the additional potential seen by the bound carrier is not independent of the wave function of the bound carrier, but that it changes self-consistently with $e_c{}^*$. We do not really need to know how this happens, but it is useful to have an explicit model. One possible model, which is consistent with the importance of valence energies, asserts that it is the strain field which sets up a repulsive potential proportional to r^{-3} near the impurity which stabilizes the ground state (Phillips, 1969a). Thus the potential near the impurity has *two* parts, an attractive well which binds a carrier in the first place, and a repulsive shell around this well, which keeps the binding energy small. The net binding is related to differences in valence bond energy between the impurity and the host which determine the point of stability.

At this point it would be helpful if we knew how to calculate these differences. As we indicated in discussing alternative crystal structures in Sec. 9, such a calculation appears to require extremely accurate determinations of each bond energy separately, and this is just what no one has done. (The situation is even more complicated here, because one must determine the local strain field as well.) One can therefore have recourse to the same maneuver here as we used in discussing crystal structures. One *assumes* that $E^0 - E_H{}^0$ is related to the differences in E_h and C for the two bonds, or equivalently is some function of the differences in E_g and φ for the two bonds. Because these can be estimated from the formulae given in Sec. 5, it is then a relatively simple matter to examine the eight known experimental values of $E^0 - E_H^0$ and find the desired functional relationship. It is

$$|E^0 - E_H^0| \propto \Delta E_b \propto |E_g(\text{host}) - E_g(\text{impurity})| \quad (13.6)$$

for a given host. To improve the agreement of (13.6) with experiment, it turns out that one can make a small correction ΔE_{pd} for the effects of p–d hybridization, which we saw in Sec. 10 influences cohesive energies, especially in the Si row. The theoretical values (Phillips, 1970) for ΔE_b and ΔE_{pd} are shown in Fig. 18, and they are compared with experiment in Fig. 19. In only one case (Bi in Si) is agreement with experiment poor.

We remarked previously that a satisfactory theory would attribute an approximately constant volume to the region over which the impurity potential departed from the asymptotic hydrogenic form (13.3). The way to test this point using the relative chemical model (13.6) is to introduce $P_c = |\Psi_H(0)|^2$ as a macroscopic scaling factor for each host crystal, Si or Ge. Here $\Psi_H(r)$ is the hydrogenic wave function which characterizes the ground-state wave function in presence of the simple Coulomb potential (13.3). When this is done, one finds that the remaining constant of proportionality required in (13.6) for donor impurities in a Ge host is about five times larger than it is when Si is the host. This is a satisfying result, because the conduction-band edge in Ge is of the symmetry type L_1, while that in Si belongs to Δ_1. Studies of these band edges, either in Ge–Si alloys or by hydrostatic pressure, have shown (Phillips, 1966) that the L_1 edge in Ge is indeed about five times more sensitive than the Δ_1 edge in Si. An attractive feature of this result is that it reduces the calculation of all eight values of $E^0 - E_H^0$ to two free parameters, while simultaneously avoiding the question of the detailed nature of the volume affected by the impurity potential and the impurity strain field.

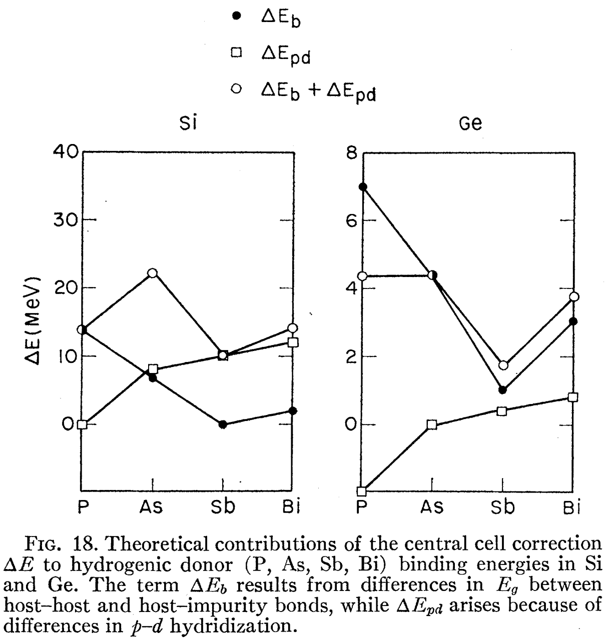


FIG. 18. Theoretical contributions of the central cell correction ΔE to hydrogenic donor (P, As, Sb, Bi) binding energies in Si and Ge. The term ΔE_b results from differences in E_g between host–host and host-impurity bonds, while ΔE_{pd} arises because of differences in p–d hydridization.

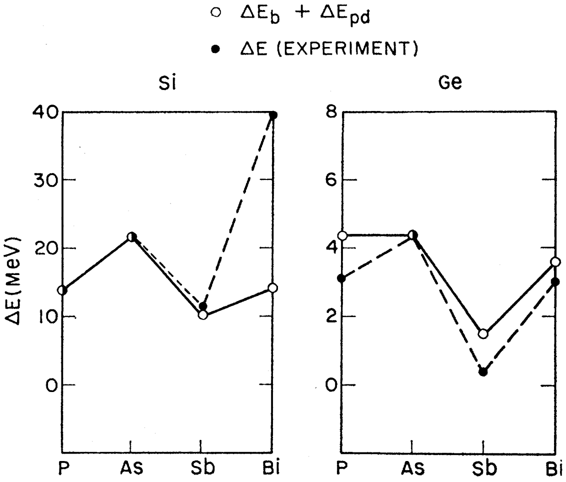


FIG. 19. Comparison of theory (see Fig. 18) and experiment for central cell corrections ΔE to hydrogenic donor binding energies. The agreement is good for all cases save Bi in Si, where there must be a giant contribution arising from differences in size between the Si core and the Bi core.

An interesting feature of (13.6) is that it is the first time that we have found a chemical trend which depends on E_g rather than on the ionicity phase angle φ. This is probably because this is the first case which involves a departure from having eight electrons per unit cell. Because of the compensating effects of the strain field, however, there remain nearly eight electrons per unit cell, i.e., $e_c^* \ll e$. With a greater departure from the normal complement of eight valence electrons, one would probably have to solve the entire quantum-mechanical problem, as one usually does in treating the effects of heteroatoms in hydrocarbons (Streitweiser, 1961). However, in the crystal the degree of localization is so small that a simple result such as (13.6) is not unexpected. Indeed the fact that ϵ_0 depends only on E_g, and not on φ, means that the appearance of only E_g in (13.6) qualitatively describes a central cell interaction which can be interpreted as arising from dielectric mismatch between the impurity and the host lattice. Such a concept is appropriate for *small* degrees of charge localization, because the dielectric constant describes the response of the crystal to weak electric fields.

Note added in proof: The spatial dimension of wave packets associated with shallow impurity states are large compared to an atomic radius. When these wave packets are transformed to momentum space (Kohn, 1957), the amplitudes are large only for crystal momentum very near the conduction-band edge (donor states, $\Delta Z = 1$) or valence-band edge (acceptor states,

$\Delta Z=-1$). Accordingly, one might expect that E_g, which represents conduction-valence-band energy differences averaged over the entire Brillouin zone, would not be so appropriate to the description of central cell corrections as band differences at the **k** corresponding to the appropriate band edge. This has recently been shown to be the case (Phillips, 1970c) for acceptors in Si and Ge.

14. CLASSICAL ELECTROSTATIC THEORY

The classical model of the polarizability of ionic crystals predates quantum mechanics. The model was introduced in the middle of the nineteenth century by Mosotti (1850) and Clausius (1879). A brief description of the theory as applied to alkali halide and related ionic crystals follows.

Each ion is idealized as a sphere of radius R of continuous uniformly polarizable dielectric. Allowing for depolarization effects the field E produced inside the sphere by an applied field F is

$$\mathbf{E}=[3/(\epsilon_0+2)]\mathrm{P}. \tag{14.1}$$

The electric moment $\mathbf{m}_s$ of the sphere is given by

$$\mathbf{m}_s=\alpha_s\mathbf{F}, \tag{14.2}$$

where α_s is the polarizability of the sphere. The volume of the sphere is $\Omega=4\pi R^3/3$, and $\mathbf{P}$ is the electric moment per unit volume such that

$$\mathbf{m}_s=\mathbf{P}\Omega=(\epsilon_0-1)\mathbf{E}\Omega/4\pi, \tag{14.3}$$

from which one obtains

$$(\epsilon_0-1)/(\epsilon_0+2)=(4\pi/3\Omega)\alpha_s=\alpha_s/R^3. \tag{14.4}$$

For a conducting sphere $\epsilon_0\rightarrow\infty$, and $\alpha_s\rightarrow R^3$. When $\epsilon_0\gg1$, (14.4) can be used to estimate R, and this was useful before x-ray and electron diffraction were available for precise structural determinations. When more than one kind of molecule is present in a total volume V, one has

$$\alpha=\sum\alpha_i=(3V/4\pi)(\epsilon_0-1)/(\epsilon_0+2) \tag{14.5}$$

as a natural generalization of (14.4). In writing the additive relation (14.5), we have assumed that the presence of other molecules does not alter the polarizability of any single molecule. This means that the charge densities of the molecules cannot overlap appreciably if (14.5) is to hold.

Empirically it has been found that the additivity assumption works well for the *total* polarizability (electronic plus ionic) in crystals such as those of the rocksalt type in which each ion is a center of inversion symmetry, at least with respect to its nearest neighbors. From a chemical point of view this is what one would expect, since the absence of inversion symmetry ordinarily is explained in terms of the formation of hybridized sp^3 or sp^2 directed valence orbitals. Such orbitals imply a sharing of valence electrons between atoms, and in that case the condition of nonoverlapping charges is violated.

If the total polarizability α_i^t of an ion is a constant independent of its environment, then one should have

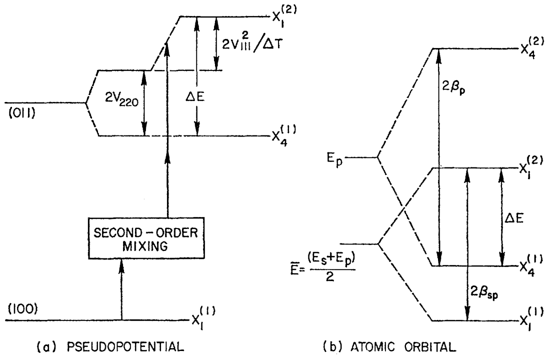
$$\alpha_{\mathrm{AC}}^t+\alpha_{\mathrm{BD}}^t=\alpha_{\mathrm{AD}}^t+\alpha_{\mathrm{BC}}^t=\alpha_{\mathrm{A}}^t+\alpha_{\mathrm{B}}^t+\alpha_{\mathrm{C}}^t+\alpha_{\mathrm{D}}^t. \tag{14.6}$$

For most alkali halides, (14.6) is valid to a few percent (Roberts, 1949).

There is still a problem in separating α_{AB}^t into α_{A}^t and α_{B}^t. So long as one considers only diatomic crystals, a constant can be added to all the anion polarizabilities α_{B}^t and subtracted from all cation polarizabilities α_{A}^t and leave the compound polarizabilities α_{AB}^t unchanged. Triatomic crystals such as Cu_2O resolve this ambiguity. One may also choose a reference value for one ion and show (Roberts, 1949) that this gives halide polarizabilities roughly proportional to their volume [cf. (14.4)].

A similar analysis of electronic polarizabilities only has been carried out by Tessman, Kahn, and Shockley (TKS) (1953) who found agreement for quite ionic crystals to within a few percent by assigning fixed polarizabilities to each ion. Their results were similar to those obtained earlier by Fajans and Joos (1924), Born and Heisenberg (1924), and Pauling (1927). Their work contained two new features:

(1) A statistical analysis was made of the suitability of the spherical or Lorentz–Lorenz depolarization factor which occurs in (14.1). More generally, (14.4) may be



(a) PSEUDOPOTENTIAL (b) ATOMIC ORBITAL

FIG. 20. A sketch showing the parentage of the energy-band levels in diamond-type crystals at $\mathbf{k}=\mathbf{X}$ in the Brillouin zone near the energy gap between the valence and conduction bands. The highest valence-band level is $X_4^{(1)}$; the lowest conduction-band level $X_1^{(2)}$. In (a) the levels are described in pseudopotential language; in (b) they are described in atomic orbital or tight-binding language. The levels $X_4^{(1)}$ and $X_1^{(1)}$ are each occupied by two electrons per atom. Thus (b) shows how the initial C atomic valence configuration of $2s^22p^2$ becomes $2s2p^3$ in diamond. On the other hand, (a) shows how simply the crystal energy gap ΔE between valence and conduction bands is calculated in pseudopotential language.

written

$$\alpha_s=\Omega[(\epsilon_0-1)/(L\epsilon_0-L+4\pi)], \qquad (14.7)$$

where L is a depolarization factor. One may write

$$L=\gamma 4\pi/3, \qquad (14.8)$$

and the value $\gamma=1$ is the Lorentz–Lorenz choice corresponding to (14.1). It was found that for the alkali halides, $\gamma=1$ did indeed give the best fit, but $\gamma=0.5$ or 2 raised the mean-squared error (normalized by the mean polarizabilities) by less than 15%. This is most disappointing, and it rather suggests that the success of the model rests in part on the large number of parameters used.

(2) Having determined "best" values for alkali and halide polarizabilities, an attempt was made to estimate divalent anion and radical polarizabilities in more complex structures by subtraction. Again rather variable results were obtained, but one striking result emerged for the $O^=$ ion. There the polarizability was found to vary by a factor of three. This variation was explained by taking the measured molecular volume in a number of oxides and subtracting the cation volumes as estimated from ionic radii, to yield effective $O^=$ ion volumes. As shown in Fig. 20, this yields quite good results, with $\alpha(O^=)$ linear in $\Omega(O^=)$ for about two-thirds of the cases studied.

The microscopic Clausius–Mosotti theory is now nearly 120 years old, and it is still robust and vigorous. Perusal of citations to TKS (1953) in the year 1968 by only those authors whose names begin with A or B shows two papers discussing surface relaxation of ionic crystals, one paper discussing the energetics of ionic powders, another calculating lattice energies of layer and chain transition-metal salts, another paper using dielectric properties to determine the solvated structure of $ZnBr_2$ in ether, another discussing collisions between atoms and diatomic molecules, and several others which were at the bindery. Thus the utility of the classical model is beyond question. We now attempt to understand this model in quantum-mechanical terms.

15. DISPERSION THEORY OF IONIC CRYSTALS

The dispersion theory of A^NB^{8-N} crystals discussed in Sec. 5 and applied to a variety of problems in other sections of this article starts from completely covalent diamond-type crystals in which the ionic energy C described by Eq. (5.20) is zero. One then treats crystals in which $C\neq 0$. It might have been expected that the treatment was valid only for predominantly covalent crystals for which $C\ll E_h$ and $f_i\ll 1$. In fact, C is defined by Eqs. (5.10) and (5.16), which are not based on perturbation theory. Thus E_h and C remain useful parameters for describing A^NB^{8-N} crystals up to about $f_i=0.9$, i.e., up to C about equal to $3E_h$. For more ionic crystals, dispersion theory based on the Penn model is probably not so useful.

Although the region $0\leq f_i\lesssim 0.9$ encompasses all tetrahedrally coordinated crystals and some crystals of the rocksalt type, there are many ionic crystals which lie outside this range, and there are also many ionic crystals which are not binary. In order to develop a meaningful spectroscopic theory of these crystals, one needs at least two observables which can be manipulated algebraically to determine optical model parameters. (For A^NB^{8-N} crystals, the two observables used were the lattice constant a and $\epsilon_1(0)$, which determined E_h and C, respectively.) It has been suggested (DiDomenico and Wemple, 1969; Wemple and DiDomenico, 1969) that the Sellmeier formula

$$\epsilon_1(\omega)=1+\mathfrak{F}/[\epsilon_0^2-(\hbar\omega)^2] \qquad (15.1)$$

may prove useful in this connection. Of course (15.1) is valid only for $\omega\lesssim\omega_1$, where ω_1 is the lowest frequency for which appreciable absorption associated with electronic transitions of the pure crystal take place. For each crystal there are already two parameters, $\mathfrak{F}$ and ϵ_0, which can be obtained by fitting experimental data to (15.1). Thus one does not need to have an ideal reference crystal (e.g., the diamond-type crystals), and one is not restricted to binary compounds.

At first one might expect to find no similarities between values of $\mathfrak{F}$ and ϵ_0 in such diverse materials as MgO and $Ba_2NaNb_5O_{15}$. However, all strongly ionic materials ($f_i\gtrsim 0.9$) are characterized by localization of most of the *s–p* valence electrons around the anions, $O^=$ in these two crystals. Moreover in each case the cation coordination number N_c is six. Thus the configuration seen by most of the $O^=$ valence electrons in $Ba_2NaNb_5O_{15}$ is not very different from that in MgO, which has the rocksalt structure. With a suitably chosen optical model, the differences between the two structures could turn out to be small.

The Wemple–Didomenico model (1969) for ionic crystals assumes that

$$\begin{aligned} \omega\epsilon_2&=4\pi\sigma_0 && \epsilon_g<\hbar\omega<b\epsilon_g, \\ \epsilon_2&=0 && \text{otherwise}, \end{aligned} \qquad (15.2)$$

i.e., a constant or rectangular conductivity, $\sigma_0=\omega\epsilon_2/4\pi$ over a frequency band between ϵ_g and $b\epsilon_g$. These two energies are related to ϵ_0 and a dispersion energy ϵ_d defined by

$$(\epsilon_d\epsilon_0)^2=\mathfrak{F}^2 \qquad (15.3)$$

by the equations derived from taking a Kramers–Kronig transform of (15.2):

$$\epsilon_d=\sqrt{3}8\sigma_0(b-1)/(1+b+b^2)^{1/2}, \qquad (15.4)$$

$$(\epsilon_0/\epsilon_g)^2=3b^2/(1+b+b^2)^{1/2}. \qquad (15.5)$$

The values of ϵ_0, ϵ_d, ϵ_g, and b are computed from the experimental data. The results for more than 50 crystals, most of them strongly ionic, are:

(1) ϵ_d is much more nearly constant within a group of chemically similar crystals than is either ϵ_0 or ϵ_g. In fact the relation

$$\epsilon_d = \beta N_c Z_a N_e \tag{15.6}$$

is found to hold with about 15% accuracy for all crystals containing a single anion species. Here Z_a is the valence of the anion, and N_e is the number of s–p valence electrons per anion. The energy parameter β has the value

$$\beta = \beta_i = 0.26 \pm 0.04 \text{ eV} \tag{15.7}$$

for ionic compounds (halides and oxides), and the value

$$\beta = \beta_c = 0.39 \pm 0.04 \text{ eV} \tag{15.8}$$

for more covalent nonhalide zincblende compounds. Again N_c is the cation coordination number.

(2) The results (15.6)–(15.8) can be reformulated in terms of a "broad" covalent optical band width

$$b_c = 3.4, \tag{15.9}$$

and a "narrow" ionic band width

$$b_i = 2.1, \tag{15.10}$$

and a "universal" conductivity constant

$$(\sigma_0 / N_c Z_a N_e) = 80 \pm 12 (\Omega \cdot \text{cm})^{-1}. \tag{15.11}$$

The striking feature of the analysis of Wemple and Didomenico is that the constant anion and cation polarizabilities of the Clausius–Mossotti model have been replaced by anion polarizabilities which exhibit dependences on the discrete environmental factors N_c and Z_a of the kind that one might expect from Pauling's resonating bond theory. Thus this model represents the first advance over the classical Clausius–Mossotti model in more than a century.

One may criticize the W–D model by noting that it is accurate to only 15%, whereas the C–M model is accurate to a few percent. However, as TKS (1953) shows, the accuracy of the C–M model stems primarily from the large number of parameters employed. The W–D model is accurate to 15% using three parameters for 50 crystals, so that with a score or more of free parameters it is not surprising that the C–M model works well. On the other hand, if the discrete variables N_c and Z_a of the W–D model could be replaced by functions of suitably determined continuous variables such as bond length and ionicity (as is the case for the dispersion theory of binary compounds), then very good accuracy might be achieved. This is a promising problem for further research.

16. MODELS AND MOMENTS

The dispersion theory discussed in the preceding section employs a model optical spectrum described by Eq. (15.2) which is quite different from the one associated with the Penn model illustrated in Fig. 1. The spectrum of the Penn model is quite complicated, but to a good approximation it is given by (Bardasis and Hone, 1967)

$$\epsilon_2(\omega) = \frac{16}{3a_0 k_F} \frac{E_F{}^2 E_g{}^2}{(\hbar\omega)^4} \frac{\hbar\omega}{[(\hbar\omega)^2 - E_g{}^2]^{1/2}}, \tag{16.1}$$

where $a_0 = \hbar/me^2$ is the Bohr radius, and k_F and E_F are the Fermi momentum and energy, respectively, of a free electron gas with density equal to that of the valence electrons. In (16.1) it is assumed that $\hbar\omega > E_g$; for $\hbar\omega < E_g$, of course, $\epsilon_2 = 0$.

Because (16.1) has been generated by an isotropic energy-band model, there is an analytic singularity in $\epsilon_2(\omega)$ of the one-dimensional type, i.e., as $\hbar\omega \to E_g$, $\epsilon_2(\omega)$ becomes infinite in proportion to $(\hbar\omega - E_g)^{1/2}$. This differs from the weaker singularities found in realistic three-dimensional band models (Phillips, 1966).

The question arises whether the model spectrum (16.1) or the model spectrum (15.2) provides a more useful basis for describing the over-all properties of the optical spectra of a broad class of crystals, be they semiconductors or insulators. This question is not quite the same as the much debated one as to the best choice of a set of basis functions in which to expand the trial function in the Schrödinger equation for a crystal or some other system. In the latter case, any set of complete functions used in the same equation should eventually give the same answer, so that the problem of choosing one representation over another reduces to the questions of which method converges more rapidly, or of which gives better answers at an early stage of calculation. It has often been supposed (but never demonstrated) that good answers at an early stage are indicative of a good choice of basis functions, whose matrix elements have particular physical significance.

Criteria of this type must be used to decide between different forms of model spectra. One of the physical properties which is characteristic of crystals of high polarizability [$\epsilon(0) \gg 1$] is that each atom is approximately neutral (see Sec. 3). This means that it is important that the model incorporate plasma oscillations which give rise to charge fluctuations, even though the plasma energy $\hbar\omega_p$ is much larger than the average energy gap E_g. Detailed comparison between experiment and theory on this point is hindered by the fact that most experimental data on semiconductors, for example, exhibit excess absorption near and above the plasma frequency because of contamination of the surface by an oxide layer. In this spectral region such a layer is much more absorptive than the semiconductor itself. There may also be many-electron effects (Bardasis and Hone, 1967) which enhance oscillator strength for ω above the plasma frequency ω_p, but these easily obscured by the surface oxide. In any event, it has been shown (Bardasis and Hone, 1967) that the Penn model gives a good qualitative account of the shape of $\epsilon_2(\omega)$ in semiconductors for $\hbar\omega \gtrsim 1.2E_g$. On the other hand, the Penn model fails completely to describe the absorption in spectral region $\hbar\omega < E_g$.

How important is the region $\hbar\omega < E_g$? The answer to this question depends on the physical observable being studied and the nature of the bonding. Obviously if one is interested in finding the energy E_0 of the threshold for intrinsic electronic absorption, then it is just the anisotropy of the energy gap (which is neglected entirely in the Penn model) which is of crucial importance. Because of the cubic symmetry of the crystal and the tetrahedral configuration of the nearest neighbors of each atom, there are many properties of semiconductors for which Penn's model is adequate. An exception is the heats of formation (Sec. 10), which are particularly delicate because they involve small differences between the cohesive energies of metals and the nearly equal cohesive energies of semiconductors. Here too the chemical trends in E_g are important, however, after allowance is made for the dehybridizing effects associated with the low-energy structure labeled E_0 and E_1.

The situation is different for the more ionic crystals. In the case of semiconductors, interactions between the electron and hole which are excited by photon absorption have little effect on $\epsilon(0)$, because the Coulomb interaction between the electron and hole is screened by $\epsilon(0)$. In particular, the oscillator strength associated with the formation of bound electron–hole states (called excitons) is small. Indeed in a hydrogenic model it can be shown (Elliott, 1970) that this oscillator strength is proportional to $[\epsilon(0)]^{-3}$.

In ionic crystals, $\epsilon(0)$ is much smaller. This greatly increases the oscillator strength of excitons, which are prominent (Phillips, 1966) in the spectra of crystals of the NaCl type. One can associate the exciton states with a shift of interband oscillator strengths to lower energies. Such a shift also occurs for all states in the continuum, and it can even produce resonances in the continuum (Phillips, 1966). In bonding language these spectroscopic effects are usually described in terms of greater localization of valence charge near the anions, a localization which becomes much more pronounced as the anion valence electron configuration approaches the closed-shell limit. In this limit, the electronic configuration may approximate for example Na^+Cl^-, which means that the neutral binding picture appropriate to more covalent crystals has been abandoned in favor of a classical valence picture. This may account for the appearance in (15.6) of the discrete variables N_c and Z_a which are important in Pauling's resonating bond theory.

Qualitative discussions of this type can be made more quantitative by considering spectral moments. Define the rth moment of the optical spectrum of a crystal by the relation

$$M_r = (2/\pi) \int_{\omega_g}^{\infty} \omega^{3-2r} \epsilon_2(\omega)\, d\omega. \tag{16.2}$$

In the following discussion use atomic units, i.e., $\hbar = 1$. In (16.2), ω_g is the threshold frequency for optical absorption.

One can insert either (15.2) or (16.1) into (16.2), compute the moments, and compare the two models algebraically. The f— sum rule (4.12) is written

$$M_1 = \omega_p^2, \tag{16.3}$$

and the Kramers–Kronig relation (4.6) gives for $\epsilon_1(0)$ the result

$$\epsilon_1(0) = 1 + M_2. \tag{16.4}$$

On the other hand, the parameters $\mathfrak{F}$ and ϵ_0 in (15.1) are given by (DiDomenico and Wemple, 1969; Wemple and DiDomenico, 1969)

$$\epsilon_d^2 = (\mathfrak{F}/\epsilon_0)^2 = M_2^3/M_3, \tag{16.5}$$

$$\epsilon_0^2 = M_2/M_3. \tag{16.6}$$

Thus the dispersion model of Sec. 5 involves M_1 and M_2, while that of Sec. 15 involves M_2 and M_3. Because $\omega_p^2 = 4\pi Ne^2/m$, where N is the valence electron density, one would expect that M_1 should be a necessary element in any general theory. The fact that this moment is omitted from the model of Sec. 15 may explain the somewhat limited range of applicability of the theory to primarily ionic crystals. The same is true of the Clausius–Mosotti theory.

The Penn spectrum (16.1) is derived from an energy-band model. This gives dispersion theory based on the Penn model a more solid foundation than would otherwise be the case, for some physical assumptions must be included in the theory beyond algebraic relations among the parameters of the theory and moments of the spectrum.

Hopfield (private communication) has derived an interesting relationship between the M_0 moment and the total *bare* ionic pseudopotential denoted by $V_b(\mathbf{r})$. Let the number of valence electrons per unit cell be n, and separate the valence charge density $\rho(\mathbf{r})$ into a constant part ρ_0 and a spatially varying part $\delta\rho(\mathbf{r})$. Then Hopfield's relation, derived from an argument involving translational invariance, is

$$M_0 = -\frac{\omega_p^2 e}{3mn} \int_{\text{all space}} \delta\rho(\mathbf{r})\, \nabla^2 V_b(\mathbf{r})\, d^3r. \tag{16.7}$$

By separating both $\delta\rho$ and V_b into symmetric and antisymmetric parts $\delta\rho_s$, $\delta\rho_a$, $V_b{}^s$, and $V_b{}^a$, respectively, one can rearrange (16.7) and utilize (16.3) to obtain

$$\langle\omega^2\rangle \equiv M_0/M_1 = -(e/3mn)\int(\delta\rho_s\nabla^2V_b{}^s+\delta\rho_a\nabla^2V_b{}^a)\,d^3r. \tag{16.8}$$

In pseudopotential theory (Phillips, 1970), one simplifies (16.8) further by transforming to **k** space. In this space, the Fourier components of $\delta\rho_s(\mathbf{k})$ are linearly related to those of $V_b{}^s(\mathbf{k})$, and similarly for $\delta\rho_a$. Thus (16.8) contains only terms of the form $V_b{}^s\nabla^2V_b{}^s$ and $V_b{}^a\nabla^2V_b{}^a$, with no cross terms or terms of higher than second order. This makes (16.8) closely resemble (5.10), the basic relation $E_g{}^2=E_h{}^2+C^2$ of dispersion theory.

This formal parallel may be appealing to readers seeking a more mathematical justification of the basic relation (5.10). Notice that the use of translational invariance to derive (16.7) is analogous to using the structure factors $\cos\mathbf{G}\cdot\boldsymbol{\tau}$ in (5.4). The condition of translational invariance is customarily used in lattice dynamics to prove that phonon frequencies ω tend to zero like ck, where c is a suitable sound velocity. However, c depends (Phillips, 1970) on ion–ion forces as well as electron–ion forces of the type (16.7), so that there is no direct relation between M_0 and the interatomic forces discussed in Sec. 11.

17. DISPERSION THEORY AND BAND THEORY

Most studies of the electronic structure of crystals have utilized the full symmetry of the crystal to classify and describe states in the one-electron approximation. In this approximation the states form energy bands specified by $E_n(\mathbf{k})$, where **k** is the crystal momentum or eigenvalue associated with translational invariance, and n is a band index. By contrast valence bond theory describes certain properties of the crystal which involve *averages* over the one-electron band states. Although quantum mechanics gives us a prescription for obtaining the band states, at least after making plausible simplifications of the one-electron type, there is no general set of rules governing the *kind* of average that should be used to determine valence bond parameters.

The situation is similar (Phillips, 1970b) to the one involved in deriving thermodynamic relations from classical mechanics. Newton's laws do not contain any prescription for determination of thermodynamic variables such as entropy. The connection between these two topics is provided only by an entirely new approach, statistical mechanics.

The view taken here is that the proper choice of averages depends on the kind of bonding involved. For example, different approaches are probably required to treat binary metallic compounds AX, where A is a nontransition element and X is a transition element, than are required for binary compounds AB. Even among the binary nontransition compounds A^NB^{M-N}, it appears that different values of M give rise to quite different structural properties.

All of this is in accord with classical valence theory, much of which is unfamiliar to most physicists who have studied the electronic structure of solids. What has been shown by dispersion theory is that at least for the large family of semiconductors and insulators with $M=8$, the variable N is not important. Indeed in the ionicity range $0\leq f_i\lesssim 0.9$, even the question of nearest-neighbor coordination number is not essential, as is shown by extending (Van Vechten, 1969; Penn 1962) the treatment of bond ionicity in tetrahedrally coordinated crystals to crystals of the sixfold coordinated NaCl type.

In spite of the success of dispersion theory in phenomenological terms, many theorists may welcome a discussion relating the results to those of band theory. It should be noted in advance, however, that band theory in its present state cannot tell us what kind of average we should take. It may, however, for a given average give a more explicit relation between the bond parameters and the ionic potentials.

A. Pseudopotentials and the Jones Zone

The feature that distinguishes metals from nonmetals in the energy-band picture is that in the case of nonmetals there is an energy gap between the valence bands (which are full) and the conduction bands (which are empty). In the language of Brillouin zones, the Fermi surface is congruent with the mth Brillouin-zone boundary, where $m=M/2$, and M is the number of valence electrons per unit cell. The volume of **k** space so occupied is called the Jones zone (Heine and Jones, 1969).

From Fig. 3 one can see that in diamond-type crystals the largest peak in $\epsilon_2(\omega)$, labeled E_2, falls quite close to the average energy gap E_h introduced in dispersion theory. The same holds true for E_2 and E_g for zincblende and wurtzite crystals (Phillips, 1970), and this coincidence apparently reflects the suitability of an sp^3-hybridized bond model for describing these crystals. Heine and Jones (1969) have proposed to relate the energy of the peak E_2 to the actual energy levels $E_n(\mathbf{k})$ by making the following approximations:

(i) E_2 is identified with the energy $E(X_1)-E(X_4)$. Here $\mathbf{k}=X$, and X_1 and X_4 label the lowest conduction and highest valence bands at X. The justification for this is that the Jones zone in this case is a dodecahedron, and X lies at the center of each face. Actually band studies have shown (Van Vechten, 1970) that $E(X_1)-E(X_4)$ is always 0.2–0.3 eV below E_2. This is a small error compared to the basic one of identifying E_2 with E_g, and in itself need cause no concern.

(ii) The energies $E(X_1)$ and $E(X_4)$ would be degenerate in the free-electron limit with the crystal potential $V=0$. This suggested that one carry out quasidegenerate perturbation theory to estimate

$$E(X_1)-E(X_4)=\Delta E\approx E_2\approx E_h. \qquad (17.1)$$

There are four plane waves at X which must be treated exactly: $(0, \pm1, \pm1)$ in units of $2\pi/a$. Two others $(\pm1, 0, 0)$ are treated by second-order perturbation theory. The result is expressed in terms of $V(\mathbf{G})$, which is the pseudopotential form factor evaluated at the reciprocal lattice vector $\mathbf{G}$. One finds

$$\Delta E=2[V(220)+\{V(111)\}^2/\Delta T], \qquad (17.2)$$

where the difference in free electron energies is

$$\Delta T=(\hbar^2/2m)(2\pi/a)^2[(110)^2-(001)^2]. \qquad (17.3)$$

The values of ΔE calculated from values of $V(\mathbf{G})$ adjusted to fit experiment and obtained by solving a *complete* secular equation (about 100 plane waves) agree fairly well (Heine and Jones, 1969) with those obtained from (17.2).

The first term in (17.2) is the only one that is retained in simplified theories of metals (Phillips, 1970). The second term is the one characteristic of covalent bonding. The Heine–Jones model shows how covalency enters pseudopotential theory.

One can ask whether it is possible to extend this model to include ionic effects as well. To do so one introduces symmetric and antisymmetric pseudopotentials (Cohen and Bergstresser, 1966; Walter and Cohen, 1969) according to

$$2V_s(\mathbf{G})=V_\mathrm{A}(\mathbf{G})+V_\mathrm{B}(\mathbf{G}), \qquad (17.4)$$

$$2V_a(\mathbf{G})=V_\mathrm{A}(\mathbf{G})-V_\mathrm{B}(\mathbf{G}). \qquad (17.5)$$

The structure factor for $\mathbf{G}=(220)$ is zero for V_a, and is 90° out of phase for $V_a(111)$ compared to $V_s(111)$. Thus if by ΔE we mean the energy difference between the average of the two conduction-band energies and the two valence-band energies, the generalization of (17.2) is $\Delta E=|\Delta E_1+i\Delta E_2|$ with

$$\Delta E_1/2=V_s(220)+\{V_s(111)\}^2+\{V_a(111)\}^2/\Delta T, \qquad (17.6)$$

$$\Delta E_2/2=2V_s(111)V_a(111)/\Delta T. \qquad (17.7)$$

In the limit $V_a\rightarrow0$ we see that $\Delta E\rightarrow\Delta E_1$ which is to be identified with E_h. Thus ΔE_2 is to be identified with C. From (17.7) we see that ΔE_2 is odd with respect to interchange of atoms A and B because it is first order in V_a. One can evaluate (17.7) numerically very simply by noting that $\Delta T=(2\pi a_0/a)^2$ Ry, where a_0 is the Bohr radius, a the cubic lattice constant, using a_0/a from (Van Vechten, 1969; Penn, 1962), and taking $V_s(111)$ and $V_a(111)$ from (Cohen and Bergstresser, 1966; Walter and Cohen, 1969). The result for GaAs is $\Delta E_2=2.5$ eV, compared to $C=2.9$ eV as determined spectroscopically.

A more severe test of the utility of the truncated expressions (17.6) and (17.7) is obtained by expanding $(\Delta E)^2$ in powers of $C^2=\Delta E_2^2$

$$(\Delta E)^2=(\Delta E_0)^2+a_2(\Delta E_2)^2+a_4(\Delta E_2)^4, \qquad (17.8)$$

where a term-by-term comparison with (17.6) and (17.7) gives

$$\Delta E_0=V_s(220)+\{V_s(111)\}^2/\Delta T, \qquad (17.9)$$

$$a_2=1+\Delta E_0\Delta T/8\{V_s(111)\}^2, \qquad (17.10)$$

$$a_4=(\Delta T)^2/\{4V_s(111)\}^4. \qquad (17.11)$$

A careful survey of the behavior of all optical peaks in diamond and zincblende crystals as a function of C^2 has been made (Van Vechten, 1970). The result for the E_2 peak is

$$E_2^2=E_{2h}^2+C^2. \qquad (17.12)$$

This corresponds in (17.8) to

$$a_2=1, \qquad a_4=0. \qquad (17.13)$$

At first one is inclined to dismiss the differences between (17.10), (17.11) on the one hand, and (17.13) on the other, as artifacts of the truncation process [confining the calculation to (110) plane waves perturbed to second order by (001) plane waves]. However, there is a good deal more to be said. Notice that the additional terms in (17.10) and (17.11) both increase ΔE above what is found experimentally and described by (17.13). If one compares ZnSe with GaAs, one experimentally finds $C(\mathrm{ZnSe})=2C(\mathrm{GaAs})$. [Other horizontal sequences such as BN–BeO and InSb–CdTe also give $C(\text{II–VI})=2C(\text{III–V})$.] However, the extra terms in (17.10) and (17.11) increase ΔE more rapidly for IV–VI compounds than for III–V ones. This means that if the band calculation were to be stopped in the $(110)+(001)=(*)$ subspace the values of $C^*(\mathrm{ZnSe})$ and $C^*(\mathrm{GaAs})$ required to fit the observed values of E_2 would satisfy the inequality

$$C^*(\mathrm{ZnSe})/C^*(\mathrm{GaAs})<2. \qquad (17.14)$$

The band calculations of Cohen and Bergstresser (1966) on the other hand, have been carried to convergence (∞ subspace). Because $V_s(111)$ and ΔT are the same in GaAs and ZnSe, ΔE_2 and effectively C are proportional to $V_a(111)$, which takes the values 0.07 and 0.18 Ry, respectively. Thus in an obvious notation

$$C^\infty(\mathrm{ZnSe})/C^\infty(\mathrm{GaAs})=2.3. \qquad (17.15)$$

When we compare (17.14) and (17.15) with the experimental ratio derived by dispersion theory

$$C^{\mathrm{exptl}}(\mathrm{ZnSe})/C^{\mathrm{exptl}}(\mathrm{GaAs})=2.0, \qquad (17.16)$$

we arrive at a very interesting result. By choosing as truncation parameter the size of our subspace, it should be possible to adjust this in such a way as to satisfy (17.16). The size of the subspace should be greater than the * subspace (Heine and Jones, 1969), but less than the ∞ subspace (Cohen and Bergstresser, 1966; Walter and Cohen, 1969).

The conclusion that (17.16) can be satisfied in this way is of particular interest if we choose to take the values of $V_{\mathrm{A}}(\mathbf{G})$ and $V_{\mathrm{B}}(\mathbf{G})$ not from a fit to semiconductor spectra, but instead from free atom term values as given in tables by Animalu and Heine (1965). These values also satisfy (17.16). This suggests that there may be something inherently unsound about the semiempirical pseudopotential form factor values $V_\alpha{}^\infty(\mathbf{G})$, where $\alpha=\mathrm{A}$ or B, and the superscript ∞ means that adjustments to experiment have been made based on convergent secular equations.

The use of a partially truncated plane-wave expansion in pseudopotential calculations is not new. In the original work (Phillips, 1958) on diamond-type crystals, the expansion was confined to about 30 plane waves, partly for reasons of computational economy and partly because this was the smallest subspace which gave accurate results for all $E_n(\mathbf{k})$ of interest. More recently studies of $\mathrm{Ga}_x\ \mathrm{In}_{1-x}$ As alloys appeared to indicate (Jones and Lettington, 1969) that the Heine–Animalu atom-based form factors gave much better results in a truncated subspace than in a convergent one. Again the optimal cutoff was found to be at about 30 plane waves. However, in both cases the results are rather fragmentary.

Is there anything special about 30 plane waves? Apparently this is about the number of plane waves required to synthesize wave functions for states near the energy gap in the region outside the atomic cores. Adding further plane waves "improves" the wave functions in two regions of the crystal: in the region where the atomic potentials overlap most strongly, which is where the atomic radii touch, and also in the core regions. This second effect however is not an improvement at all. Particularly in pseudopotential calculations but actually in all other approaches as well, it is assumed that the regions of the atomic cores are the same in the crystal as in the free atoms. For homopolar semiconductors or metals this assumption never causes any great difficulty. For partially ionic crystals, however, problems do arise.

The point is that the energy level shifts in a calculation based on $V_a{}^\infty(\mathbf{G})$ relative to $V_a{}^{30}(\mathbf{G})$ are largely based on charge transfer from the region of the cation core to the region of the anion core. This effect is probably spurious, because through the exclusion principle the ion core electrons keep the valence electrons from accumulating in the core regions. In empirically adjusted orthogonalized plane-wave calculations, it has been found (Stukel *et al.*, 1969) that satisfactory results can be achieved only through the introduction of so-called "core (energy) shifts." So far no explanation of these core shifts has been given, but from the present point of view it appears that such shifts may be used to simulate the changes in the cores which take place to prevent valence charge transfer.

The foregoing remarks can be used to formulate a program for treating the effects of ionicity on crystalline energy bands. First notice that truncation of the secular equation to a fixed number of plane waves has certain disagreeable features, in particular it introduces different apparent degrees of convergence at different points $\mathbf{k}$ in the Brillouin zone. This difficulty can be minimized as follows. Truncation of the function space to N plane waves is equivalent to replacing the off-diagonal matrix elements H_{ij} in the original secular equation by $H_{ij}\theta_j\theta_i$, where i and j label reciprocal lattice vectors $\mathbf{G}_i$ arranged in some sequence. Here $\theta_i=1$ for $i\leq N$, and $\theta_i=0$ for $i>N$. One may refine this process by using the cutoff function

$$\begin{aligned}\theta(x)=\theta_i(\mathbf{G}_i)&=1 \quad \text{if} \quad x=|\mathbf{k}+\mathbf{G}_i|^2\leq\epsilon_c\\ &=0 \quad \text{if} \quad x=|\mathbf{k}+\mathbf{G}_i|^2>\epsilon_c,\end{aligned} \tag{17.17}$$

where $\hbar^2\epsilon_c/2m$ is a cutoff energy.

The weakness of (17.17) is its discontinuous character. Therefore, replace the step function $\theta(x)$ by a smooth function such as the Fermi function. In such a function, one could use ϵ_c for both the Fermi energy and a multiple g of the Fermi temperature, i.e., replace (17.17) by something like

$$F(\mathbf{G}_i)=\{1+\exp[g(x-\epsilon_c)/\epsilon_c]\}^{-1}, \tag{17.18}$$

where g is 5 or 10. One then adjusts ϵ_c and g to reproduce (17.16) as well as possible.

The reason that this point has been discussed at such length is that it shows how dispersion theory can be used to derive a criterion for efficiently introducing self-consistency into energy-band calculations in partially ionic crystals. The discussion given here has been rather abstract, and there are no doubt some readers who will feel that it has also been unnecessarily complicated. Why not settle all these matters simply by looking at atomic charge densities and representing the wave functions by atomic orbitals? What could be simpler?

In fact, the straightforward atomic approach is not simple at all, and it soon gets bogged down in convergence problems and unphysical parameters. (We have already seen an example of this in Sec. 7 and Table V.) As a result the literature on the atomic approach is much less coherent than is the pseudopotential literature (Cohen and Heine, 1970). For the benefit of readers who would appreciate a short summary of the difficulties involved, a partial discussion is given in the following sections.

B. Atomic Potentials and Energy Levels

It has been customary in almost all energy-band calculations to use the one-electron approximation in setting up the crystal potential. Of course everyone knows that the effective potential seen by valence electrons is not merely the average Coulomb potential of the other electrons. To achieve the level of accuracy required to reproduce differences in electronic structure from one crystal to another (e.g., Si and Ge, or GaP and GaAs) one must also make allowance for the exchange interactions between electrons of parallel spin required by Fermi statistics, and dynamical correlations between electrons of both parallel and antiparallel spins which at high energies give rise to collective plasma oscillations.

Unfortunately no prescription is known at present for treating these effects accurately in covalent systems. Even if one were known, incorporating it into actual one-electron calculations would be difficult, because exchange and correlation effects are described by a nonlocal energy-dependent potential $\mathcal{V}(\mathbf{r}, \mathbf{r}', \omega)$, whereas for practical purposes one-electron calculations use a local, energy-independent potential $V(\mathbf{r})$ corresponding to $\mathbf{r}' = \mathbf{r}$, and $\omega = 0$.

Some idea of the magnitude of these effects in covalent systems can be obtained by examining the ionic terms in the potential more closely. In molecular orbital theory, as we saw in Sec. 7, efforts have been made to incorporate exchange and correlation energies into estimates of V_{ionic} by using ionization energies and electron affinities of free atoms. These are nearly all polyvalent or open-shell cases, and consequently somewhat misleading. For whenever an electron is ionized from a polyvalent atom, the remaining electrons alter their orbitals considerably, and their energies are lowered. Moreover, in the original orbitals electron–electron interactions are much greater than in the crystal, because in the latter case the electrons are delocalized and their interactions are reduced by plasma screening by the other delocalized valence electrons.

Because the one-electron approximation works so well in crystals, better results are perhaps obtainable by replacing the ionization energies used to estimate $\alpha_A(0) - \alpha_B(0)$ by differences in one-electron atomic energies. These energies, calculated with the so-called Slater exchange terms included, are given in (Herman and Skillman, 1963). When one tries to use these energies, one immediately encounters the problem of whether to use the *ns* or *np* valence energy. Because the levels near the top of the valence band are of *p* character, the *p* valence energy seems more suitable. For atoms from Columns I or II this requires calculations of excited states, not given in Herman and Skillman (1963). We therefore quote values of $\alpha_A(0) - \alpha_B(0)$ in two $A^N B^{8-N}$ crystals with $N = 3$. These two cases are BN and GaAs, and they are sufficient to illustrate what is happening. We let ΔE_{AB} represent $E_A(np) - E_B(np)$, and compare this with $\alpha_A(0) - \alpha_B(0) = \Delta\alpha_{AB}$ (cf. Table V) and C_{AB}. The results for BN are

$$\Delta\alpha_{AB} = 16.5 \text{ eV}, \quad \Delta E_{AB} = 9.5 \text{ eV}, \quad C_{AB} = 7.8 \text{ eV}. \tag{17.19}$$

The values of C_{AB}, which are based on observed spectra, are self-consistent thanks to nature herself. Thus the one-electron value ΔE_{AB} looks quite good; unlike $\Delta\alpha_{AB}$ it is already almost self-consistent. This may be an accident, however, so consider the numbers for GaAs,

$$\Delta\alpha_{AB} = 11.6 \text{ eV}, \quad \Delta E_{AB} = 6.0 \text{ eV}, \quad C_{AB} = 2.9 \text{ eV}. \tag{17.20}$$

Again ΔE_{AB} is about half as large as $\Delta\alpha_{AB}$, and again it is closer to being self-consistent. However, in this case $C_{AB} \approx \Delta E_{AB}/2 \approx \Delta\alpha_{AB}/4$, so that for larger atoms the one-electron estimates of V_{ionic} are still far from being self-consistent.

If one now considers other compounds, the same general pattern holds, as one could have inferred from the formula (5.20) for C_{AB}. The agreement between ΔE_{AB} and C_{AB} is best for small first-row atoms and poorer for larger atoms from other rows. As suggested by (5.20), screening plays an important role, and for first-row atoms this is similar in the crystal as in the free atom. For other rows this is not the case, presumably because the larger atoms are more polarizable.

C. Atomic Orbitals

If we consider the two quantities E_h and C, we see from (5.20) that C has a simple semiclassical interpretation in terms of differences in potential of screened point ions. On the other hand, E_h is a measure of a kind of interference effect produced by V_{covalent} which leads to an energy gap between bonding and antibonding states. Although the dependence of $E_h \propto a^{-2.5}$ on the lattice constant a given by (5.19) is quite simple, no semiclassical interpretation of this behavior has yet been found. This might lead us to think that, in view of the difficulties the atomic picture encounters in estimating C quantitatively, it might even fail to aid us in understanding qualitatively the effects of V_{covalent}. This indeed proves to be the case: billiard ball models tell us little about interference effects and thus little about covalent bonding.

The best description of the atomic orbital approach to the electronic structure of crystals has been given by Slater and Koster (1954). The electronic energy levels are described in terms of overlap integrals between atomic orbitals which involve the crystal potential which is represented as a sum of atomic potentials. There are two kinds of such overlap integrals, depending on whether the orbitals and the atomic potential have a common center or not. In the latter case it is assumed that the largest terms arise when the two orbitals are centered on nearest-neighbor atoms, and the potential is also centered on one of the atoms. The average energy for a given band is determined by the one-center term, and its $\mathbf{k}$ dependence is determined by the two-center terms.

In realistic calculations based on atomic orbitals it has been found that a great many two-center terms are required to produce satisfactory convergence for $E_n(\mathbf{k})$ in covalent crystals. For this reason Slater and Koster (1954) suggested that an alternate approach would be to determine empirically the overlap parameters for the first few neighbors. This proposal has since been carried out (Dresselhaus and Dresselhaus, 1967) but the resulting parameters in the only cases studied (Si and Ge) were rather unphysical (Phillips, 1970). One can demonstrate this point in more analytic terms by studying the form in which ionic effects appear in the overlap parameters.

In order to make the discussion specific and also to facilitate comparison with results previously discussed, attention is focused on the energy levels at $\mathbf{k}=\mathbf{X}=(2\pi/a)(1,0,0)$. From the viewpoint of atomic orbitals $\mathbf{X}$ is significant because it makes equal angles with all (111) valence bonds. Thus it constitutes the Bloch analog of hybridized sp^3 orbitals. Moreover, the valence level (labeled X_4 in the diamond structure, X_5 in the zincblende) contains p_y and p_z valence orbitals, while the conduction level (labeled X_1 in the diamond structure and twofold degenerate there, and labeled X_1 and X_3 in the zincblende structure) contains s and p_x orbitals. It is a special consequence of the symmetry of the diamond structure which does not hold for zincblende cases that these s and p orbitals are degenerate in the diamond structure. However, the energy gap that concerns us is

$$\Delta E=E(X_1)-E(X_4) \qquad \text{(diamond)},$$
$$2\Delta E=E(X_1)+E(X_3)-2E(X_4) \qquad \text{(zincblende)}, \qquad (17.21)$$

which means that so far as ΔE is concerned, we can treat the conduction levels in zincblende as degenerate as well.

To make clearer the connection between the pseudopotential and atomic orbital descriptions of levels at X, the two pictures are compared in Fig. 20. First, note that the scale of energy is about twice as large in the pseudopotential picture as in the AO one. Second, note that while ΔE arises quite simply in the Jones zone, or nearly free electron picture, through splitting of degenerate states, the AO description is much more complex. This is because the atomic energy levels E_s and E_p are such that the ground-state configuration of a four-valent atom is s^2p^2, whereas the crystalline configuration is sp^3. This means that as the atoms are brought together the s and p levels split, and a level crossing must occur. This is exactly what happens at X, as shown in Fig. 20(b). The splitting parameter for the p_y, p_z levels is denoted by β_p. One can see from the fact that $\mathbf{X}$ makes equal angles with (111) bonds that

$$\beta_p=\{(pp\sigma)+2(pp\pi)\}/3, \qquad (17.22)$$

while the splitting parameter β_{sp} involves the more complicated combination

$$\beta_{sp}=\{(ss\sigma)+\beta_p+2(sp\sigma)\}/2. \qquad (17.23)$$

The notation for overlap integrals in (17.22) and (17.23) is the standard one (Slater and Koster, 1954).

What happens to the energy level scheme shown in Fig. 20(b) when we pass from the AA diamond to the AB zincblende structure? This question has not been studied systematically because the AO representation is so clumsy. However, qualitatively one can see by comparison with the discussion of Sec. 3 that β_p and β_{sp} should change little, but E_p and $\bar{E}=(E_s+E_p)/2$ should split into $E_p{}^{\mathrm{A}}$, $E_p{}^{\mathrm{B}}$ and $\bar{E}^{\mathrm{A}}$, $\bar{E}^{\mathrm{B}}$ so that we have an X_4 and an X_1, 2×2 matrix to solve, analogous to Eqs. (3.3)–(3.5).

It follows then that the ionic contribution to ΔE should appear indirectly through the ionic contributions to E_p and $\bar{E}$. Clearly, each of these energies can be expanded to order C^2, but equally clearly when this is done one will find terms in ΔE of order C^2, C^4, C^6, etc. Thus the basic experimental fact that ΔE depends only on C^2 and not C^4, C^6, etc., is not apparent in the AO approach. Indeed because of the level crossing shown in Fig. 20(b), ΔE appears to be a sensitive difference of several large and nearly equal energies E_s, E_p, $ss\sigma$, $sp\sigma$, $pp\sigma$, and $pp\pi$. In the AO representation, the fact that ΔE behaves so simply as a function of C^2 seems to be nothing short of miraculous.

From the physical point of view it should not be surprising that the description of crystal bonding in terms of atomic orbitals is so poor. After all, the specific features of atomic orbitals arise from the spherical symmetry of the atom. But the covalent bonds which we seek to understand represent directional effects which are not present in the free atom. On the other hand, plane waves in the crystal are at least already partially directional, which makes possible the relative simplicity of the energy level scheme in Fig. 20(a) compared to Fig. 20(b).

D. Phase Shifts

The difficulties encountered by the atomic orbital method have been well known for more than 30 years. In 1937 Slater proposed a method that appeared to overcome most of these difficulties. Spheres are inscribed around each atom A of radius r_{A}. Inside the spheres wave functions of atomic (spherical) symmetry are used to represent the wave function, while outside the spheres plane waves are used. The eigenvalue condition is that Ψ'/Ψ is continuous at the spherical boundary.

Slater's idea was later cast into the language of scattering theory by Korringa (1947), and later Kohn and Rostoker (1954) (KKR). By using partial-wave expansions, KKR were able to present the eigenvalue condition in L (angular momentum) space, and evaluate the matrix elements of their secular equation using Ewald summation techniques. Because the geometrical basis of Slater's augmented plane waves (APW) and the KKR scattering or Green's function method are the same, it has been possible to show that the two methods are essentially equivalent (Ziman, 1965).

If the KKR method is used with real spherical harmonics, the matrix elements W_{LL} of the secular determinant have the form (Ziman, 1965)

$$W_{LL'}=A_{LL'}+K\cot\eta_l\delta_{LL'},\tag{17.24}$$

where the phase shift η_l is that associated with a single atom A with potential truncated at $r=R_A$. (For simplicity one ordinarily considers only monatomic crystals.) Here $K=\bar{E}^{1/2}$ in atomic units, and the structure of $A_{LL'}$ involves reciprocal lattice vectors $\mathbf{G}$, Bessel functions, and spherical harmonics.

At first sight the form of (17.24) is rather appealing. Our old friends, the η_l, are present, and perhaps we could introduce symmetric and antisymmetric combinations

$$\eta_l^s=(\eta_l^A+\eta_l^B)/2,\tag{17.25}$$

$$\eta_l^a=(\eta_l^A-\eta_l^B)/2,\tag{17.26}$$

to describe covalent and ionic effects. We again concentrate on $\mathbf{k}=\mathbf{X}$ and attempt to compute $\Delta E=E(X_1)-E(X_4)$ in the approximation that only terms with $\mathbf{k}+\mathbf{G}=(011)$ or (100) are retained. This recovers the plane-wave simplicity of Fig. 20(a), while retaining the phase shifts η_l to describe the atomic potentials.

The snag in the prescription is that we need to know $j_l(|\mathbf{k}-\mathbf{G}|R_A)$ for each $\mathbf{G}$ and for each atom A. Of course in principle we can take these numbers from tables, but in practice if we seek *analytic* expressions for ΔE similar to (17.8), etc., we soon find ourselves bogged down in algebra. Thus for monatomic transition metals where one has to do with a d-band resonance crossing a nearly free electron s–p band, a great deal of algebraic manipulation is required to establish a simple analytic picture (Heine, 1967; Jacobs, 1968). In a diatomic lattice where the energy level diagram is more complicated (Fig. 20) a great deal of effort would be required to separate covalent and ionic effects. It would be problematical whether this would lead to a relationship of the $\Delta E^2=E_h^2+C^2$ type. What would be more likely would be a criterion for choosing r_A and r_B, which would be interesting but not essential. Another problem implicit in the phase shift approach is determination of the zero of energy used to define $K=\bar{E}^{1/2}$. This represents some kind of average of the crystal potential in the region outside the atomic spheres, and how this average would depend on Z_A, r_A and Z_B, r_B is not known. Calculations to date have treated the zero of energy as an adjustable parameter which gives rise to rather large discontinuities in potential (of the order of 1 Ry) at the atomic spheres (Phillips and Sandrock, 1968).

18. PROSPECTS FOR FUTURE WORK

Predictions of the course of future research are unlikely to be accurate if the problems involved are as difficult as those in solid state physics. Nevertheless to conclude this article it seems appropriate to review what is known so far and what the implications of this knowledge are for future work.

Anyone who surveys the properties of crystals is immediately struck by the many interesting relations and trends which are obvious even in the macroscopic properties. Our knowledge of the quantum structure of crystals has increased very greatly in the last decade, and now trends are apparent in these properties as well. In this article we have attempted to relate both macroscopic and quantum trends largely to trends in ionicity. We have been successful in discussing many properties, and our success confirms many of the assumptions of valence bond theory as formulated by Pauling and by Coulson. We have been able to discuss band structures, ionization potentials, and work functions (Van Vechten, 1970); nonlinear optical properties (Phillips and Van Vechten, 1969), alloys (Van Vechten and Bergstresser, 1970; Jones and Lettington, 1969), heats of formation and cohesive energies (Phillips and Van Vechten, 1970), interatomic forces (Martin, 1970), effective charges, impurity energies (Phillips, 1970a), and so on.

The discussion so far has focused chiefly on tetrahedrally coordinated A^NB^{8-N} semiconductors. This is because the experimental data are far more complete for this family of materials. But as shown in Mooser and Pearson (1959) and Pearson (1962), regularities similar to those of the A^NB^{8-N} family are found in many other binary compounds.[3] Although the prescriptions employed for A^NB^{8-N} compounds cannot be applied without further thought to these other families, it would appear that there are lessons to be learned from the case about which we have the most information, and that some of these lessons can be applied to the other families as well.

In closing, I should like to direct an appeal to younger theorists to involve themselves in problems of this kind. The number of man-years spent by younger

---

[3] See the interesting collection of semiempirical models discussed in Rudman, Stringer, and Jaffee (1967).

theorists on formal problems such as the Kondo effect or the properties of the free electron gas must now number in the hundreds. Such mathematical exercises eventually take on a Byzantine quality which enriches neither the practitioners nor the scientific community as a whole. On the other hand the richness of patterns of crystal structures and properties is now accessible to explanation in quantum-mechanical terms by those who are willing to consult and survey experimental data on a wide scale. It is my hope that this article has made work of this kind seem attractive and interesting.

Postscript: I am often asked the question: Can the spectroscopic ideas developed here for crystals be used to discuss bonding in molecules? Because of the structural similarities between crystals and molecules, the answer in a larger sense must be yes. Speculations about how this might be done can be found in Ref. 1. There is a close similarity in over-all view between my ideas and those of Pauling (Ref. 2), but his empirical basis is thermochemical, whereas I believe the spectroscopic approach is more accurate for crystals.

If one tries to apply the spectroscopic approach to molecules, one immediately encounters difficulties connected both with paucity of data and with the definition of bond polarizabilities in molecules containing more than one kind of bond. (Similar problems arise in ternary, quaternary, etc., crystals.) Traditionally, these problems have been resolved for qualitative purposes by assuming that bond polarizabilities are additive and transferable. These two assumptions are only roughly valid, and their use tends to vitiate many of the advantages of the spectroscopic approach.

It appears that there is no universal prescription which can be applied to all structural problems. Perhaps the best way to appreciate the significance of the present approach is to regard the A^NB^{8-N} crystals as a kind of ideal system. This system is used to give an absolute definition of ionicity, just as the ideal gas is used to construct an absolute scale of temperature. (See my article in *Physics Today*, February 1970, p. 23.) In measuring temperature using a thermometer made out of real materials, one must make corrections to the Kelvin scale for calibration purposes. Working out these corrections is a serious task in each case.

On a more practical level, there is one important lesson to be learned from the present work. In estimating ionic effects, it is natural to calculate an ionic energy. In many cases there is a tendency to overlook the covalent energy, which is equally important, but which has no simple classical meaning. The results for A^NB^{8-N} crystals demonstrate how important it is to treat *both* effects *on an equal footing*.

APPENDIX

Can spin–orbit splittings serve as an accurate measure of ionicity?

The pedagogical object of this review has been to make the subject of ionicity in crystals *appear* simple. Treated in the proper spectroscopic perspective the *results* are indeed strikingly simple. However, some of my colleagues have drawn the mistaken conclusion that the subject itself really *is* simple. The purpose of this Appendix is to disabuse the reader of such a naive notion by making explicit some of the real quantum-mechanical problems inherent in attempts to define ionicity which arise with definitions based on over-simplified atomic models. In order to make the discussion specific, I shall use as an example spin–orbit splittings at $\mathbf{k}=0$ in cubic A^NB^{8-N} crystals.[4]

At first sight the spin–orbit splitting of the threefold orbitally degenerate levels at $\mathbf{k}=0$ (which correspond to the p states of the free atoms) appears to provide an excellent measure of the parameter λ in the valence-band wave function discussed in Sec. 3. This wave function may be written in the AB unit cell in the atomic orbital approximation as

$$\psi=\psi_A+\lambda_{AB}\psi_B, \tag{A1}$$

where ψ_A is a valence p orbital centered on atom A, and similarly for ψ_B. Denote the spin–orbit operator by

$$\mathcal{H}_{so}=U\mathbf{L}\cdot\mathbf{S}, \tag{A2}$$

where $\mathbf{L}$ and $\mathbf{S}$ denote orbital and spin angular momentum. The coupling factor U will be discussed in more detail below. The spin–orbit splitting is determined by

$$E_{so}=\langle\psi\mid\mathcal{H}_{so}\mid\psi\rangle/\langle\psi\mid\psi\rangle. \tag{A3}$$

If ψ_A and ψ_B were orthogonal, and cross terms of U between atoms A and B could be neglected, (A3) would reduce to

$$E_{so}{}^{AB}=(\mathcal{H}_{so}{}^{A}+\lambda_{AB}{}^2\mathcal{H}_{so}{}^{B})/(1+\lambda_{AB}{}^2), \tag{A4}$$

where

$$\mathcal{H}_{so}{}^{A}=\langle\psi_A\mid\mathcal{H}^{so}\mid\psi_A\rangle/\langle\psi_A\mid\psi_A\rangle,$$

and similarly for $\mathcal{H}_{so}{}^{B}$. Given experimental values of $E_{so}{}^{AB}$, if some way could be found to evaluate $\mathcal{H}_{so}{}^{A}$ and $\mathcal{H}_{so}{}^{B}$ as invariant parameters characteristic of the respective atoms only (i.e., independent of the nearest-neighbor atoms), then (A4) could be solved for λ_{AB}, and the definition of ionicity given in Sec. 7 could be used.

From the foregoing it is clear that unless some independent way can be found to evaluate $\mathcal{H}_{so}{}^{A}$, this procedure requires at least as many independent parameters (one for each atom) as the Clausius–Mosotti theory. One might imagine that these numbers could

---

[4] Braunstein and Kane (1962) were careful to note that using the same sharing time for all III–V compounds gives good agreement with experiment, and so does not reflect the different ionic character among the compounds. See Sec. 7, where the Coulson theory gives a similar result.

be taken from free-atom term values, but it turns out that in elemental crystals (e.g., Ge) $E_{\mathrm{so}}{}^{\mathrm{AA}}$ differs from the atomic value by 50%. Even this, however, is not the most serious shortcoming of this approach.

Let the nuclei of atoms A and B be located at $\mathbf{R}_{\mathrm{A}}$ and $\mathbf{R}_{\mathrm{B}}$, respectively, and let $r_{\mathrm{A}} = |\mathbf{r} - \mathbf{R}_{\mathrm{A}}|$, and similarly for r_{B}. Let $W_{\mathrm{A}}(r_{\mathrm{A}})$ and $W_{\mathrm{B}}(r_{\mathrm{B}})$ represent the parts of the *total* crystal potential, including the core potentials, which are spherically symmetric about atoms A and B. One can then show that $\mathcal{H}_{\mathrm{so}}{}^{\mathrm{A}}$ depends on $\langle \psi^{\mathrm{A}} | \partial^2 W_{\mathrm{A}}/\partial r_{\mathrm{A}}{}^2 | \psi^{\mathrm{A}} \rangle / \langle \psi^{\mathrm{A}} | \psi^{\mathrm{A}} \rangle$, and similarly for $\mathcal{H}_{\mathrm{so}}{}^{\mathrm{B}}$. If we approximate W_{A} by Ae^{-kr}/r, where A is the total charge on the A nucleus, then $\partial^2 W_{\mathrm{A}}/\partial r_{\mathrm{A}}{}^2$ involves several terms, the largest of which is proportional to $r_{\mathrm{A}}{}^{-3}$. The energy gap between bonding and antibonding states, as we have seen, depends on the valence pseudopotential near the atomic radius. The spin–orbit splittings, on the other hand, involve the second derivative of the *total* potential, and the largest contribution comes from the core region where $r_{\mathrm{A}}{}^{-3}$ is large.

This discussion can be made more quantitative by expanding ψ_{A} in terms of the core states of atom A and a smooth outside or valence function. Such an expansion has been carried out in great detail for Ge (Liu, 1962). The reader may refer to this paper for quantitative instances which illustrate the effects discussed.

Because $\partial^2 W_{\mathrm{A}}/\partial r_{\mathrm{A}}{}^2$ and $\partial^2 W_{\mathrm{B}}/\partial r_{\mathrm{B}}{}^2$ are large only in the nonoverlapping regions of the atomic cores, neglect of cross terms of U between atoms A and B, mentioned in connection with Eq. (A3), is justified. However, neglect of cross terms of ψ_{A} and ψ_{B} is not justified. Let

$$S_{\mathrm{AB}} = \langle \psi_{\mathrm{A}} | \psi_{\mathrm{B}} \rangle, \qquad \text{(A5)}$$

then (A4) becomes

$$E_{\mathrm{so}}{}^{\mathrm{AB}} = (\mathcal{H}_{\mathrm{so}}{}^{\mathrm{A}} + \lambda_{\mathrm{AB}}{}^2 \mathcal{H}_{\mathrm{so}}{}^{\mathrm{B}}) / (1 + 2\lambda_{\mathrm{AB}} S_{\mathrm{AB}} + \lambda_{\mathrm{AB}}{}^2). \qquad \text{(A6)}$$

To simplify the situation, consider an AA elemental crystal, where $\lambda = 1$. Then

$$E_{\mathrm{so}}{}^{\mathrm{AA}} = \mathcal{H}_{\mathrm{so}}{}^{\mathrm{AA}} / (1 + S_{\mathrm{AA}}). \qquad \text{(A7)}$$

The magnitude of S_{AA} may be obtained as follows. The bonding state is given by (A1) with $\lambda = 1$; the antibonding AA state by (A1) with $\lambda = -1$. As a result the antibonding spin–orbit splitting is

$$\tilde{E}_{\mathrm{so}}{}^{\mathrm{AA}} = \mathcal{H}_{\mathrm{so}}{}^{\mathrm{AA}} / (1 - S_{\mathrm{AA}}). \qquad \text{(A8)}$$

From (A7) and (A8) we see that the spin–orbit splitting of the antibonding state is larger than that of the bonding state by the factor

$$f = (1 + S_{\mathrm{AA}}) / (1 - S_{\mathrm{AA}}), \qquad \text{(A9)}$$

with $S_{\mathrm{AA}} > 0$. In Ge, however, these two splittings at $\mathbf{k} = 0$ are known from cyclotron resonance experiments (Hensel and Suzuki, 1969). The spin–orbit splitting of the antibonding state is only half that of the bonding state.

I have led the reader down the garden path to this paradox in order to dramatize how misleading apparently simple arguments based on atomic orbitals can be. The atomic concept so popular in ancient Athens still enjoys an intuitive appeal to many physicists in situations where they ought to know better.

The paradox presented here was first recognized and resolved (Phillips and Liu, 1962), long before experimental data conclusively demonstrated that antibonding spin–orbit splittings are actually smaller than bonding ones. The paradox arises because of the initial assumption that the bonding and antibonding wave functions can be confined to the $L = 1$ subspace of atomic functions. In actuality the crystal wave functions involve considerable p–d hybridization. The valence wave functions involve (p bonding–d antibonding) states, the conduction ones (p antibonding–d bonding) ones. Clearly the admixture of d states is greater in the level of higher energy, which makes its spin–orbit splitting smaller. This is what one would expect on general grounds from the fact that $\partial^2 W/\partial r^2$ is large in the core region. But equally clearly d admixture must be allowed for, and the amount of this admixture will depend significantly on the various energy gaps in the AB crystal. Thus the parameters $\mathcal{H}_{\mathrm{so}}{}^{\mathrm{A}}$ and $\mathcal{H}_{\mathrm{so}}{}^{\mathrm{B}}$ really have no invariant significance.

In closing this short lesson, I should mention that there *are* qualitative trends in E_{so} which correspond roughly to the values of λ_{AB} obtained in the Coulson model. For the reasons discussed above, however, one should not attempt to make these trends the basis of an *accurate* scale of ionicity. Ultimately the reason for this is that $\partial^2 W_{\mathrm{A}}/\partial r_{\mathrm{A}}{}^2$ is a poor weighting factor compared to V_{A}, because the former is too singular near $r_{\mathrm{A}} = 0$. The same remark is doubly applicable to values of the nuclear contact interaction which depends on the even more singular operator $\delta(\mathbf{r} - \mathbf{R}_{\mathrm{A}})$, and great care should be exercised in basing even qualitative estimates of ionicities on values of $|\psi(0)|^2$ measured in nuclear magnetic resonance experiments.

REFERENCES

Allred, A. L., 1961, J. Inorg. Nucl. Chem. **17**, 215.

Animalu, A. O. E., and V. Heine, 1965, Phil. Mag. **12**, 1249.

Bardasis, A., and D. Hone, 1967, Phys. Rev. **153**, 849.

Bell, R. P., and D. A. Long, 1950, Proc. Roy. Soc. **A203**, 364.

Born, M., and W. Heisenberg, 1924, Z. Physik. **23**, 388.

Braunstein, R., and E. O. Kane, 1962, J. Phys. Chem. Solids **23**, 1423.

Burstein, E. *et al.*, 1967, J. Quantum Chem. **15**, 759.

Callen, H. B., 1949, Phys. Rev. **76**, 1394.

Clausius, R., 1879, *Die Mechanische Warmetheorie* (Vieweg Braunschweig, Germany), Vol. 2, p. 94.

Cochran, W., and R. A. Cowley, 1962, J. Phys. Chem. Solids **23**, 447.

Cohen, M. L. and T. K. Bergstresser, 1966, Phys. Rev. **141**, 789.

——, and V. Heine, 1970, Solid State Phys. (to be published).

Coulson, C. A., L. B. Redei, and D. Stocker, 1962, Proc. Roy. Soc. (London) **270**, 352.

Di Domenico, M., Jr., and S. H. Wemple, 1969, J. Appl. Phys. **40**, 720.

Dresselhaus, G. and M. S. Dresselhaus, 1967, Phys. Rev. **160**, 649.

Ehrenreich, H. and M. H. Cohen, 1959, Phys. Rev. **115**, 786.

Elliott, R. J., 1957, Phys. Rev. **108**, 1384.

Fajans, K., and G. Joos, 1924, Z. Physik. **23**, 1.

Faulkner, R. A., 1969, Phys. Rev. **184**, 713.

Heine, V., 1967, Phys. Rev. **153**, 673.

——, and R. O. Jones, 1969, J. Phys. C **2**, 719. (The Jones referred to is H. Jones, who first stressed the usefulness of zone counting for insulators in the 1930's.)

Hensel, J. C., and K. Suzuki, 1969, Phys. Rev. Letters **22**, 838.

Herman, F. and S. Skillman, 1963, *Atomic Structure Calculations* (Prentice-Hall, Englewood Cliffs, N.J.).

Hopfield, J. J. (private communication).

Jacobs, R. L., 1968, J. Phys. C **1**, 492.

Jones, D., and A. H. Lettington, 1969, Solid State Commun. **7**, 1319.

Keating, P. N., 1966, Phys. Rev. **145**, 637.

Kohn, W., 1957, Solid State Phys. **5**, 257.

——, and J. M. Luttinger, 1955, Phys. Rev. **97**, 883.

——, and N. Rostoker, 1954, Phys. Rev. **94**, 1111.

Korringa, J., 1947, Physica **13**, 392.

Landau, L. D., and E. M. Lifshitz, 1960, *Electrodynamics of Continuous Media* (Pergamon Press, New York), p. 259.

Liu, L. 1962, Phys. Rev. **126**, 1317.

Martin, R. M., 1970, Phys. Rev. (to be published).

Moffitt, W. E., 1949, Proc. Roy. Soc. (London) **A196**, 510.

——, 1950, Proc. Roy. Soc. (London) **202**, 548.

Mooser, E. and W. B. Pearson, 1959, Acta Cryst. **12**, 1015.

Mosotti, O. F., 1850, Mem. Soc. Ital. Sci. (Modena) **14**, 49.

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

Mullikan, R. S., 1949, J. Chim. Phys. **46**, 497.

Nozieres, P., and D. Pines, 1959, Phys. Rev. **113**, 1254.

Pauling, L., 1927, Proc. Roy. Soc. (London) **A114**, 191.

——, 1960, *The Nature of the Chemical Bond* (Cornell U. P., Ithaca, N.Y.).

Pearson, W. B., 1962, J. Phys. Chem. Solids **23**, 103.

Penn, D. R., 1962, Phys. Rev. **128**, 2093.

Philipp, H. R., and H. Ehrenreich, 1963, Phys. Rev. **129**, 1550

Phillips, J. C., 1958, Phys. Rev. **112**, 685.

——, 1966, Solid State Phys. **18**, 56.

——, 1968, Phys. Rev. **166**, 832.

——, 1968a, J. Chem. Phys. **48**, 5740.

——, 1968b, Phys. Rev. Letters **20**, 550.

——, 1969, Chem. Phys. Letters **3**, 286.

——, 1969a, Phys. Rev. Letters **22**, 285.

——, 1969b, Phys. Rev. Letters **22**, 645.

——, 1969c, Phys. Rev. Letters **23**, 482.

——, 1970, *Covalent Bonding in Crystals, Molecules, and Polymers* (University of Chicago Press, Chicago).

——, 1970a, Phys. Rev. (to be published).

——, 1970b, Phys. Today **23**, 23.

——, 1970c, Phys. Rev. Letters **24**, 1114.

——, and L. Liu, 1962, Phys. Rev. Letters **8**, 94.

——, and R. Sandrock, 1968, Phys. Rev. **174**, 663.

——, and J. A. Van Vechten, 1969, Phys. Rev. **183**, 709.

——, and J. A. Van Vechten, 1969a, Phys. Rev. Letters **23**, 1115.

——, and J. A. Van Vechten, 1970, Phys. Rev. (to be published).

Pritchard, H. O., and H. A. Skinner, 1955, Chem. Rev. **55**, 745.

Roberts, S., 1949, Phys. Rev. **76**, 1215.

Robinson, F. N. H., 1968, Phys. Letters **26A**, 435.

Rudman, P. S., J. Stringer, and R. I. Jaffee, Eds., 1967, *Phase Stability in Metals and Alloys* (McGraw-Hill, New York).

Schactschneider, J. H., and R. G. Snyder, 1963, Spectrochim. Acta **19**, 117.

Slater, J. C., 1937, Phys. Rev. **51**, 846.

——, 1953, Phys. Rev. **92**, 603.

——, and G. F. Koster, 1954, Phys. Rev. **94**, 1498.

Streitwieser, A. W., Jr., 1961, *Molecular Orbital Theory* (Wiley, New York).

Stukel, D. J., *et al.*, 1969, Phys. Rev. **179**, 740.

Tessman, J. R., A. H. Kahn, and W. Shockley, 1953, Phys. Rev. **92**, 890.

Thomas, D. G., and J. J. Hopfield, 1966, Phys. Rev. **150**, 680.

Tosi, M. P., 1965, Sol. State Phys. **16**, 68.

Van Vechten, J. A., 1969, Phys. Rev. **182**, 891.

——, 1969a, Phys. Rev. **187**, 1007.

——, and T. K. Bergstresser, 1970, Phys. Rev. (to be published).

Walter, J. P., and M. L. Cohen, 1969, Phys. Rev. **183**, 763.

Wemple, S. H., and M. DiDomenico, Jr., 1969, Phys. Rev. Letters, **23**, 1156.

Ziman, J. M., 1965, Proc. Phys. Soc. (London) **86**, 337.