

Electronic Structure from the Point of View of the Local Atomic Environment

VOLKER HEINE

Cavendish Laboratory, University of Cambridge, Cambridge, England

I.	Fundamental Ideas	1
1.	Throwing Out \mathbf{k} Space.....	1
2.	The Invariance Theorem.....	9
3.	Brief Appendix on the Mathematics of the Green Functions.....	17
II.	The Method of Matching the Green Functions.....	19
4.	Derivation of the Green Function	19
5.	Boundary Correction Terms in the Invariance Theorem	24
6.	Applications to Alloys, Surfaces, and Molecules.....	27
III.	LCAO: From under a Cloud to Out in the Sun	35
7.	Historical Perspective and Basic Clarifications	35
8.	Electron Correlation in Solids and Molecules	42
9.	The Modern Anderson-Bullett Scheme for the Tight-Binding Method	47
10.	From Atomic Resonances to LCAO Parameters for Transition Metals.....	57
IV.	Calculations with Atomic Orbitals, Atomic Vibrations, and Other Localized Basis Sets	69
11.	Working with Clusters, Moments, and Continued Fractions.....	69
12.	The Haydock Recursion Scheme	87
13.	Total Energy Calculations.....	92
14.	Various Types of Perturbation	99
V.	Miscellaneous Properties from the Local Point of View.....	105
15.	A Local Computational Scheme for the Density Functional Formalism	105
16.	Pressure, Force, and Energy	110
17.	Self-Consistency in the Potential	121
VI.	Historical Note	123
18.	The Development of the Concepts of Local Electronic Structure	123

I. Fundamental Ideas

1. THROWING OUT \mathbf{k} SPACE

Every textbook on solid state physics begins with a discussion of the translational symmetry of crystals, and the whole subject is based on this model of

a perfect crystal with perfect lattice—translational symmetry, periodic boundary conditions even being invented so that the perfect symmetry is not lost at the surface of the specimen. As a result, one has mathematically rigorously Bloch's theorem, the concept of the \mathbf{k} vector with band structures and Brillouin zones to describe electrons, phonons, magnons, x-ray diffraction, and most else in solids. Perfect periodicity and \mathbf{k} space have been the corner stones of an enormously successful development of solid state physics for nearly 50 years.

However, any strongly rolling bandwagon tends to sweep under the rug the awkward little questions that cannot conveniently be attached to it. In recent years with the development of surface science and the study of amorphous materials, some of these questions have been creeping out from under the rug.

This volume, in brief, is devoted to aspects of solid state physics where the model of perfect lattice periodicity is simply not appropriate. This theme has influenced strongly the research of the solid state theory group in Cambridge, England, over the last ten years. While the same is true to a considerable extent worldwide, the particular development of the subject and selection of topics in this volume will reflect the point of view of the Cambridge group, as will many of the applications. No slight toward the work of others is intended; unfortunately pages and hours are finite.

There are broadly two types of situation. The first is where an essential lack of periodicity exists^{1,2}: surfaces and interfaces, random alloys, amorphous materials, defects in solids, liquid metals, spin glasses, molecules, clusters and microcrystallites, etc. At a perfect surface one loses the periodicity perpendicular to the surface, retaining a two-dimensional periodicity. The latter is also lost at an adsorbed atom, and at steps and kink sites believed to be important for molecular reactions at surfaces. To make a model with three-dimensional lattice periodicity is to throw out the baby with the bathwater. Nevertheless a steady stream of papers is published treating some such systems in terms of Bloch-like functions, albeit with finite decay length: in some cases these will turn out to be useful approximations but one cannot prejudge that issue until one has dealt with the lack of periodicity properly.

In the second type of situation one may have perfect crystal symmetry, but one is asking questions about the local environment of one atom, e.g., its bonding to its neighbors. By definition a Bloch function is uniformly spread over the whole solid, so that a local property such as the charge density, bond order, or atomic magnetic moment has to be obtained as a summation over $\sim 10^{23}$ electrons. Computers can do it, but physical insight is lost, unless the applicability of some perturbation theory enables the summation

¹ R. Haydock, Chapter III, this volume.

² M. J. Kelly, Chapter IV, this volume.

to be carried out analytically, e.g., in the pseudopotential theory of phonons and crystal structures.³ An intermediate case between the two situations is provided by the ordered Laves phases of transition metal alloys with rather large unit cells. Bloch's theorem makes no statement about the wave function in one unit cell, only its relation to other unit cells that are rather far away in the case of 24 and 36 atoms per unit cell, respectively. Surely the physics of the material is as much concerned with the intracell interactions. Certainly the methods of \mathbf{k} space become unwieldy⁴: the Brillouin zone collapses to a small size containing 180 separate d bands like a can of spaghetti.

At the risk of overkill, let us review briefly a specific example for which, incidentally, we include some results later on. The ferromagnetic ordered alloy Fe₃Al has a body-centered cubic (bcc) arrangement of atoms, the same as ordinary iron except that every fourth atom is replaced by an aluminum atom. Thus Fe₃Al contains two types of iron site: the D site, in which all eight nearest neighbors are also iron atoms, and the A site, where four nearest neighbors are iron atoms and four aluminum atoms. The magnetic moments on the D and A sites are 2.14 and 1.46 Bohr magnetons, (μ_B), respectively.⁵ What is remarkable is how close the moment on the D site is to that of pure iron, namely, 2.2 μ_B . In other words, one shell of nearest neighbors is enough to make a D site iron atom feel it is almost in pure iron. Why? There is no real answer in conventional band theory. (I take it as almost universally accepted now that all the electrons in iron are on the metallic side of the Mott transition, so that a band picture of the ferromagnetism is appropriate, at least at 0 K.) Of course one can do a band structure calculation, putting in the same exchange splitting that gives the right moment in pure iron, and sample the Brillouin zone to sum over all occupied states. This turns out to give the right numbers for the magnetic moments^{6,7} which verifies that quantum mechanics gives a correct description of nature, or rather that our methods form a valid approximation, but it does not carry us much further in answering the question.

Instead of making physics fit the mathematics of perfect periodicity and \mathbf{k} space, can we not develop the theory in a way closer and more appropriate to the physics we want to describe? We can, but it is necessary to start again at the beginning. Traditional quantum mechanics revolves around solving the Schrödinger equation in the form

$$\mathcal{H}\psi_n = E_n\psi_n, \quad (1.1)$$

³ V. Heine and D. Weaire, *Solid State Phys.* **24**, 250 (1970).

⁴ R. L. Johannes, R. Haydock, and V. Heine, *Phys. Rev. Lett.* **36**, 372 (1976); see also R. Haydock and R. L. Johannes, *J. Phys. F* **5**, 2055 (1975).

⁵ R. Nathans, M. T. Piggott, and C. G. Shull, *J. Phys. Chem. Solids* **6**, 38 (1958).

⁶ S. Ishida, J. Ishida, S. Asano, and J. Yamashita, *J. Phys. Soc. Jpn.* **41**, 1570 (1976).

⁷ R. Haydock and M. V. You, to be published.

since the subject arose from the study of atomic energy levels and is best taught that way. However, the eigenfunctions ψ_n are singularly inappropriate for the discussion of large systems with consequent high degeneracy, for two reasons. As already mentioned, any physically measurable quantity involves a summation over $N \sim 10^{23}$ electrons, with a loss of physical insight and certainly the loss of anything related to any particular ψ_n . This is not an artifact of the mathematics. It is the recognition that the molecular-orbital approach gives a better description of a molecule than the Heitler London method, and that electrons do roam through the molecule as seen in the chemical properties particularly of conjugated hydrocarbons. In solids the limit $N \rightarrow \infty$ gives us a sharp division, the Mott transition,⁸ and in almost all materials we are on the “metallic” side,⁹ with correlation subsumed into Landau quasi-particles. Of course in NaCl and solids of the inert gases with completely closed shells the two approaches become identical, but in semiconductors like silicon or selenium the band picture is fundamentally the correct one, with the electron states spread out over the whole solid.

The second problem about the eigenfunctions in Eq. (1.1) is that they are fickle. Consider a free electron gas of the same density in a cubical box and in a spherical “box.” There is almost no correspondence between the individual eigenstates in the two cases, yet we believe that they will give the same Fermi level E_F and the same total kinetic energy per unit volume. Worse than that: the perturbation produced by changing one atom out of $N \sim 10^{23}$ in a solid is of order $1/N$, but so is the spacing between consecutive energy levels E_n , so that the alteration of one atom will produce a wholesale scrambling^{10a} of nearly degenerate states ψ_n throughout the band. The individual eigenstates are unstable under small perturbations, which are physically irrelevant over almost the whole of the solid beyond some distance from the perturbation (at least on the delocalized side of the Anderson^{10b} transition).

We see now that both difficulties with the ψ_n arise from having a large system with a band of quasi-degenerate delocalized states. In that sense they do not spring directly from the model of perfect lattice periodicity because the same problems apply, for example, in studying some bonding site in a large molecule. However, the mathematics of Bloch’s theorem relates to the

⁸ J. M. Ziman, “Principles of the Theory of Solids.” Cambridge Univ. Press, London and New York, 1964.

⁹ Even in a Mott insulator such as CoO it is appropriate to take the band picture as the zero-order basis and then introduce the Coulomb interaction.¹⁰

¹⁰ P. W. Anderson, *Solid State Phys.* **14**, 99 (1963).

^{10a} For a beautiful recent picture of such an eigenfunction, see S. W. McDonald and A. N. Kaufman [*Phys. Rev. Lett.* **42**, 1189 (1979)] where the scrambling is produced by the unsymmetrical, though smooth, boundary.

^{10b} P. W. Anderson, *Science* **201**, 307 (1978); *Phys. Rev.* **109**, 1492 (1958).

eigenstates ψ_n . We need to throw out \mathbf{k} space, even where we have lattice periodicity as exemplified above, because \mathbf{k} space relates to the individual eigenstates ψ_n and these are inappropriate to the systems and properties we wish to discuss.

New tools are needed, a new formulation of quantum theory. Much of what we want is contained in the *local density of states*¹¹

$$n(E, \mathbf{r}) = n(E) |\psi_E(\mathbf{r})|^2, \quad (1.2)$$

where $n(E)$ is the total density of states of the system that is modulated by $|\psi|^2$ for a “typical” state of energy E . Integrated up to E_F it will give the total electron density. In a compound it will distinguish which bands are more or less weighted on which atoms. At the surface of a solid, the density of states “seen” on a surface atom (for example, by ion neutralization spectroscopy) will differ from that in the bulk, perhaps because of surface states. In a magnetic compound or at the surface of a magnetic material, the difference between the up and down spin local density of states will describe the distribution of magnetic moment.

Mathematically Eq. (1.2) begs the question of the definition of a typical ψ_E ; more precisely we write (for each spin direction)

$$n(E) = \sum_n \delta(E - E_n), \quad (1.3)$$

$$n(E, \mathbf{r}) = \sum_n |\psi_n(\mathbf{r})|^2 \delta(E - E_n), \quad (1.4)$$

as sums over the eigenstates of (1.1). The reader to whom (1.3) is unfamiliar is asked to consider a finite system, for which the “density of states” $n(E)$ is a set of delta functions at the energy eigenvalues E_n , and every time we integrate in energy through one we “count” one more electron. For a macroscopic system the eigenvalues are very close together and $n(E)$, $n(E, \mathbf{r})$ are assumed infinitesimally smeared out into smooth continuous functions of E through the band. We have already progressed one step.

Suppose we have a set of ψ_n with equal energy. Then a unitary transformation among the ψ_n , such as is induced by a small perturbation, may change individual ψ_n by a large amount but leave the sum $\sum_n |\psi_n(\mathbf{r})|^2$ invariant. Thus (1.4) does not have the extreme sensitivity to small perturbations that wave functions have.

At this point our story develops three threads. Physically the most important one we defer to Section I,2: the local density of states has an *invariance property* that goes a long way toward explaining the magnetic moments in Fe₃Al and much else besides.¹¹

¹¹ J. Friedel, *Adv. Phys.* **3**, 446 (1954).

The second thread is to point out the mathematical step that has made all the work in this volume possible. Consider Eq. (1.4): it defines the local density of states in terms of the (extended) eigenstates ψ_n of the whole system; and if we had to evaluate it this way, then nothing would have been gained because we would first have had to calculate the ψ_n and E_n , i.e., the band structure for a periodic system and the unimaginably complicated ψ_n for a less-ordered one. The point is that $n(E, \mathbf{r})$ is related to the imaginary part of the *Green function*^{1,2} $G(\mathbf{r}, \mathbf{r}', E)$ by the equation

$$n(E, \mathbf{r}) = -\pi^{-1} \operatorname{Im} G(\mathbf{r}, \mathbf{r}', E), \quad (1.5)$$

and the Green function has many marvellous mathematical properties so that it *can be solved for directly* without ever involving oneself with the ψ_n . One of these techniques, the method of *matching Green functions*, is discussed in Section II,4; another, *Haydock's recursion method*, is discussed in Section IV,12. Most of the work to be described in this volume^{1,2} hangs on these two methods, particularly the second one.

The Green function is defined as

$$G(\mathbf{r}, \mathbf{r}', E) = \sum_n \frac{\psi_n(\mathbf{r})\psi_n^*(\mathbf{r}')}{E + iO - E_n} \quad (1.6a)$$

$$= \mathcal{P} \sum_n \frac{\psi_n(\mathbf{r})\psi_n^*(\mathbf{r}')}{E - E_n} - i\pi \sum_n \psi_n(\mathbf{r})\psi_n^*(\mathbf{r}') \delta(E - E_n), \quad (1.6b)$$

from which (1.5) immediately follows, using Eq. (1.4). The reader unfamiliar with the significance of the infinitesimal iO in Eq. (1.6a) and its equivalence to the second term in (1.6b) is referred to the opening lines of Section I,3.

The use of G is not limited to finding the local density of states. Suppose in ultraviolet photoemission spectroscopy we are interested in electron transitions to a final emerging state ψ_f of energy E_f . The emerging current $I(E_f)$ will be proportional to

$$\sum_n |\langle \psi_f | p_\alpha | \psi_n \rangle|^2 \delta(E_f - h\nu - E_n) = -\pi^{-1} \operatorname{Im} \iint G(\mathbf{r}, \mathbf{r}', E_f - h\nu) \\ \times p_\alpha p'_\alpha \psi_f^*(\mathbf{r})\psi_f(\mathbf{r}') d^3\mathbf{r} d^3\mathbf{r}' \quad (1.7)$$

as follows from (1.6b). The result involves the transition matrix elements of the relevant component p_α of the momentum operator, which have also been expressed in terms of G .

This, then, is the formulation of quantum mechanics needed: to express the physical quantities required in terms of G and solve for that directly.

^{1,2} Peace to some of my friends! (True to Cavendish tradition, "Green function theorist" is still a dirty word in parts of the laboratory.)

Sometimes the analysis can be pushed much further. For example, in (1.7), if the final state ψ_f can be represented as a plane wave, then the integrals in (1.7) give the double Fourier transform $G(\mathbf{k}, \mathbf{k}, E_f - h\nu)$, which can be calculated directly¹³ as we show in Section IV,11. The \mathbf{k} may even be complex to allow for a finite escape depth from below the surface. If ψ_f has to be expressed as a sum of a few plane waves because of “final-state scattering,” then one would need to add a few terms similar to $G(\mathbf{k}, \mathbf{k}, E)$ but the principle is the same.¹⁴ Other examples occur later.

We turn now to the third thread in our story. Often in solids and molecules it is useful to express the eigenstates ψ_n , or indeed any states, as a *linear combination of atomic orbitals* $\phi_{\alpha l}$ (LCAO):

$$\psi_n(\mathbf{r}) = \sum_{\alpha l} a_{n,\alpha l} \phi_{\alpha l}(\mathbf{r}). \quad (1.8)$$

Here l is a site index for the atomic site and α indicates the type of orbital, e.g., $3s$, $3p_x$, $3p_y$. In its crudest form, LCAO gives the tight-binding method (TB), but LCAO in the present context must be conceived much more widely than that, involving the use of any appropriately selected atomiclike basis states. We return to that point in Sections III,9 and III,10, and the chapter by Bullett¹⁵ is devoted to it. Suffice it to note that almost the whole of the work reviewed in the present volume^{1,2,15} is expressed in LCAO terms. There are historical reasons: the local point of view was stimulated greatly by the growth of surface science, where the transition metal d bands, usually described in LCAO approximation, are of special interest. Also in nearly free-electron solids one can often go a long way with second-order perturbation theory³ based on a free-electron gas. The orbitals $\phi_{\alpha l}(\mathbf{r})$ are of course localized around the atomic sites l , but this has no bearing on the fact that the ψ_n are extended as discussed above: the latter simply implies that the coefficients $a_{n,\alpha l}$ in (1.8) are of comparable magnitude throughout the solid. In terms of (1.8) the Schrödinger equation takes the matrix form

$$\mathcal{H} a_n = E_n a_n, \quad (1.9)$$

(where a_n is the column matrix column vector) of coefficients $a_{n,\alpha l}$. Also \mathcal{H} is the square matrix whose rows and columns are labeled by the index pair αl , its matrix elements $\langle \alpha l | \mathcal{H} | \alpha' l' \rangle$ in the Dirac notation representing the hopping of the electron from orbital αl to $\alpha' l'$ for $\alpha' l' \neq \alpha l$ and the energy of orbital αl for $\alpha' l' = \alpha l$. In general, the orbitals $\phi_{\alpha l}$ will not be orthogonal to those on neighboring sites, and this can be taken properly into account as discussed in Sections III,7 and III,9 and other chapters.^{1,15} But for the purposes of

¹³ R. K. C. McLean and R. Haydock, *J. Phys. C* **10**, 1929 (1977).

¹⁴ R. K. C. McLean, *J. Phys. C* **10**, 4655 (1977).

¹⁵ D. W. Bullett, Chapter II, this volume.

introducing the ideas, we assume all the $\phi_{\alpha l}$ as orthonormal throughout other parts of this chapter.

The local density of states becomes

$$n_{\alpha l}(E) = \sum_n |a_{n, \alpha l}|^2 \delta(E - E_n), \quad (1.10)$$

which is the density of states, with the coefficients $|a_{n, \alpha l}|^2$ representing the weight that the ψ_n have on a particular orbital $\phi_{\alpha l}$.

We shall prove that $n_{\alpha l}(E)$ is given by

$$n_{\alpha l}(E) = -\pi^{-1} \operatorname{Im} \langle \alpha l | [E + iO - \mathcal{H}]^{-1} | \alpha l \rangle. \quad (1.11)$$

The eigenstates ψ_n form a complete set so that in Dirac notation

$$\sum_n |\psi_n\rangle \langle \psi_n| \quad (1.12)$$

is the unit operator. Also we have from (1.8)

$$a_{n, \alpha l} = \langle \phi_{\alpha l} | \psi_n \rangle. \quad (1.13)$$

After inserting (1.12) twice into (1.11), the latter becomes

$$\begin{aligned} & -\pi^{-1} \operatorname{Im} \sum_n \sum_m \langle \phi_{\alpha l} | \psi_n \rangle \langle \psi_n | [E + iO - \mathcal{H}]^{-1} | \psi_m \rangle \\ & \times \langle \psi_m | \phi_{\alpha l} \rangle. \end{aligned} \quad (1.14)$$

Now \mathcal{H} is a diagonal matrix in the representation $\langle \psi_n | \mathcal{H} | \psi_m \rangle$ formed by its own eigenstates, and hence $[E + iO - \mathcal{H}]^{-1}$ is also diagonal with matrix elements

$$(E + iO - E_n)^{-1} \delta_{nm}, \quad (1.15)$$

which must be interpreted in accordance with Eq. (3.4) as having an imaginary part

$$-\pi \delta(E - E_n) \delta_{nm}. \quad (1.16)$$

Using (1.15), (1.16) and also (1.13) in (1.14) we obtain

$$\sum_n \sum_m a_{n, \alpha l} \delta(E - E_n) \delta_{nm} a_{m, \alpha l}^*, \quad (1.17)$$

which equals (1.10). This therefore proves the formula (1.11) from which the local density of states will be computed.

There is obviously a close analogy of (1.11) to Eqs. (1.5) and (1.6a) for the case of the continuous variables \mathbf{r} . In fact, as pointed out in Section I,3, the Green function in matrix form is the square matrix

$$G(E) = [E + iO - \mathcal{H}]^{-1}, \quad (1.18)$$

which can be transformed according to different basis functions. One point now becomes clear: suppose we are computing with a finite cluster involving 10^4 atomic orbitals, which is then the number of rows and columns in (1.18). Clearly $G(E)$ contains a tremendous amount of information (see Section I,3). However, the information (1.11) that we want, the local density of states on one orbital $\phi_{\alpha l}$, consists of only one matrix element in this matrix. The same comment can be made regarding Eq. (1.5): we select only the $\mathbf{r}' = \mathbf{r}$ part of G . We see now that the trick in the new formulation of quantum mechanics is not just to express measurable quantities in terms of G , but *in terms of some appropriate small parts of G that can be solved for and computed separately from all the unwanted remainder of G .*

Instead of proceeding from the general to the particular, we have preferred to introduce ideas in the specific (and to most readers probably the most familiar) context of electronic structure of solids and molecules. But the ideas can be generalized. Matrix equations analogous to (1.9) describe the linking of the atomic vibrations into normal modes of oscillation of the system, or the interaction of localized Heisenberg spins¹⁶ distributed on the sites l . Mookerjee¹⁷ has even used the probabilistic nature of quantum mechanics to replace the ensemble averaging over a distribution of impurities, by the solution of an enlarged pseudo-“Hamiltonian” without further averaging. Perhaps the techniques can even be applied to transport problems, e.g., the percolation conduction of electrons through a band of impurity states in a semiconductor. Some of these applications are referred to in Section IV,11 and in Kelly’s chapter.²

This chapter serves to introduce the remainder of this volume. To skim the cream still further, the reader is directed to Section I,2 and then Sections IV,11–IV,14, with a glance at Sections III,7, III,9, and III,10 on the way.

2. THE INVARIANCE THEOREM

The local density of states Eq. (1.4) is similar to a classical quantity,^{11,18,19} the density of black-body radiation $\rho(\omega, \mathbf{r})$: we change ψ to the electromagnetic field \mathcal{E} and E to the frequency ω . To make the analogy complete we would need to discuss the normalization of the normal modes $\mathcal{E}_n(\mathbf{r})$ and their degree of excitation at temperature T but that presents no difficulty.

Now it is well known that an invariance theorem applies to $\rho(\omega, \mathbf{r})$: the density of black-body radiation (per unit volume per unit frequency range)

¹⁶ T. Fujiwara, *J. Phys. C* **10**, 5035 (1977).

¹⁷ A. Mookerjee, *J. Phys. C* **6**, 1340 and L205 (1973).

¹⁸ C. Kittel, “Quantum Theory of Solids,” p. 338. Wiley, New York, 1963.

¹⁹ M. V. Laue, *Ann. Phys. (Leipzig)* [4] **44**, 1197 (1914).

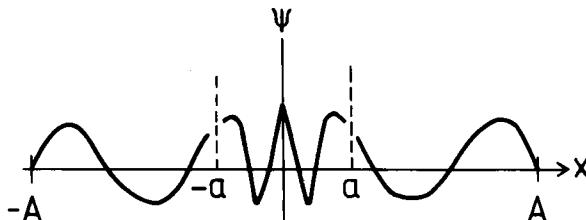


FIG. 1. Matching the wave function at the boundary of an atom ($x = \pm a$) to the surrounding medium.

in a container is independent of the size and shape of the container and of the material that the walls consist of, i.e., it is independent (mathematically speaking) of the boundary conditions. As Friedel¹¹ pointed out, the same mathematics can be used to prove the equivalent result for the quantum *local density of states*, indeed for the whole Green function: it, too, is *independent of the boundary conditions*.

We can apply this result immediately to understand in a zero-order sense the magnetic moments of the iron atoms in Fe₃Al and in pure iron: they are the same because the local density of states is the same inside any iron atom, independent of the boundary conditions on the wave function at the boundary of the atom. Note that here the “container” that we apply the theorem to is the Wigner–Seitz cell around a single atom, not the whole sample of alloy or pure metal. Let us make the logic clear by reducing the situation to one dimension in Fig. 1. The sample of Fe₃Al or iron extends, say, between $\pm A$, and the boundary condition there we might take as $\psi(\pm A) = 0$. The region (“container”) we are interested in is the atom extending from $-a$ to $+a$. To solve the Schrödinger equation for the whole system, we would integrate outwards with our computer from the origin to the boundary $|x| = a$ using some energy E , and inwards to the same boundary from the sample edge $|x| = A$. The eigenvalues E_n would be determined by varying E until the inward and outward solutions match in “logarithmic derivative” $L = \psi'/\psi$ at the boundary $|x| = a$, where ψ' is the derivative $d\psi/dx$. Rephrased slightly, the wave functions ψ_n inside our atom are obtained by solving the Schrödinger equation inside the atom, making the solutions fit a (very complicated) boundary condition $L(E)$ at the edge of the atom, where $L(E)$ is determined by the surrounding medium. Our theorem tells us that the local density of states (1.4), the whole Green function (1.6) in fact, constructed from these states, is independent of the boundary condition imposed by the medium. Rephrased further, the Green function and all physical properties of an atom are independent of the medium in which the atom finds itself.

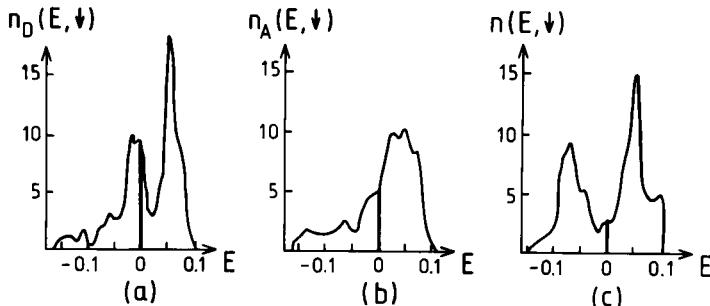


FIG. 2. Local density of states $n(E)$ for the minority-spin d band on one iron atom in atomic units.²⁰ The energy is measured relative to the Fermi level. (a) D site (eight iron n.n.) in Fe_3Al . (b) A site (from iron n.n.) in Fe_3Al . (c) Pure iron (after Haydock and You⁷).

Clearly we have proved a bit too much. Our theorem is in fact only an approximate theorem: in truth it contains the caveat *provided \mathbf{r} is a few wavelengths from the boundary*. The interpretation of “a few” depends on the accuracy desired, the further from the boundary the less is the Green function affected. This is beautifully illustrated by the Fe_3Al alloy. As mentioned in Section I,1, at an A site an iron atom has half its nearest neighbors as aluminum atoms, our boundary surrounds only that one atom, so that all points \mathbf{r} inside it are quite close to the boundary, and the magnetic moment differs by 35 % from when the atom is in a medium of pure iron. But at a D site the iron atom is surrounded by eight nearest neighbors of iron atoms and we can draw a boundary enclosing these nine iron atoms, to be compared with the corresponding cluster in pure iron. Now our central atom is a bit further from the boundary and its moment differs by less than 5 % from that of pure iron. The corrections are seen to be remarkably small. A general rule of thumb is that integrated properties arising from the whole electron gas are typically affected by the order of 10 % by changes in nearest neighbors and by 1 % by next-nearest neighbors. Differential properties like the local density of states at some given E are perturbed rather more. The similarities and differences between the densities of states in the three cases above are well seen in Fig. 2 in a model in which the 4s4p electrons on the iron atoms have been ignored (but those on aluminum in Fe_3Al included).²⁰

In speaking of “an approximate theorem,” we have not taken leave of rigorous mathematical logic: the simplified form of the result becomes approximate near the boundary, where correction terms arise that can be calculated explicitly. In the case of a localized impurity in a free-electron gas

²⁰ We use atomic units (a.u.) defined by $\hbar = m = e^2/4\pi\epsilon_0 = 1$, so that the Bohr radius of the hydrogen atom is the unit of length and the energy unit is 2 Rydbergs = 27.2 eV.

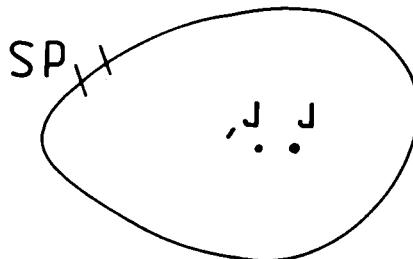


FIG. 3. Two points r and r' close together but far from the boundary (closed curve).

these are the well-known Friedel oscillations in the charge density around the impurity.^{8,11} We shall defer a proof of the theorem till Section II,4, merely pointing out here the following simple didactic demonstration. In Section I,3, below Eq. (3.8) we comment that G represents the propagation of an electron from r' to r . For the local density of states, Eq. (1.5), we want the limit $r' \rightarrow r$, and so consider in Fig. 3 the two points close to one another but far from the boundary. Since the latter is far from the two points, the propagation of an electron between them will not be affected by it: in principle there are reflection effects from the boundary but such reflections from different elements dS (Fig. 3) of the boundary will destructively interfere and cancel out since the boundary is many wavelengths away.

How small the effects are in our examples is really rather remarkable, for the wavelength of electrons is always comparable to atomic size, so that even in an atom surrounded by a shell of neighbors the distance to the boundary is only one wavelength or so. If we had "hard" boundary conditions, e.g., $\psi' = 0$ changing to an infinite potential barrier with $\psi = 0$, then the changes in local density of states and derived properties would be rather larger, but for an atom in a solid one is always matching onto a "soft" quasi-continuum, onto an electron gas whose density often hardly changes by a factor of two in altering the medium. The effects on G are then smaller.

The size of atoms illustrates both the truth and the lie in the theorem, as well as its power. The size of a given atom is remarkably constant, whether measured by bond length in open covalent structures or atomic volume in close-packed metallic materials or by hard sphere radius in ionic crystals.²¹ For example, Slater showed that from an atomic-sized point of view, even an ionic substance like NaCl, usually written as $\text{Na}^+ \text{Cl}^-$, can equally well be considered as $\text{Na}^0 \text{Cl}^0$ in terms of neutral atoms, where Na^0 has nearly the same size as in sodium metal.²¹ Of course, chemists and crystallographers

²¹ J. C. Slater, "Quantum Theory of Molecules and Solids," Vol. 2. McGraw-Hill, New York, 1965 (see pp. 95–117, esp. pp. 96–97 and 106–108).

are vitally concerned with the few percent of change in radius with the kind of bonding, but they are small.²¹ At the surface of a solid where half the atoms are totally cut away, no change of lattice spacing for the surface layer has yet been reliably measured, though changes of the order of 5% are indicated by some data. To appreciate the power of the theorem, consider how one would have to approach theoretically the question of atomic size or lattice constant in terms of the (extended) eigenstates of the whole system: one would have to solve for these states in two different alloys or compounds of the same element, and minimize the total energy so found with respect to atomic spacing. It is difficult to see any relation, such as additivity of radii, coming out of such an approach. Incidentally, if the invariance were exactly true, the argument would also be exactly self-consistent in the sense of the Hartree-Fock equations, for the same local density of states in two materials would yield the same self-consistent potential. That is another reason why the "invariance" property works so well.

The invariance property of the local density of states is least well fulfilled at energies near the bottom of a band (or indeed near any van Hove singularity⁸) because the electron wavelength tends to infinity there, so that the criterion of "several wavelengths away from the boundary" becomes less and less well satisfied. The effect is shown in Fig. 4a. Inglesfield^{22,23} considered a free-electron gas with Fermi level E_F , chosen to represent lithium metal. In it is a single spherical potential well to represent an impurity atom of magnesium. The well depth is a constant, equal to the difference in bandwidth between the two metals so that the Fermi level of the magnesium electron gas lines up with that in lithium. How much effect do the abrupt potential step and surrounding medium have? We plot the impurity density of states $n_i(E)$ defined as $n(E, \mathbf{r})$ integrated over the volume of the sphere representing the magnesium atom. In pure magnesium, $n_i(E)$ would be the usual parabola $\propto E^{1/2}$ and Fig. 4a shows that the difference from the parabola is greatest at low energy $E < 0.3E_F(\text{Li})$ as expected. But the nature of the difference is such as to conserve $n_i(E)$ as much as possible: in the dissolved case all states of the system and thus $n_i(E)$ are confined to the energy range of the lithium band, and the "spectral weight" of the parabola at $E < 0$ has been transferred to the region at small positive energy. When one calculates the wavefunctions in that energy range, one finds that their amplitude is substantially enhanced on the magnesium atom.²³ It is remarkable that $n_i(E)$ is already back to within 10% of the parabolic value for $E > 0.25E_F(\text{Li})$.

Figure 4b results when the well depth is increased to represent aluminum, being now large enough to pull out a bound state from the bottom of the

²² J. E. Inglesfield, *J. Phys. F* **2**, 63 (1972).

²³ J. E. Inglesfield, *J. Phys. F* **2**, 878 (1972).

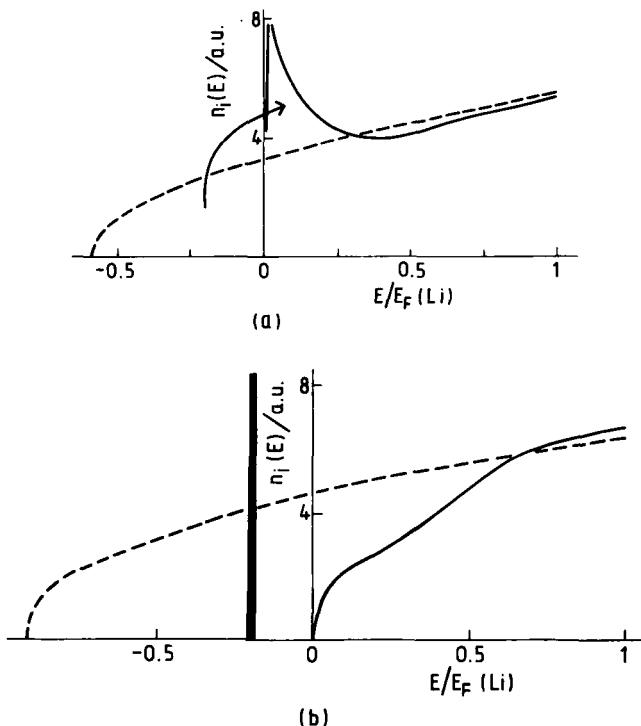


FIG. 4. Impurity density of states $n_i(E)$ (full line) for a square-well free-electron-gas model of a single impurity atom dissolved in lithium. (a) Magnesium impurity and (b) aluminum impurity. The dashed line is the density of states for one atom in pure magnesium or pure aluminum, respectively (Inglesfield²³).

band.²³ Thus $n_i(E)$ includes a delta function at negative E , which now replaces the lower portion of the free-electron parabola in pure aluminum. In fact, since it has to contain an integral number of electrons, two counting both spins, it actually robs some of the spectral weight from $E > 0$.

We may summarize this discussion by noting that the invariance theorem works quite well for the local density of states $n(E, \mathbf{r})$, but *it works even better for the local integrated density of states*

$$N(E, \mathbf{r}) = \int_{-\infty}^E n(E', \mathbf{r}) dE'. \quad (2.1)$$

Changes in boundary conditions do not so much add or subtract spectral weight from $n(E, \mathbf{r})$: rather they move it around a little so that the effects tend to cancel when one has integrated through them in (2.1) to a higher E .

The local density of states treats localized and extended states on an equal

footing. In the present instance the localized bound state can be recognized from the delta function in $n_i(E)$ (Fig. 4b). But in the case of localized surface states on a solid surface, one has a two-dimensional band of surface states giving a continuous $n(E, \mathbf{r})$ similar to that of the bulk states. One might hope to identify them by seeing whether $n(E, \mathbf{r})$ dies away as one goes away from the surface into the bulk, or one might consider $G(\mathbf{r}, \mathbf{r}', E)$ with $\mathbf{r}' \neq \mathbf{r}$. Nevertheless, the reader might feel disappointed that $n(E, \mathbf{r})$ fails to indicate such a qualitative difference in the nature of the states. Actually the difference is not as great as it seems. Returning to our example, the critical well depth D_c when a bound state first appears represents a mathematical singularity, separating a region $D > D_c$ with a bound state from a discretely different situation without one at $D < D_c$. (We consider the depth D a positive quantity.) Certainly this is so if one looks at the mathematical properties of the individual state. However Kohn and Majumdar²⁴ have proved a remarkable theorem that there is no singularity in the integrated density of states (2.1) taken up to some energy, say E_F . When the bound state appears as we increase D , there are compensating changes in the continuum states, cancelling out the singularity in any integrated property depending on the electron system as a whole. The presence or absence of localized states seems to be of less significance than might be expected.

Once aware of the invariance property, one sees examples everywhere. Soft x-ray spectra give a picture of $n(E, \mathbf{r})$ on an atom, decomposed according to different angular momenta l by segregating states of different l in Eq. (1.6) or operating on $G(\mathbf{r}, \mathbf{r}', E)$ with projection operators P_l . In any case the invariance property is not destroyed; it applies to each l component. It is therefore not surprising that in AlMg alloys the spectrum from the aluminum atoms has about the same bandwidth as in pure aluminum, and from the magnesium atoms as in pure magnesium. However, the original finding came as a puzzle.^{25,26} Certainly in the ordered concentrated alloys and even in the nonstoichiometric ones, the states of the system occupy an energy band that ought to be seen to have the same width on the two types of atom. But the spectral weight is such that the states at the bottom of the band have an enhanced weight on the aluminum atoms and a reduced one on the magnesium atoms.²⁷ Thus the Auger broadening will then give the aluminum spectrum the appearance of greater width at low E , the magnesium spectrum a smaller one.

²⁴ W. Kohn and C. Majumdar, *Phys. Rev.* **138**, A1617 (1965).

²⁵ C. Curry, in "Soft X-Ray Band Spectra" (D. J. Fabian, ed.), p. 174. Academic Press, New York, 1968.

²⁶ G. A. Rooke, in "Soft X-Ray Band Spectra" (D. J. Fabian, ed.), p. 186. Academic Press, New York, 1968.

²⁷ R. L. Jacobs, *Phys. Lett. A* **30**, 523 (1969); *J. Phys. F* **3**, L166 (1973).

The invariance theorem need not be applied always to a single atom. Friedel¹¹ first applied it in dilute alloys to the region occupied by solvent, excluding the solute atoms. The theorem shows that some properties pertaining to this region, e.g., the Fermi level and nuclear magnetic resonance, change less rapidly than linearly in the concentration.¹¹ Similarly, in cesium suboxides there are regions dominated by the ionic charge of clusters of three oxygen ions, but elsewhere the Cs-Cs interatomic spacing is similar to that of pure cesium metal, and hence in the latter regions we find an electron gas with bandwidth and density of states profile similar to that of pure cesium.²⁸ A further example is afforded by silicon surfaces. In the clean state in vacuum, a half-filled band of surface states pegs the Fermi level E_F at an energy about one-third up the gap. When a metal is evaporated onto the surface to make a rectifying contact, E_F at the surface is pegged at nearly the same energy, although no surface states can exist. Any energy level in the band gap of the insulator gives an ordinary conduction band state of the metal, but has an exponentially decaying tail stretching into the semiconductor. In a simple one-dimensional model, the spectral weight $n(E, \mathbf{r})$ of the tails was mainly in the lower half of the gap, and any charge in them would have an analogous electrostatic effect as in true surface states.²⁹ We do not enter into the subsequent discussion³⁰ as to whether or not this is the correct explanation of the observed phenomenon. Suffice it to note for present purposes that the mathematical model displayed to some extent the invariance property: the surface state with vacuum boundary conditions became tails of metal wave functions, spread out in energy but still concentrated in the lower half of the gap. (See also references in Section II,6.)

Finally a pointer to the development of Section II,4. We want a mathematical formulation of quantum theory that is in step with the physics, which starts off with the invariance property built into it as the zero-order approximation. Returning, for example, to aluminum (or magnesium) in the concentrated AlMg alloys, we want to write

$$G(\mathbf{r}, \mathbf{r}' \text{ on Al in alloy}) = G(\text{pure Al}) + (\text{correction terms}), \quad (2.2)$$

where the whole physics lies in the small correction terms. Effectively this means taking as our zero-order model of the alloy each atomic cell of aluminum identical to pure aluminum and each cell of magnesium in the alloy the same as in pure magnesium.³¹ It is a good zero-order approximation because of the invariance property, but requires sophisticated mathematics

²⁸ M. G. Burt and V. Heine, *J. Phys. C* **11**, 961 (1978).

²⁹ V. Heine, *Phys. Rev.* **138**, A1689 (1965).

³⁰ See, for example, J. C. Inkson, *Proc. Int. Conf. Phys. Semicond., 11th, 1972*, p. 1486 (1972); *J. Phys. C* **6**, 1350 (1973); S. G. Louie and M. L. Cohen, *Phys. Rev. B* **13**, 2461 (1976).

³¹ J. E. Inglesfield, *J. Phys. F* **9**, 1557 (1979).

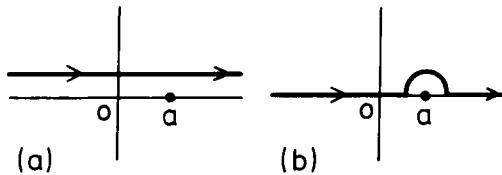


FIG. 5. Argand diagram and contours for energy integration.

because we have cobbled together pieces from two separate “unperturbed” problems, the pure aluminum metal and the pure magnesium: there is no single “unperturbed Hamiltonian.” Contrast this with the conventional pseudopotential perturbation approach: we are forced to start with an unperturbed free-electron gas of 2.5 electrons per atom for a 50% : 50% alloy, and most of the perturbation theory will be to give the aluminum and magnesium atoms their correct number of electrons, among which the difference between alloy and pure metal is lost! Such then is the method of matching Green functions that we develop in Sections II,4 and II,6.

3. BRIEF APPENDIX ON THE MATHEMATICS OF THE GREEN FUNCTIONS

We first prove the equivalence of the two forms of the Green function given in Eq. (1.6), namely,

$$G(\mathbf{r}, \mathbf{r}', E) = \sum_n \frac{\psi_n(\mathbf{r})\psi_n^*(\mathbf{r}')}{E + iO - E_n} \quad (3.1a)$$

$$= \mathcal{P} \sum_n \frac{\psi_n(\mathbf{r})\psi_n^*(\mathbf{r}')}{E - E_n} - i\pi \sum_n \psi_n(\mathbf{r})\psi_n^*(\mathbf{r}') \delta(E - E_n). \quad (3.1b)$$

The imaginary infinitesimal quantity iO in (3.1a) arises from the ambiguity of an expression like $1/(E - a)$ when it occurs in an integral,

$$\int_{-\infty}^{\infty} (E - a)^{-1} f(E) dE \quad (3.2)$$

because the integrand becomes meaningless (undefined, “infinite”) at $E = a$. Here $f(E)$ is any well-behaved function. Most simply one defines (3.2) by its principal value (denoted by \mathcal{P}) obtained by “cutting out” the offending piece from $a - \varepsilon$ to $a + \varepsilon$ from the range of integration, where ε is a very small quantity. However, one obtains richer mathematical properties if one considers E as a complex variable and integrates along a contour close to the real axis but skirting slightly above the pole at $E = a$. (Some differences of sign would arise if we chose to contour below the pole.) The contour can be the straight line $z = E(\text{real}) + ie$ (Fig. 5a), which is implied by Eq. (3.1a),

where the iO means that we take the limit as ε tends to zero. Alternatively, the contour can follow the real axis as in the definition of the principal value (Fig. 5b), plus a small semicircle along

$$z = a + \varepsilon \exp i\theta \quad (3.3)$$

carrying the contour around the pole. The semicircle is infinitesimally small but lies infinitesimally close to the pole, so that it makes a finite contribution. This is easily obtained as $-i\pi f(a)$ by making the substitution (3.3) for E and integrating from $\theta = \pi$ to $\theta = 0$. We thus obtain

$$(E + iO - a)^{-1} = \mathcal{P}(E - a)^{-1} - i\pi \delta(E - a), \quad (3.4)$$

and hence the second term of (3.1b).

A third interpretation is to absorb the iO in (3.1a) into E , which is treated as a complex variable roaming over the Argand diagram. This has its uses. We are interested in the value of G as E tends to the real axis from the positive imaginary side. Now a finite system of N atoms, which we compute with, has discrete energy levels separated by intervals $\Delta E \sim O(N^{-1})$. On the real axis these give $\text{Im } G$ the delta functions as in Eq. (3.1b), but a little further away at $\text{Im } E \gtrsim \Delta E$ they broaden into Lorentzians and add up to give a smooth continuous density of states. Thus computationally a simple (but not always best) way of calculating a local density of states is to evaluate some formula for G at a complex value of E with a suitably chosen small but nonzero imaginary part, e.g., the continued fractions in Section IV.11.

The basic equation satisfied by G is

$$(E - \mathcal{H})G(\mathbf{r}, \mathbf{r}', E) = \delta(\mathbf{r} - \mathbf{r}'), \quad (3.5)$$

where \mathcal{H} operates on \mathbf{r} . This follows readily from (3.1a): operating with $E - \mathcal{H}$ on ψ_n gives $(E - E_n)\psi_n$, thus canceling the denominator, and $\sum_n \psi_n(\mathbf{r})\psi_n^*(\mathbf{r}')$ is the unit operator $\delta(\mathbf{r} - \mathbf{r}')$, analogous to Eq. (1.12).

G , being a function of three variables $\mathbf{r}, \mathbf{r}', E$, contains a large amount of information. Consider the solution of (3.5) in a region of \mathbf{r} space excluding the point $\mathbf{r} = \mathbf{r}'$: in this region the right-hand side of (3.5) is zero, so that (3.5) reduces to the Schrödinger equation (1.1). Indeed it is evident from (3.1a) that G contains information about all the eigenstates ψ_n of the Schrödinger equation. But G contains more: it contains information about the solution of the Schrödinger equation for any value of E in (3.5), not only the energy eigenvalues corresponding to the usual solutions ψ_n satisfying appropriate boundary conditions. Thus in a solid for E in a band gap, the $\delta(\mathbf{r} - \mathbf{r}')$ in (3.4) generates in G the evanescent solutions used in describing surface states.

With a Fourier transform we can define

$$G(\mathbf{r}, \mathbf{r}', t) = (2\pi i)^{-1} \int_{-\infty}^{\infty} G(\mathbf{r}, \mathbf{r}', E) \exp(-iEt/\hbar) dE, \quad (3.6)$$

which has the following physical significance. If at time $t = 0$ we have any initial state $\psi_i(\mathbf{r})$, then its development in time for $t > 0$ is given by

$$\psi_i(\mathbf{r}, t) = \int G(\mathbf{r}, \mathbf{r}', t) \psi_i(\mathbf{r}') d^3\mathbf{r}'. \quad (3.7)$$

This result may be proved very readily by noting that for $t > 0$, Eq. (3.6) may be turned into a contour integral by adding a large semicircle below the real axis, and then evaluated in terms of the residues of (3.1a) at the poles. We obtain

$$G(\mathbf{r}, \mathbf{r}', t > 0) = \sum_n \psi_n(\mathbf{r}) \exp(-iE_n t/\hbar) \psi_n^*(\mathbf{r}') \quad (3.8)$$

from which (3.7) follows. We conclude that $G(\mathbf{r}, \mathbf{r}', t)$ represents *the propagation of an electron*. Indeed, we could as a special case take the initial state as the delta function at $\mathbf{r} = \mathbf{R}$ corresponding to an electron placed into the system at \mathbf{R} : then $G(\mathbf{r}, \mathbf{R}, t)$ would literally be its subsequent spreading out.

When wave functions are expressed in terms of a linear combination of atomic orbitals, as in Eqs. (1.8)–(1.11), we obtain the *matrix formulation of quantum mechanics*. The Green function becomes the matrix defined as

$$G(E) = [(E + iO)I - \mathcal{H}]^{-1} \quad (3.9)$$

by analogy to (3.1a), where I is the unit matrix and usually omitted as understood. G satisfies the matrix equation

$$[E - \mathcal{H}]G(E) = I, \quad (3.10)$$

in complete analogy to (3.5). Indeed (3.5) can be interpreted as saying that G is the inverse of the operator $(E - \mathcal{H})$: while \mathcal{H} is a differential operator, G in (3.7) acts as an integral operator.

II. The Method of Matching the Green Functions

4. DERIVATION OF THE GREEN FUNCTION

We already outlined at the end of Section I,2 the type of theory we need. Our system, say a solid with a surface, consists of two regions A and B (Fig. 6), the perfect periodic solid in A and vacuum in B . We know the solution of²⁰

$$[-\frac{1}{2}\nabla^2 + V(\mathbf{r}) - E]G(\mathbf{r}, \mathbf{r}', E) = -\delta(\mathbf{r} - \mathbf{r}') \quad (4.1)$$

in the perfect solid, G_A say, and the solution G_B in pure vacuum. Surely it must be possible to match the two solutions together to obtain a complete solution to (4.1), directly analogous to matching solutions of the Schrödinger

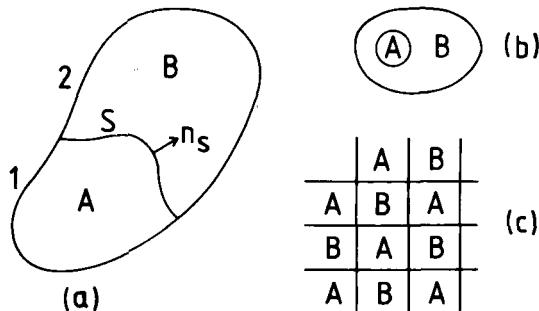


FIG. 6. Schematic view of a two-part system. (a) The boundary S divides the system between regions A and B . The outer line denotes the outer boundary of the whole system, usually taken at infinity. (b, c) Different types of boundary between A and B (after Inglesfield³⁵).

equation in two regions to give an energy value for a discrete state or a wave function in a continuum.

Besides the surface of a solid, examples of such systems might be a diatomic molecule³² or larger molecule³³ by extension of the method, an isolated impurity in a solid^{22,23} (Fig. 6b), or an ordered alloy³¹ in which case the regions A and B become interlacing checkerboard patterns (Fig. 6c). A solid with an adsorbed atom would be another case, or a microcrystal of some material.

To be precise, G_A is the solution of

$$[-\frac{1}{2}\nabla^2 + V_A - E]G_A = -\delta(\mathbf{r} - \mathbf{r}'), \quad (4.2)$$

with medium A filling the whole system: and similarly for G_B . These are relatively simple to obtain, for example, from the Bloch functions in the infinite solid³⁴ or the plane waves in empty space, the vacuum. Now (4.2) is identical to (4.1) in region A so that G_A also satisfies (4.1) there, i.e., for \mathbf{r} and \mathbf{r}' in A we have

$$[-\frac{1}{2}\nabla^2 + V - E]G_A(\mathbf{r}, \mathbf{r}', E) = -\delta(\mathbf{r} - \mathbf{r}'). \quad (4.3)$$

However G and G_A are not the same because they would satisfy different boundary conditions on the boundary S (Fig. 6), and the difference between G and G_A would be, as usual, a solution F of the corresponding homogeneous equation (in A)

$$G = G_A + F, \quad (4.4)$$

$$[-\frac{1}{2}\nabla^2 + V - E]F = 0. \quad (4.5)$$

³² J. E. Inglesfield, *Mol. Phys.* **37**, 873 (1979).

³³ J. E. Inglesfield, *J. Chem. Phys.* **67**, 505 (1977).

³⁴ F. Garcia-Moliner, V. Heine, and J. Rubio, *J. Phys. C* **2**, 1797 (1969).

Now (4.5) in region A is identical to (4.1) in region A if we take the point \mathbf{r}' outside A , i.e., F is also a Green function. Thus in one form or another the Green functions contain all that is necessary.

We now proceed to derive the solution of (4.1) in terms of G_A and G_B , following Inglesfield,^{22,23,35} and Velicky and Bartos,³⁶ extending the earlier ideas of Garcia-Moliner and Rubio.^{37,38} After the change of variable \mathbf{r} to \mathbf{r}'' in Eq. (4.1) and \mathbf{r} to \mathbf{r}'' and \mathbf{r}' to \mathbf{r} in (4.3) these equations become

$$[-\frac{1}{2}\nabla_{\mathbf{r}''}^2 + V(\mathbf{r}'') - E]G(\mathbf{r}'', \mathbf{r}', E) = -\delta(\mathbf{r}'' - \mathbf{r}'), \quad (4.6a)$$

$$[-\frac{1}{2}\nabla_{\mathbf{r}''}^2 + V(\mathbf{r}'') - E]G_A(\mathbf{r}, \mathbf{r}'', E) = -\delta(\mathbf{r} - \mathbf{r}''), \quad (4.6b)$$

for $\mathbf{r}, \mathbf{r}', \mathbf{r}''$ all in A , where in (4.6b) we have also used the property

$$G(\mathbf{r}, \mathbf{r}', E) = G(\mathbf{r}', \mathbf{r}, E), \quad (4.7)$$

which is the reciprocity theorem of wave propagation between two points. Equation (4.7) follows from the definition (3.1a) by noting that the ψ_n are real if one has real boundary conditions such as $\psi = 0$, or in the case of periodic boundary conditions to every solution ψ_n there exists a "time-reversed" solution ψ_n^* . The property can also be verified for other types of boundary conditions such as "outward-going waves." Now multiply (4.6a) by $G_A(\mathbf{r}, \mathbf{r}'')$ (we leave the variable E as understood) and (4.6b) by $G(\mathbf{r}'', \mathbf{r}')$, subtract, and integrate \mathbf{r}'' over region A using Green's theorem

$$\iiint (\phi \nabla^2 \psi - \psi \nabla^2 \phi) d^3 \mathbf{r} = \int (\phi \nabla \psi - \psi \nabla \phi) d\mathbf{S} \quad (4.8)$$

to obtain

$$G(\mathbf{r}, \mathbf{r}') = G_A(\mathbf{r}, \mathbf{r}') + \frac{1}{2} \int dS \left[\frac{\partial G_A}{\partial n_S}(\mathbf{r}, \mathbf{r}_S) G(\mathbf{r}_S, \mathbf{r}') - G_A(\mathbf{r}, \mathbf{r}_S) \frac{\partial G}{\partial n_S}(\mathbf{r}_S, \mathbf{r}') \right], \quad [\mathbf{r}, \mathbf{r}' \text{ both in } A], \quad (4.9a)$$

where $\partial/\partial n_S$ denotes the normal component of grad across the surface S (Fig. 6) and \mathbf{r}_S is a point on S . The surface integral over the outer boundary 1

³⁵ J. E. Inglesfield, *J. Phys. C* **4**, L14 (1971).

³⁶ B. Velicky and I. Bartos, *J. Phys. C* **4**, L104 (1971).

³⁷ F. Garcia-Moliner and J. Rubio, *J. Phys. C* **2**, 1789 (1969).

³⁸ F. Garcia-Moliner and J. Rubio, *Proc. R. Soc. London, Ser. A* **324**, 257 (1971).

in Fig. 6a vanishes by assuming G and G_A to have the same boundary condition there, e.g., $G = 0$, or $\partial G / \partial n_S = 0$. Similarly, for \mathbf{r}, \mathbf{r}' in region B we have

$$\begin{aligned} G(\mathbf{r}, \mathbf{r}') &= G_B(\mathbf{r}, \mathbf{r}') - \frac{1}{2} \int dS \left[\frac{\partial G_B}{\partial n_S}(\mathbf{r}, \mathbf{r}_S) G(\mathbf{r}_S, \mathbf{r}') \right. \\ &\quad \left. - G_B(\mathbf{r}, \mathbf{r}_S) \frac{\partial G}{\partial n_S}(\mathbf{r}_S, \mathbf{r}') \right], \quad [\mathbf{r}, \mathbf{r}' \text{ both in } B], \end{aligned} \quad (4.9b)$$

where the change of sign from (4.9a) is due to the fact that n_S is now an *inward* pointing normal into region B (Fig. 6a). To complete the specification of the Green function we also have to consider \mathbf{r} and \mathbf{r}' in different regions. With \mathbf{r}'' in B and \mathbf{r}' still in A , Eq. (4.6a) becomes

$$[-\frac{1}{2}\nabla_{\mathbf{r}''}^2 + V(\mathbf{r}'') - E]G(\mathbf{r}'', \mathbf{r}') = 0, \quad (4.10a)$$

where $\delta(\mathbf{r}' - \mathbf{r}'')$ is always zero on the right-hand side when $\mathbf{r}', \mathbf{r}''$ are in different regions. Instead of (4.6b) we have with \mathbf{r} in B :

$$[-\frac{1}{2}\nabla_{\mathbf{r}''}^2 + V(\mathbf{r}'') - E]G_B(\mathbf{r}, \mathbf{r}'') = \delta(\mathbf{r} - \mathbf{r}''). \quad (4.10b)$$

A procedure similar to the one before gives the analog to (4.9a):

$$\begin{aligned} G(\mathbf{r}, \mathbf{r}') &= -\frac{1}{2} \int dS \left[\frac{\partial G_B}{\partial n_S}(\mathbf{r}, \mathbf{r}_S) G(\mathbf{r}_S, \mathbf{r}') \right. \\ &\quad \left. - G_B(\mathbf{r}, \mathbf{r}_S) \frac{\partial G}{\partial n_S}(\mathbf{r}_S, \mathbf{r}') \right], \quad [\mathbf{r} \text{ in } B, \mathbf{r}' \text{ in } A], \end{aligned} \quad (4.9c)$$

with a similar relation for \mathbf{r} in A , \mathbf{r}' in B .

Equation (4.9) gives us a formal solution for G in the desired form (2.2) but is useless as it stands because it also contains $G(\mathbf{r}_S, \mathbf{r}')$ and $\partial G / \partial n_S$ on the right-hand side, for which we must first solve. One relation for them is obtained by letting \mathbf{r} tend to the boundary in (4.9a), where we put $\mathbf{r} = \mathbf{r}_S''$:

$$\begin{aligned} G(\mathbf{r}_S'', \mathbf{r}') &= G_A(\mathbf{r}_S'', \mathbf{r}') + \frac{1}{2} \int dS \left[\frac{\partial G_A}{\partial n_S}(\mathbf{r}_S'', \mathbf{r}_S) G(\mathbf{r}_S, \mathbf{r}') \right. \\ &\quad \left. - G_A(\mathbf{r}_S'', \mathbf{r}_S) \frac{\partial G}{\partial n_S}(\mathbf{r}_S, \mathbf{r}') \right]. \end{aligned} \quad (4.11)$$

The other relation comes from using (4.9c) with \mathbf{r} in region B and letting \mathbf{r} tend to \mathbf{r}_S'' from that side of the boundary, which is how the matching of solutions comes in:

$$\begin{aligned} G(\mathbf{r}_S'', \mathbf{r}') &= -\frac{1}{2} \int dS \left[\frac{\partial G_B}{\partial n_S}(\mathbf{r}_S'', \mathbf{r}_S) G(\mathbf{r}_S, \mathbf{r}') \right. \\ &\quad \left. - G_B(\mathbf{r}_S'', \mathbf{r}_S) \frac{\partial G}{\partial n_S}(\mathbf{r}_S, \mathbf{r}') \right]. \end{aligned} \quad (4.12)$$

The two simultaneous integral equations (4.11) and (4.12) can be solved for $G(\mathbf{r}_S, \mathbf{r}')$ and $\partial G / \partial n_S$, which are then substituted in the right-hand side of (4.9).

It helps to write Eqs. (4.11) and (4.12) symbolically in operator form^{35,37}:

$$\mathcal{g} = \mathcal{g}_A + \frac{1}{2}\mathcal{G}'_A\mathcal{g} - \frac{1}{2}\mathcal{G}_A\mathcal{g}', \quad (4.13a)$$

$$\mathcal{g}' = -\frac{1}{2}\mathcal{G}'_B\mathcal{g} + \frac{1}{2}\mathcal{G}_B\mathcal{g}', \quad (4.13b)$$

where $\mathcal{g}, \mathcal{g}_A, \mathcal{g}'$ are vectors in a space $\{\mathbf{r}_S\}$ with components $G(\mathbf{r}_S, \mathbf{r}')$, $G_A(\mathbf{r}_S, \mathbf{r}')$, $\partial G(\mathbf{r}_S, \mathbf{r}') / \partial n_S$ with \mathbf{r}' fixed, while $\mathcal{G}_A, \mathcal{G}'_A$, etc., are operators in the space with matrix elements $G_A(\mathbf{r}_S'', \mathbf{r}_S)$, $\partial G_A(\mathbf{r}_S'', \mathbf{r}_S) / \partial n_S$, etc. The solution of (4.13a) and (4.13b) solved as a pair of linear equations with care about the order of operators is

$$\mathcal{g} = [\mathcal{G}_A^{-1}(1 - \frac{1}{2}\mathcal{G}'_A) + \mathcal{G}_B^{-1}(1 + \frac{1}{2}\mathcal{G}'_B)]^{-1}\mathcal{G}_A^{-1}\mathcal{g}_A, \quad (4.14a)$$

$$\mathcal{g}' = [(1 - \frac{1}{2}\mathcal{G}'_A)^{-1}\mathcal{G}_A + (1 + \frac{1}{2}\mathcal{G}'_B)^{-1}\mathcal{G}_B]^{-1}(1 - \frac{1}{2}\mathcal{G}'_A)^{-1}2\mathcal{g}_A, \quad (4.14b)$$

which may be inserted in (4.9) to give our final answer.

Of course, the solution (4.14) exists meaningfully only if the operators in square brackets actually have inverses. It is easy to show by taking out factors of $\mathcal{G}_A^{-1}, (1 - \frac{1}{2}\mathcal{G}'_A), \mathcal{G}_B^{-1}, (1 + \frac{1}{2}\mathcal{G}'_B)$ in the right order that the operators in square brackets in (4.14a) and (4.14b) are in fact proportional to one another, so that we may refer to either one indiscriminately, or to any other intermediate form obtained by using some of the factors. However, no inverse exists if the determinant vanishes:

$$\det |[\text{operator in (4.14a) or (4.14b)}]| = 0. \quad (4.15)$$

When this happens, \mathcal{g} and \mathcal{g}' from (4.14) and hence G from (4.9) have a pole. Thus from (3.1), Eq. (4.15) is the condition for a discrete energy level of the Schrödinger equation for the combined $A + B$ system. Incidentally, both \mathcal{G}_A and \mathcal{G}'_A have poles at the energy levels of the system containing medium A everywhere, so that these poles always cancel in the operators in (4.14) and similarly for the poles of $\mathcal{G}_B, \mathcal{G}'_B$.

The group of Garcia-Moliner^{34,37,38} and Inglesfield^{22,23,31} have shown that Eqs. (4.14) are not as formidable as they look. For a plane boundary all the quantities may be expanded in plane waves in terms of the reciprocal lattice vectors, resulting in the need to invert a matrix of modest size. This case applies both to a plane boundary at a surface and to the faces of the Wigner-Seitz cells as in Fig. 6c. The inversion is even simpler for a spherical boundary, such as an atomic sphere (Fig. 6b): one expands²² all quantities in terms of spherical harmonics, there is no mixing between different l, m , and no matrix algebra is required. In some cases it is even convenient to make further analytic simplifications.²² We can define G_A in terms of states ψ_n satisfying the boundary condition $\partial\psi_n / \partial n_S = 0$ on surface S instead of any

particular boundary condition on boundary 2 in Fig. 6a, so that $\mathcal{G}'_A = 0$. Similarly, we can make $\mathcal{G}'_B = 0$, and then Eqs. (4.14) simplify enormously. In particular, from (4.14b) the condition (4.15) for an energy eigenvalue becomes

$$\det |\mathcal{G}_A + \mathcal{G}_B| = 0, \quad (4.16)$$

which Garcia-Moliner and Rubio originally derived for a plane boundary.³⁷ When region A is completely surrounded by B (Fig. 6b) the outer boundary “1” in Fig. 6a disappears and with it the requirement of G and G_A having the same boundary condition there, thus adding a further freedom to our choice of G_A such as the possibility of using running outgoing waves.²²

5. BOUNDARY CORRECTION TERMS IN THE INVARIANCE THEOREM

Equation (4.9) constitutes our *proof of the invariance theorem*, because it has the required form

$$G(\mathbf{r}, \mathbf{r}') = G_A(\mathbf{r}, \mathbf{r}') + \text{boundary corrections}, \quad (5.1)$$

where \mathbf{r}, \mathbf{r}' are in region A and the boundary terms can be evaluated by substituting for G and $\partial G / \partial n_S$ on the surface from (4.14). The form of (5.1) has already been presented in Eq. (2.2) and discussed there, the connection with the local density of states being defined by (1.5).

It remains to consider the *boundary correction terms*. Clearly, no general statement about their size can be made. As an extreme case, if we throw up an infinite potential barrier, it will reduce ψ to zero there and hence produce a 100% change in $n(E, \mathbf{r})$ at the barrier. Examples of less drastic changes in boundary condition have already been shown in Figs. 2 and 4, and more are contained in Section II,6.

What we can do is to show that the correction terms decrease as we retreat away from the boundary and to derive their asymptotic form.³⁹ We proceed by cavalier approximations to obtain an answer as speedily as possible, and then consider the question of its wider applicability. In accordance with Eq. (1.5) we set $\mathbf{r} = \mathbf{r}'$, which enters the surface integral in (4.9) through $G_A(\mathbf{r}, \mathbf{r}_S)$ directly and through $\mathcal{G}_A = G_A(\mathbf{r}_S, \mathbf{r})$ in (4.14). For an electron in a constant (zero) potential we have

$$G_A(\mathbf{r}, \mathbf{r}', E) = \frac{\exp(i\kappa|\mathbf{r} - \mathbf{r}'|)}{4\pi|\mathbf{r} - \mathbf{r}'|}, \quad (5.2)$$

³⁹ The author knows of no general discussion in the literature.

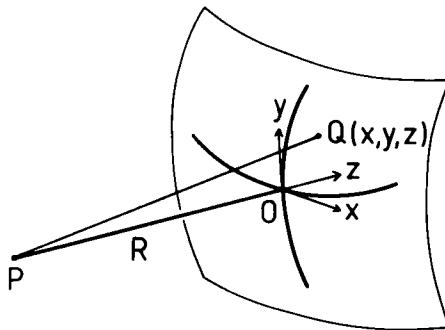


FIG. 7. Construction for integration of correction terms from the boundary surface.

where $E = \frac{1}{2}\kappa^2$ in atomic units,²⁰ and we see that $\partial G/\partial n_s$ has the same form as (5.2) for large $|\mathbf{r} - \mathbf{r}'|$ apart from factors of κ and angular terms all of which we drop. Thus the surface integral in (4.9) has the structure

$$\iint G_A(\mathbf{r}, \mathbf{r}_s) K(\mathbf{r}_s, \mathbf{r}'_s) G_A(\mathbf{r}'_s, \mathbf{r}) d^2\mathbf{r}_s d^2\mathbf{r}'_s, \quad (5.3)$$

where K is the integral representation for the combination of surface operators preceding φ_A in (4.14). Let us approximate K by a delta function! The surface correction then becomes

$$I(\mathbf{r}) = \int [G_A(\mathbf{r}, \mathbf{r}_s)]^2 d^2\mathbf{r}_s. \quad (5.4)$$

In Fig. 7, P is the point \mathbf{r} , PO the perpendicular of length R to the surface, $Q(x, y, z)$ the surface point \mathbf{r}_s with respect to origin O , and the x and y axes are taken along the principal directions of curvature on the surface. Near O the shape of the surface is defined by

$$z(x, y) = (x^2/2R_1) + (y^2/2R_2), \quad (5.5)$$

where R_1 and R_2 are the principal radii of curvature. With (5.2), Eq. (5.4) becomes (dropping further constant factors)

$$I(R) = \iint \frac{\exp(i2\kappa|PQ|)}{|PQ|^2} dx dy, \quad (5.6)$$

with

$$\begin{aligned} |PQ| &= [(R + z)^2 + x^2 + y^2]^{1/2} \\ &= R + z + (x^2 + y^2)/2R + O(z^2/R) + O(R^{-2}). \end{aligned} \quad (5.7)$$

Substituting (5.5) we obtain

$$I(R) = \frac{\exp(i2\kappa R)}{R^2} \iint \frac{\exp[i\phi(x, y)]}{D} dx dy, \quad (5.8)$$

where

$$\phi(x, y) = \frac{\kappa x^2}{R'_1} + \frac{\kappa y^2}{R'_2}, \quad (5.9)$$

with

$$\frac{1}{R'_1} = \frac{1}{R_1} + \frac{1}{R}, \quad \frac{1}{R'_2} = \frac{1}{R_2} + \frac{1}{R}, \quad (5.10)$$

$$D = 1 + O(x^2/R^2). \quad (5.11)$$

D is approximated as unity and the integral in (5.8) becomes

$$\begin{aligned} & \int_{-\infty}^{\infty} \exp(i\kappa x^2/R'_1) dx \int_{-\infty}^{\infty} \exp(i\kappa y^2/R'_2) dy \\ &= [(1 + i)(\pi R'_1/2\kappa)^{1/2}] [(1 + i)(\pi R'_2/2\kappa)^{1/2}]. \end{aligned} \quad (5.12)$$

Two principal cases now arise. If we are dealing with an atom or small cluster, R_1 and R_2 are of atomic dimensions and dominate R'_1 and R'_2 in (5.10) so that (5.12) becomes a constant and the asymptotic behavior of the boundary corrections is given by the prefactor $\exp(i2\kappa R)/R^2$ in (5.8). If, however, we have a plane surface, R_1 and R_2 are infinite, so that $R'_1 = R'_2 = R$, giving from (5.12) an extra factor of R . The term $\exp(i2\kappa R)$ in (5.8) will be multiplied by some constant that will be a complex number, yielding from (1.5) a sinusoidal variation with a phase shift γ . *Our final result for the asymptotic behavior away from the boundary is therefore*

$$n(E, \mathbf{r}) = n_A(E, \mathbf{r}) + (\text{const}) \frac{\sin(2\kappa R + \gamma)}{R \text{ or } R^2}, \quad (5.13)$$

where in the denominator R refers to a plane boundary and R^2 to a curved one. The former behavior (plane boundary) can easily be checked for a free-electron gas near a plane infinite barrier, whereas the Friedel oscillations^{8,11} around an impurity atom form an example of a curved boundary.

We now go back to inquire into the effect of our approximations. The basic point is that only the central area around O (Fig. 7) contributes significantly to the integrals in (5.3) and (5.8). In (5.12), the integrand oscillates more and more rapidly as x increases, consecutive positive and negative half cycles canceling. Thus it is only the contribution from the first half-cycle that effectively determines the value of the integral, and its order of magnitude

can be estimated that way. The situation is completely analogous to the construction of Fresnel half-period zones or the Cornu spiral in the traditional theory of diffraction. Slowly varying factors will not alter the approximate magnitude of the integral, e.g., the denominator in (5.8), angular and area factors in (5.4) that we have dropped or the higher order terms in (5.7). For a small boundary such as the impurity atom, the integrals in (5.3) and (5.4) only extend as far as the first or second Fresnel half-period zone. Going back now to (5.3), we have each factor of G_A contributing a factor of $\exp(i\kappa R)/R$ from the variation along PO , and the remainder of the variation over the surface being described by the same set of half-period Fresnel zones. The size of the latter are independent of R (for large R) on a curved surface so that the surface integrals in (5.3) contribute some result independent of R , thus giving the same asymptotic behavior as before. With a plane surface the half-period zones increase as $R^{1/2}$, which becomes large compared with the range of K , the latter determined by the electron wavelength $2\pi/\kappa$: thus K behaves as a delta function compared with the size of the zones and again we recover our previous result. In the presence of a periodic potential, κ must be replaced by $|\phi|$, where ϕ is determined by the band structure $E(\phi)$ as the wave vector for propagation in the direction PO with energy E . In the limit of large R , the line PQ (Fig. 7) makes an infinitesimally small angle with PO when Q lies within the first few half-period zones, so that the shape of the energy contours $\mathbf{k}(E)$ is immaterial. *With this interpretation for κ , our result (5.13) is therefore applicable in the presence of a periodic potential and is of quite general validity.* In the case of a *disordered medium* with scattering, the Green function decreases exponentially and the boundary term in (5.13) must be multiplied by a further factor of $\exp(-2R/\Lambda)$, where Λ is the mean free path. However, when E tends to a band edge or other Van Hove singularity, κ effectively goes to zero, the range of propagation described by K and the size of the Fresnel zones become large together, and the variation of the phase becomes slow instead of rapid, so that *a more sophisticated analysis is required, particularly to establish the magnitude of the boundary corrections as a function of E near band edges and band saddle points.*

6. APPLICATIONS TO ALLOYS, SURFACES, AND MOLECULES

In this section we survey how the method of matching Green functions has been developed in order to serve various types of applications to alloys, surfaces, and molecules, in that order. The method represents an approach, a formalism, which can be shaped in a variety of directions. Broadly speaking, the calculations carried out to date only demonstrate the possibilities of the method: much remains to be done to exploit it fully.

Inglesfield has considered *dilute alloys* (magnesium and aluminum in lithium,²³ copper in aluminum⁴⁰) and an *ordered concentrated alloy*³¹ (LiMg). The calculations have all involved some element of approximation or model, as is natural in the first applications of a new method, but it is usually clear how one can do better. In any case some of the approximations are probably very good, e.g., matching over a Wigner–Seitz sphere instead of an atomic polyhedron for the dilute alloy. There is always an arbitrariness about where to define the “boundary” between dissimilar atoms, which in this case is the surface over which to make an abrupt change of potential. The atomic sphere and polyhedron are probably equally good: what has not yet been explored is the practicality of Green function perturbation theory to obtain a smooth, self-consistent potential. In any case the approximations used may be entirely adequate for some purposes, e.g., to understand the systematic changes in electronic structure as seen in soft x-ray spectra.

With this caveat about the use of approximations, the main impression one obtains from the papers is how manageable Eqs. (4.9) and (4.14) are, despite their ferocious look. Bound states from (4.15) can be handled, as well as band states.²² Various tricks help: since region A in the sense of Fig. 6b, c has no boundary at infinity, any boundary condition can be chosen and the choice $\partial\psi_A/\partial n_S = 0$ eliminates $\partial G_A/\partial n_S$ from (4.9) and G'_A from (4.14). For nearly free-electron elements, G_A can be expressed to first order in the pseudo-potential by perturbation theory in closed form,²³ aside from one numerical integral, which can be tabulated since it applies to any element with the same crystal structure. Some integrals with respect to energy can also be carried out analytically.²² As already remarked, all Green functions can be expanded in spherical harmonics on a matching sphere, which simplifies the inversions in (4.14) enormously. Such then are some of the technical tricks of the trade.

In Fig. 4 we have already referred to calculations of magnesium and aluminum impurities in lithium using a free-electron model. We now present some results for the analogous calculations with the full atomic pseudo-potential and band structure included.²³ Figure 8 shows the actual $n(E, \mathbf{r})$ at a given E , spherically averaged but not integrated over the atomic cell. In Fig. 8a for magnesium in lithium and E in the middle of the band, we see $n(E, \mathbf{r})$ resembling that of pure magnesium and pure lithium but joined smoothly together, as we would expect from Fig. 4 and all that has been said in Section I,2. However, for aluminum in lithium at $E = E_F(\text{Li})$, Fig. 8b shows a new effect (also seen for magnesium in lithium at this energy but less pronounced). The energy is close to the bottom of the band gap in lithium, where the Bloch states are the p-like combination. These are standing waves

⁴⁰ J. E. Inglesfield, *J. Phys. F* **3**, 1921 (1973).

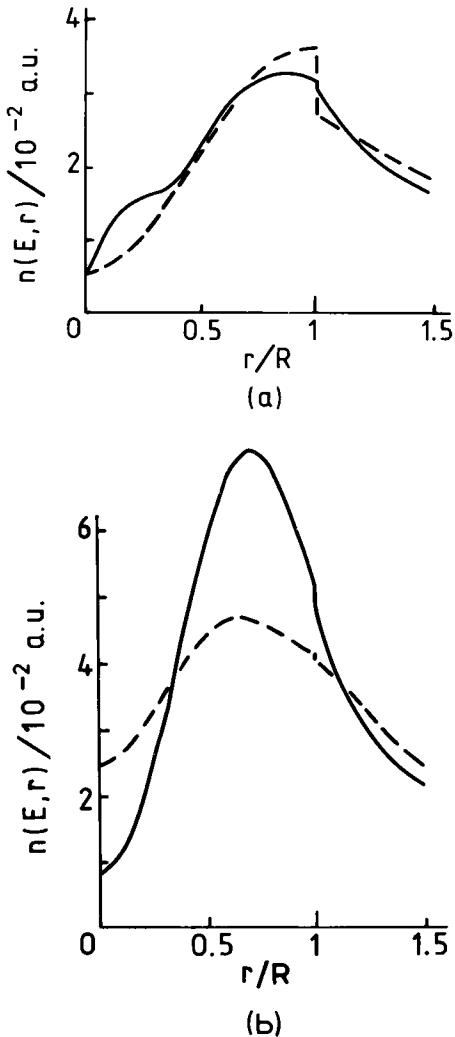


FIG. 8. Local density of states $n(E, r)$ (full line) for an impurity atom dissolved substitutionally in lithium, compared with the values (broken line) in the pure constituents. Note that the small discontinuity in the full line at the radius R of the Wigner Seitz sphere ($R = 3.26$ a.u.) results from approximations in the calculation. (a) For magnesium impurity at $E = 0.5E_F(\text{Li})$. (b) For aluminum impurity at $E = E_F(\text{Li})$ (after Inglesfield²³).

locked to the atomic positions of the lithium lattice and cannot be phase-shifted like free waves. They therefore induce a similar p-like wave function on the aluminum impurity, so that we see $n(E, \mathbf{r})$ at this energy more heaped up at the cell boundary and smaller at the origin than in pure aluminum. There are no soft x-ray data for aluminum in lithium but one would expect to see such an effect in $n(E, \mathbf{r})$, which is reminiscent of the kind of changes in band shape often seen in alloy systems.

When considering the total charge density $\rho(\mathbf{r})$ one finds what at first sight appears to be substantial charge shifts, e.g., of $0.22e$ for magnesium in lithium. But in reality the charge is taken from just inside the magnesium atomic cell and transferred by only about 1 a.u. (0.53×10^{-10} m) to just outside the cell, resulting in a dipole layer with a potential drop of 0.02 a.u. across it.²³ This is a contribution to the self-consistent potential in the alloy and is approximately the magnitude needed to match up the Fermi levels of the two constituents. Our picture of the magnesium impurity atom is therefore like an atomic sphere of pure magnesium surrounded by a thin screening dipole layer over its surface whose thickness (the Thomas Fermi screening length) is small compared with the atomic radius⁴¹ (3.3 a.u. in lithium). Thus the original model with an abrupt change of potential is not too far from self-consistency. For this reason the heats of solution come out small, too small to be calculated successfully by the approximations employed in these first calculations, but Inglesfield could analyze clearly why they are small.²³ The same applies to copper impurities in aluminum.⁴⁰

Inglesfield's study of an ordered alloy³¹ (LiMg) is most interesting and pregnant with ideas for development. In the very simplest approximation for matching the Green functions between adjacent atoms, he obtains equations analogous to those of the tight-binding (LCAO) method but with parameters determined by the calculated G_A for the atom or pure solid. Thus some of the calculational methods of Sections IV,11 and IV,12, which have been developed for the tight-binding case, can be utilized. The simple tight-binding form is just a first step toward matching and can be extended. It seems one can determine either the complete $G(\mathbf{r}, \mathbf{r}', E)$ or for the ordered alloy extract the band structure $E(\mathbf{k})$. It should be noted, however, that the quantities being joined together are Green functions G_A , G_B , and not atomic wave functions. The Green functions contain much more information, and an "empty lattice test," for example, gives a very good approximation to a free-electron parabola, not an $E(\mathbf{k})$ having cosine form as is usual in the tight-binding approximation in band theory. Another trick one can play is to develop G to represent the LiMg alloy as an effective medium and then to calculate the electronic structures of lithium and magnesium as "impurities"

⁴¹ See also the discussion in Heine and Weaire,³ p. 427.

in such a medium. Inglesfield finds this quite a good approximation to the correct results except for energies near the bottom of the band, a region that turns out to be very important for charge transfer and hence self-consistency.³¹

The method of matching Green functions has found some application in the theory of *semiconductor-vacuum surfaces*, the rectifying *semiconductor-metal junctions*, and *semiconductor-oxide interfaces*. The secular equation (4.15) can be used to calculate *surface states*,^{34,37,38,42-49} often with the simplified form (4.16) or a multiple of it. In any particular problem one usually expects at the end to solve the same numerical equation whatever theory one uses to set it up, and that is in a sense true here between the Green function method and the more traditional one of matching individual wave functions.⁵⁰ However, the route is quite different because the Green functions G_A and G_B are expressed in terms of the bulk bands with real \mathbf{k} , whereas the wave function method uses the solutions of the Schrödinger equation in the band gap that have complex \mathbf{k} .⁵⁰ The Green function approach can, for example, resolve ambiguities about truncating the number of waves in the wave function method. It has also been extended to deal with three media, e.g., an oxide layer between semi-conductor and vacuum^{42,48} or just a transition region for the potential between solid and vacuum.^{38,44,47} Most of the above references use extensions of the nearly free-electron approach but one can also develop the formalism for tight-binding.⁴⁵

For establishing the electrical neutrality of the surface or interface, the dipole moments set up there, calculating the potential self-consistency, etc., $n(E, \mathbf{r})$ is clearly an important quantity, as is the total density of states $n(E)$ and counting the number of electrons $N(E)$ up to energy E . One can obtain closed formulae for $n(E)$ and $N(E)$:

$$n(E) = \pi^{-1} \operatorname{Im} \frac{d}{dE} \log \det G(E + iO), \quad (6.1)$$

$$N(E) = \pi^{-1} \operatorname{Im} [\log \det G(E + iO)]_{-\infty}^E. \quad (6.2)$$

The first of these formulas follows easily from Eq. (14.5) and then the second by integration. Note that $\operatorname{Im} \log z = \arg z$ and (6.2) can be interpreted as

⁴² F. Yndurain and J. Rubio, *Phys. Rev. Lett.* **26**, 138 (1971).

⁴³ M. Elices and F. Yndurain, *J. Phys. C* **5**, L146 (1972).

⁴⁴ F. Flores, E. Louis, and J. Rubio, *J. Phys. C* **5**, 3469 (1972).

⁴⁵ E. Louis and F. Yndurain, *Phys. Status Solidi B* **57**, 175 (1973).

⁴⁶ F. Garcia-Moliner and F. Flores, *J. Phys. C* **9**, 1609 (1976).

⁴⁷ E. Louis, F. Undurain, and F. Flores, *Phys. Rev. B* **13**, 4408 (1976).

⁴⁸ Y. Osaka, *J. Phys. Soc. Jpn.* **41**, 865 (1976).

⁴⁹ B. Velicky and J. Kudrnovsky, *Surf. Sci.* **64**, 411 (1977).

⁵⁰ See, for example, V. Heine, *Surf. Sci.* **2**, 1 (1964); R. O. Jones, *Phys. Rev. Lett.* **20**, 18 (1968).

follows: near an energy eigenvalue, $G(E)$ is dominated by the pole at the eigenvalue, and the argument (phase) of $G(E + iO)$ changes by π as one goes halfway around the pole by walking past it along a path just above the real axis. Garcia-Moliner and Rubio³⁸ have shown how these expressions may be split up into terms from the bulk bands on the two sides of the surface and the effects of the surface itself. Clearly, this is possible from (4.9) and the corresponding formula for the other side of the surface. It is interesting that in (6.1) and (6.2), one can use \mathcal{G} instead of G , where \mathcal{G} is the operator G on the boundary surface in the sense of Section II,4. In the notation of Eq. (3.1), if the boundary surface is the plane $z = 0$, then $\det \mathcal{G}$ has poles $a_n/(E - E_n)$ with strength a_n instead of unity, where a_n is given by

$$a_n = \iint |\psi_n(x, y, z = 0)|^2 dx dy. \quad (6.3)$$

It is not difficult to show that the a_n drop out of (6.1) because of the logarithm and differentiation, or alternatively because a_n is real and does not affect the argument of $\det \mathcal{G}$ near the pole.

In such a formulation of $n(E)$, $N(E)$, the surface contributions include, and treat on an equal footing,⁵¹ the distortions in the bulk bands near the surface, localized surface states, surface resonances, and the tails of the metal wave function (in a metal–semiconductor junction) that stretch a little way into the semiconductor for metal band states with energies in the semiconductor band gap.²⁹ The Green function formulation should therefore be useful for proving some systematic cancellations and invariances that have been shown to exist among these different contributions,^{46,52–55} and to some extent the method has been.^{46,47,51} However, most of the results have only been established in one dimension^{52,53,55} in some cases by numerical studies,⁵³ so that a Green function treatment that can be extended to three dimensions would be useful. For example, the total amount of charge in the “Friedel oscillations” of a *full* band (charge deficit) is quantized in units of *half* an electron (per spin),^{38,53} half of which appears to be associated with the top and bottom of the band, respectively.^{46,53} Similarly, the charge in the tails of the metal wave functions integrated across a whole band gap appears to amount to *half* an electron (per spin)³⁸ compared to one electron for the corresponding surface states on a vacuum–semiconductor surface. The reason for these quantizations is that $G(E)$ for a semiconductor is real for E

⁵¹ F. Yndurain, *J. Phys. C* **4**, 2849 (1971).

⁵² V. Heine, *Phys. Rev.* **145**, 593 (1966).

⁵³ J. A. Appelbaum and D. R. Hamann, *Phys. Rev. B* **10**, 4973 (1974).

⁵⁴ L. Kleinman, *Phys. Rev. B* **11**, 858 (1975).

⁵⁵ F. Claro, *Phys. Rev. B* **17**, 699 (1978).

in the band gap, with a cut for the energy band ending in a square root singularity at the band edge, which results in a particular phase change in the argument (Im log) of G there. This can be seen, for example, in a one-dimensional two-band model.³⁴ Moreover, the distribution in energy of the charge from the tails of the metal wave functions seems very insensitive to the assumed form of the potential in the transition region between metal and semiconductor.⁴⁷ Another cancellation occurs between the actual distribution with depth of the charge deficit (Friedel oscillations) from the surface in a full band,⁵³ and either a whole band or half-full band of surface states. The former case, when surface states are split off from a band, is an example of the Kohn–Majumdar theorem²⁴ (see Section I,2): the surface state is genuinely sucked out of the band, leaving a deficit of charge density in space equal and opposite to its own. The latter case of cancellation by a half-full band of surface states appears to obtain at a group IV semiconductor surface, its effect on the self-consistency of the potential being to push the surface states always toward midgap.⁵³ These quantization and cancellation effects are remarkable and certainly well established for one-dimensional models, sometimes the Kroenig–Penney model. It would be nice to have simpler analytic proofs for simple two-band models of band gaps, which can be extended to three dimensions and give greater insight into which cancellation occurs when (e.g., whole- or half-band of surface states).

We turn finally to another underdeveloped area of application: the theory of *molecules and band structures of solids*. Some simple calculations by Inglesfield have demonstrated the utility of the method.³² Consider a molecule like Li_2 of two identical atoms (Fig. 9). For the sake of simplicity let us here also make the “Wigner trick” assumption (8.3) about the potential (see Section III,8 for an extended discussion of this): the total potential including correlation hole felt by an electron when in the A (or B) half-space of

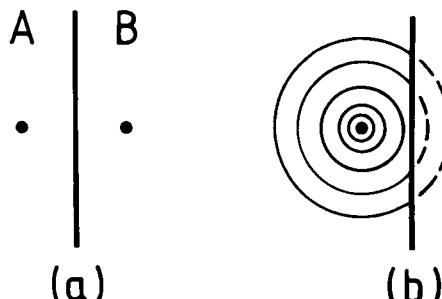


FIG. 9. (a) Diatomic molecule of two identical atoms, with dividing surface in the sense of Fig. 6. (b) Potential contours in region A (full lines), and their extension into region B (dotted lines) for the definition of G_A .

Fig. 9 is that of the Li^+ ion in that half-space, all other contributions cancelling. The potential V_A in (4.2) defining G_A is obtained by augmenting V from the left half-space (Fig. 9) in any way we like, to fill the whole space. The obvious way shown in Fig. 9b is by completing the potential in a spherically symmetric way so that V_A is simply the potential in the whole of space of a Li^+ ion. The corresponding G_A can be constructed from the atomic energy levels and wave functions (including the continuum at positive energy). Two approaches suggest themselves. Inglesfield³² constructed around the ion a sphere of suitably large radius, but not too large, on which he applied the boundary condition $\psi = 0$. The continuum becomes a series of discrete levels that can be calculated and summed. Since inside the sphere the Green function G_A is approximately invariant (Sections II,2 and II,5) under the change of boundary conditions from infinity to the large sphere, it will serve well enough for calculating the matching over the important central area of the matching plane. An alternative approach would lie in the spirit of Anderson⁵⁶ and the quantum defect method.⁵⁷ Since G_A will only be required on the matching plane outside the Li^+ core, it is uniquely determined from the free-atom energy levels by integrating inwards from infinity, in the case of Li^+ with the tabulated Coulomb wave functions, but that would not be quite correct for polyvalent elements. One can therefore envisage a semianalytic representation of G_A in terms of, say, the $l = 0$ and 1 quantum defects, perhaps together with their energy derivatives.

The importance of this approach would lie in the fact that one has a single entity G_A to represent an element in anything from a small molecule to a complex alloy.⁵⁸ For example, Inglesfield has calculated both the energy levels and cohesion of C_2 , and the band structure of the diamond crystal,³² where he obtains rather encouraging results. Similarly he considered the molecules LiNa ³² and Li_3 ³³ as well as Li_2 and Na_2 . One could, for example, combine the atomic G_A of Fig. 9 with the G_B of a bulk solid to obtain an atom adsorbed on a solid surface: such a calculation does not appear to have been made yet. It is not clear whether such methods would be superior to others for the last wrinkle in, say, the band structure of a solid: probably not. But what we lack is a fundamental quantum theory of the cohesion and structure of a wide diversity of solids. Here we integrate over all occupied states, often over filled bands, in which case the wrinkles in the band structure integrate out because they are higher Fourier components from distant atomic inter-

⁵⁶ P. W. Anderson, *Phys. Rev. Lett.* **20**, 413 (1968); H. S. Fricker and P. W. Anderson, *J. Chem. Phys.* **55**, 5028 (1971); H. S. Fricker, *ibid.* p. 5034.

⁵⁷ F. S. Ham, *Solid State Phys.* **1**, 127 (1955).

⁵⁸ For some applications one would also need to parameterize changes in G at the boundary due to changes in the self-consistent potential when one varies the bond length or atomic volume.

actions in the Wannier expansion of $E(\mathbf{k})$. Nor does an on-off numerical calculation advance our understanding very much. We need a theory to explain trends and differences. I do believe that a G_A representation of an atom, matched onto other atoms, is the correct basis for an “atoms in molecules” or “atoms in solids” type of theory.

III. LCAO: From under a Cloud to Out in the Sun

7. HISTORICAL PERSPECTIVE AND BASIC CLARIFICATIONS

The LCAO or tight-binding method^{59,60} dates from the beginnings of quantum solid state physics, from Bloch's historic paper⁶¹ in 1928. Yet only in the last few years has any substantial amount of *quantitative* calculation been done with it, as described in the next chapter.¹⁵ The purpose of this and the next two sections is to review the reasons for this gap, what the obstacles were, and how they have been overcome.

Certainly the LCAO method has been very important throughout its history as a conceptual picture of the electronic structure of solids.⁵⁹ By LCAO we mean, of course, expressing the wave function of an electron in terms of an appropriate set of atomic orbitals

$$\phi_\alpha = \phi_\alpha(\mathbf{r} - \mathbf{l}), \quad \psi = \sum_\alpha a_\alpha \phi_\alpha. \quad (7.1)$$

Note that we have here abbreviated the notation introduced in Section I,1: in this section the type of orbital α and the site l on which it is centered are subsumed into the single index α . The method is particularly appropriate when the overlap between neighboring atoms is relatively small, e.g., for describing the p bands of alkali halides and d bands of transition metals. The ϕ_α need not be strictly atomic orbitals: for example, in diamond or organic molecules ϕ_α might be a tetrahedrally oriented, directed valence bond orbital expressible in terms of the 2s and 2p carbon atomic orbitals. Linear combinations of such orbitals on neighboring atoms would then suffice to describe bonding and antibonding states between them. Alternatively if one only wished to represent the valence band, one might choose for ϕ_α a bonding orbital straddling two atoms, and ignore antibonding states.⁶²

⁵⁹ N. F. Mott and H. Jones, “The Theory of Metals and Alloys.” Oxford Univ. Press, London and New York, 1936.

⁶⁰ J. C. Slater, “Quantum Theory of Molecules and Solids.” Vol. 2, p. 203. McGraw-Hill, New York, 1965.

⁶¹ F. Bloch, *Z. Phys.* **52**, 555 (1928).

⁶² G. G. Hall, *Philos. Mag.* [8] **3**, 429 (1958).

The *Ansatz* (7.1) applied to the Schrödinger equation $\mathcal{H}\psi = E\psi$ leads to the secular determinants

$$\det |\mathcal{H}_{\alpha\beta} - ES_{\alpha\beta}| = 0 \quad (7.2)$$

for the energy values E , where $\mathcal{H}_{\alpha\beta}$ and $S_{\alpha\beta}$ are the Hamiltonian and overlap matrices

$$\mathcal{H}_{\alpha\beta} = \langle \alpha | \mathcal{H} | \beta \rangle, \quad S_{\alpha\beta} = \langle \alpha | \beta \rangle. \quad (7.3)$$

In general, orbitals on different sites are not mutually orthogonal, so that $S_{\alpha\beta} \neq 0$ for $\alpha \neq \beta$, although we assume the orbitals normalized, giving $S_{\alpha\alpha} = 1$. For an arbitrary molecule or amorphous cluster the size of the matrices \mathcal{H} and S equals the total number of orbitals, while in a periodic solid the use of Bloch's theorem allows simplification to a much smaller size (the number of orbitals per unit cell) with the $\mathcal{H}_{\alpha\beta}$ and $S_{\alpha\beta}$ in (7.2) becoming functions of \mathbf{k} to generate the band structure $E(\mathbf{k})$. We do not elaborate on the expression of $\mathcal{H}_{\alpha\beta}$ in terms of one-, two-, and three-center integrals, some of which have the significance of on-site energies, hopping integrals, and crystal field splittings.⁶³

Intermediate between a qualitative picture and a quantitative ab initio calculation lies the *semiempirical fitting, parametrization, or interpolation scheme*.⁶⁴ If some information is known from experiment or accurate computation about the energy levels of a molecule or the bands of a crystal, e.g., the energies at points of high symmetry in the Brillouin zone, then one can fit a few parameters in (7.2) and hence interpolate the remainder of the energy bands. The early history applying such a procedure was not entirely encouraging. For example, Hall⁶² developed a model of the valence bands of diamond, silicon, and germanium in terms of bonding orbitals. While the overall shape of the bands was reasonable, the effective masses at Γ at the top of the valence band came out of order unity instead of order 0.1 as observed. The reason is not hard to seek. The LCAO method gives $E(\mathbf{k})$ (or rather, each $\mathcal{H}_{\alpha\beta}$ and $S_{\alpha\beta}$) as a Fourier series⁵⁹ in \mathbf{k} , with successively higher terms corresponding to more and more distant neighbors. An effective mass of 0.1 corresponds to a very sharp curvature of $E(\mathbf{k})$ near Γ and hence to very distant interactions. However, a sharp curvature can be obtained in another way, through the interaction of two bands as in the nearly free-electron method. Thus Dresselhaus and Dresselhaus⁶⁵ achieved a good fit to the $E(\mathbf{k})$ of silicon and germanium with a model of atomic s and p orbitals describing both valence and conduction states. However, 13 parameters were

⁶³ J. Callaway, "Energy Band Theory," p. 103. Academic Press, New York, 1964.

⁶⁴ J. C. Slater and G. F. Koster, *Phys. Rev.* **94**, 1498 (1954).

⁶⁵ G. Dresselhaus and M. S. Dresselhaus, *Phys. Rev.* **160**, 649 (1967).

required, six for nearest neighbors and seven for next-nearest neighbors (plus spin-orbit coupling). By contrast, the same degree of fit can be obtained by three parameters with the pseudopotential approach.⁶⁶⁻⁶⁸ More successful are the Hückel⁶⁹ and PPP⁷⁰ schemes for organic molecules. The fitting scheme was also much more successful applied to the d bands of copper.^{71,72} Here four parameters were needed for five bands. (Further parameters described the sp band and its interaction with the d bands.) Yet this success also raised a theoretical embarrassment. The tight-binding method is intended to apply to "tightly bound" electrons, i.e., in atomic orbitals at negative energies with exponentially decreasing tails. But the copper d bands are at positive energy relative to the bottom of the nearly free-electron sp band, and at or above the "muffin-tin zero"⁸ in a band structure calculation. Thus a computer integrating outwards from an atomic center will produce a 3d orbital with an unnormalizable tail oscillating to infinity!

Put more generally, in using (7.2) one is fitting the matrix elements in the secular equation: nothing is known about the orbitals ϕ_α . In most of the fitting procedures cited above,^{62,65,71,72} the $S_{\alpha\beta}$ were set equal to $\epsilon_{\alpha\beta}$, implying orthogonalized orbitals. (Actually this implication does *not* follow, as we see below, but it was usually taken as implied at the time. However, see Hall.⁶²) Certainly there are procedures starting from an arbitrary set of orbitals and generating an orthogonal basis: it corresponds to multiplying (7.2) from the left and from the right by $S^{-1/2}$. However, such orbitals or Wannier functions⁸ usually have rather longer tails to give the orthogonality, thus necessitating rather unphysical $\mathcal{H}_{\alpha\beta}$ parameters to more distant neighbors. If, on the other hand, one decides to keep the $S_{\alpha\beta}$ as fitting parameters in (7.2), one has problems of nonuniqueness because the $-ES_{\alpha\beta}$ for $\alpha \neq \beta$ has sign opposite to the $\mathcal{H}_{\alpha\beta}$ and thus tends to cancel it: many pairs of $\mathcal{H}_{\alpha\beta}$, $S_{\alpha\beta}$ give rather similar energy bands. We are somewhat groping in a theoretical darkness.

There is one way out that we eschew: to take leave of physics and embrace pure computing. One can set up enough orbitals on each atomic site, or indeed located between atoms, to constitute effectively a complete basis set for any wave function likely to be relevant. Then one simply computes the

⁶⁶ M. L. Cohen and T. K. Bergstresser, *Phys. Rev.* **141**, 789 (1966).

⁶⁷ M. L. Cohen and V. Heine, *Solid State Phys.* **24**, 37 (1970).

⁶⁸ Not "chemical pseudopotentials" in the sense of Section III.9 and the next chapter.

⁶⁹ L. Salem, "The Molecular Orbital Theory of Conjugated Systems." Benjamin, New York, 1966.

⁷⁰ G. Klopman and R. C. Evans, in "Electronic Structure Calculations" (G. A. Segal, ed.), p. 29. Plenum, New York, 1977.

⁷¹ L. Hodges, H. Ehrenreich, and N. D. Lang, *Phys. Rev.* **152**, 505 (1966).

⁷² F. M. Mueller, *Phys. Rev.* **153**, 659 (1967).

energy levels from a given Hamiltonian. This will not work as a fitting procedure because there would be more parameters than fitting information, but there is nothing wrong with it in principle as an ab initio calculation for some specific system. Other band structure methods are generally considered faster, but that is a practical matter. However, a different philosophy underlies all that is contained in this volume (and in a companion volume⁷³). We wish to have parameters, whether fitted or calculated, that have "physical significance," since in this way "understanding" is built up. Happily or unhappily, the world of reality is always one jump or more ahead of the world of theoretical physics! We are always reaching out, tentatively adapting knowledge already gained to new situations where hopefully it will be appropriate. In other words we *want to transfer information*; we want to apply parameters from the C-C bond in diamond to similar bonds in molecules, interactions between neighboring nickel atoms in pure nickel to similar atoms found at a similar interatomic spacing in an alloy or amorphous metal system or metal cluster molecule. Such transfers will only be successful if we have previously cast our theoretical formalism into a form such that its building bricks, its parameters, are characteristic of the atom or bond in question. We are seeking a minimal basis set with a minimum number of parameters in (7.2) so that there remain no systematic numerical cancellation effects for instance. We are seeking a good description of electronic structure with atomic interactions having as short a range as possible. We require a formalism that works because we know it must, not one that may suffice on some occasions for unknown extraneous, coincidental reasons.

Clearly a deeper understanding is required to make further progress, and so we turn to the simplest systems to which LCAO has been applied, the hydrogen molecular ion H_2^+ and the hydrogen molecule H_2 . The fog has only lifted in recent years, mainly due to two separate contributions by Anderson and co-workers^{56,74} (see Section III,9). Work previous to that is well described by Slater,⁷⁵ and from it one might conclude that the LCAO technique is not likely ever to be very successful. We can discern three problem areas that we shall discuss in turn, namely, the large overlap between the atomic orbitals, the suitability of atomic orbitals as basis functions, and the treatment of electron correlation.

The first issue concerns the overlap integrals $S_{\alpha\beta}$ in (7.2). At the equilibrium interatomic spacing, we have $S_{12} = 0.6$ at $R = 2$ a.u. in H_2^+ and $S_{12} = 0.7$

⁷³ M. L. Cohen and V. Heine, *Solid State Phys.* **24**, 38, 39, and 228 (1970) (see also the remainder of that volume).

⁷⁴ P. W. Anderson, *Phys. Rev. Lett.* **21**, 13 (1968).

⁷⁵ J. C. Slater, "Quantum Theory of Molecules and Solids," Vol. 1. McGraw-Hill, New York, 1963.

at $R = 1.5$ a.u. in H_2 , both clearly very large, too large to be ignored. As already mentioned, one would appear to have the choice of retaining S_{12} and nonorthogonal orbitals resulting in large cancellation in (7.2) and indeterminacy in fitting parameters, or of orthogonalizing the orbitals at the expense of generating unphysical long-range tails, both choices being unsatisfactory for different reasons. However, there is a third way, used in the Anderson-Bullet scheme⁷⁴ of Section III,9 and the next two chapters.^{1,15} If we write

$$\mathcal{H}\phi_\beta = \sum_\alpha H_{\alpha\beta} \phi_\alpha, \quad (7.4)$$

where \mathcal{H} is the Hamiltonian operator, and substitute in (7.1), then the Schrödinger equation becomes

$$\sum_\alpha \left[\sum_\beta (H_{\alpha\beta} - E \delta_{\alpha\beta}) a_\beta \right] \phi_\alpha = 0. \quad (7.5)$$

This equation represents the expansion of the quantity zero on the right-hand side, in terms of a set of functions ϕ_α , which need not be orthogonal and which are not overcomplete. Hence each coefficient in (7.5) must be zero, giving

$$\sum_\beta (H_{\alpha\beta} - E \delta_{\alpha\beta}) a_\beta = 0, \quad \text{for all } \alpha. \quad (7.6)$$

As usual, the consistency condition is

$$\det |H_{\alpha\beta} - E \delta_{\alpha\beta}| = 0, \quad (7.7)$$

where we appear to have rederived (7.2) while dispensing with $S_{\alpha\beta}$ for $\alpha \neq \beta$! However, the matrix $[H_{\alpha\beta}]$ can easily be shown from (7.4) to be

$$[H] = [S^{-1}][\mathcal{H}], \quad (7.8)$$

so that (7.7) is equivalent to (7.2): we have simply multiplied (7.2) by S^{-1} from the left. The appearance of $\delta_{\alpha\beta}$ in place of $S_{\alpha\beta}$ is a real boon in (7.7), but a price has to be paid for all good things: the matrix $H_{\alpha\beta}$ is in general no longer Hermitian as is clear from (7.8). However, and this is an important point, $[H_{\alpha\beta}]$ is Hermitian between geometrically equivalent orbitals α, β . This can be verified easily from the case of H_2^+ or H_2 where α, β refer to the 1s orbital on the left and right proton respectively. Writing (7.8) in terms of 2×2 matrices, we obtain

$$\begin{bmatrix} aE_0 + bh & ah + bE_0 \\ ah + bE_0 & aE_0 + bh \end{bmatrix} = \begin{bmatrix} a & b \\ b & a \end{bmatrix} \begin{bmatrix} E_0 & h \\ h & E_0 \end{bmatrix} \quad (7.9)$$

and see that $[H]$ is symmetric like $[S^{-1}]$ and $[\mathcal{H}]$. The argument can be extended to the d-band case: although we have five orbitals on each atom and

hence apparently many $S_{\alpha\beta}$ between inequivalent orbitals on different sites, if we use the line of centers of the two atoms as axis of quantization, we find all $S_{\alpha\beta}$ and $\mathcal{H}_{\alpha\beta}$ are zero except for $dd\sigma$, $dd\pi$, and $dd\delta$ integrals between equivalent orbitals, and that is sufficient to establish the result. This therefore explains why no explicit S was required in the fitting scheme for the copper d bands, although the Hamiltonian matrix was short range and apparently Hermitian.^{71,72} The same applies to the Hückel scheme⁶⁹ for the π orbitals of benzene,⁷⁶ etc.

The second question encountered with the hydrogen molecule is whether the hydrogen atomic orbitals

$$\phi(\mathbf{r}) = (\text{const}) \exp(-\alpha r), \quad (7.10)$$

with $\alpha = 1$ in atomic units,²⁰ are really an adequate basis set for the bonding (or antibonding) molecular orbital. We return to this in Section III,9, but deal with one aspect here. The experience with H_2^+ and H_2 is that use of the straight hydrogen orbitals (i.e., with $\alpha = 1$) are not very good for the binding energy, but considerable improvement is obtained by varying α as a variational parameter. Certainly it is true that the bonding molecular orbital must fall away more steeply than the atomic orbital in a direction away from the other atom⁷⁷ (and the antibonding orbital less so). The reason is simply that its energy level is lower (higher) than the atomic one. If one uses LCAO wave functions with variable α , the wave function achieves this steepness in the only way it can, by taking α greater (less) than unity to minimize the energy, in fact, $\alpha = 1.25$ for the bonding orbital. Indeed, Ruedenberg⁷⁸ has made much play of this “contraction of the wave function” to describe the origin of the binding in H_2 . However, it is now known to be spurious. For H_2^+ the exact wave function is available and Pettifor⁷⁹ has expanded it in spherical harmonics about the proton position. The spherically symmetric part has the *same* exponential form ($\alpha = 1$) as the atomic orbital, the steepness of ψ in the nonbonding direction arising from the admixture of an $l = 1$ component that shifts the weight of ψ from the nonbonding direction to the opposite side, i.e., to the bond in the center of the molecule.⁸⁰ Such a lopsidedness in the localized orbitals would appear essential to a good calculation for H_2 and the alkali dimers Na_2 , etc., although even here we must not exaggerate its importance. Using straight hydrogen atom orbitals ($\alpha = 1$) one obtains 95% of the electronic energy for H_2^+ at equilibrium⁷⁵ ($R = 2.1$

⁷⁶ P. W. Anderson, *Phys. Rev.* **181**, 25 (1969).

⁷⁷ See Fig. 1-4 of Slater,⁷⁵ p. 7.

⁷⁸ K. Ruedenberg, *Rev. Mod. Phys.* **34**, 326 (1962).

⁷⁹ D. G. Pettifor, *J. Chem. Phys.* **59**, 4320 (1973).

⁸⁰ The importance of this lopsidedness in describing the bond at large atomic separation was already recognized by C. Herring, *Rev. Mod. Phys.* **34**, 631 (1962).

a.u.) although the error appears rather larger (50%) in the binding energy because most of the negative electronic energy is lost in the positive electrostatic energy of the two protons. In electrically neutral molecules and solids, the nuclear Coulomb repulsion is, of course, cancelled by the negative charge of the electron density.

In systems other than those just mentioned, the lopsidedness in the atomic-like orbitals ϕ_x is probably of less significance. In carbon, silicon, and germanium, for instance, Bullett^{15,81} uses directed valence orbitals that are already very lopsided toward the bond and rather small in the nonbonding direction, which is probably not very significant. In the π orbitals of benzene,⁷⁶ the $2p_z$ orbital binds to neighboring carbon atoms in both directions, and so no lopsidedness arises. In transition metals the situation is somewhat different. The hopping matrix element between two orbitals ϕ sited at O and I is

$$\int \phi^*(\mathbf{r}) V(\mathbf{r}) \phi(\mathbf{r} - \mathbf{l}) d^3\mathbf{r} \quad (7.11a)$$

$$= \frac{1}{2} \int \phi^*(\mathbf{r}) [V(\mathbf{r}) + V(|\mathbf{r} - \mathbf{l}|)] \phi(\mathbf{r} - \mathbf{l}) d^3\mathbf{r}, \quad (7.11b)$$

where V is the atomic potential. Here $\phi(\mathbf{r})$ and $\phi(\mathbf{r} - \mathbf{l})$ decrease exponentially in opposite directions toward each other, exactly so in hydrogen orbitals and roughly so in other cases. Thus the product $\phi^*(\mathbf{r})\phi(\mathbf{r} - \mathbf{l})$ may be taken as a constant along the line of centers, and the relative importance of different regions of space in (7.11b) depends on the magnitude of $V(\mathbf{r})$. In H_2 , the regions in the center of the bond and near the nuclei are both important, and the same applies to p bonds in the first row of the periodic table. For other sp-bonded atoms such as silicon it is best to think of V as a pseudopotential equal to the atomic potential outside the atomic core and cancelled off more or less to zero inside the core.⁸² Clearly, the region halfway between the atoms dominates the integral (7.11b) in accordance with the intuitive idea of the bond. However, the opposite situation is true in transition metal atoms, where V is very strong quite deep inside the atom: for the 3d series there is no pseudopotential cancellation while for the 4d series only the very inner 3d region is cancelled and similarly for 5d. A rather small tail from $\phi(\mathbf{r} - \mathbf{l})$ in (7.11a) overlaps a large $V(\mathbf{r})$ well inside the other atom, which explains the paradox that the interatomic bonding in transition metals (e.g., the bandwidth) can be related to an intra-atomic integral.^{84,85} We may sum-

⁸¹ D. W. Bullett, *J. Phys. C* **8**, 2695 (1975).

⁸² V. Heine, *Solid State Phys.* **24**, 20 (1970).

⁸³ V. Heine, *Solid State Phys.* **24**, 1 (1970).

⁸⁴ V. Heine, *Phys. Rev.* **153**, 673 (1967).

⁸⁵ J. W. Moriarty, *J. Phys. F* **5**, 873 (1975).

marize this discussion as follows. We have dismissed the “contraction of the wave function” [$\alpha > 1$ in (7.10)] as the origin of binding: straight atomic orbitals are better. Left open is the question of their lopsidedness, which in solid-state applications one might envisage as “lopsidedness in many directions at once,” i.e., an expansion of the orbital. We return to this in Section III.9.

Finally, in our review of the hydrogen molecule we encounter the long and thorny question of what potential our electron is supposed to be moving in, the question of *electron correlation*, which we take up next.

8. ELECTRON CORRELATION IN SOLIDS AND MOLECULES

The question of how to treat electron correlation and how to set up a potential for an electron in solid or molecule is a wide one arising in any ab initio approach to electronic structure. We discuss it here in relation to the LCAO method because localized orbitals form the basis of all the remainder of this volume and because we partially follow the theme of the previous section of inquiring what can be learned from the detailed study of H_2 and other small molecules. In fact, the traditional chemical literature⁷⁵ on H_2 is little help because chemists tend to limit themselves to Hartree–Fock, which is disastrous in solids,⁸⁶ and configurational interaction, which is inapplicable to a large system. All that we can learn from H_2 is that Hartree–Fock, which is equivalent to the simple molecular orbital approximation, is energetically not as good as the Heitler–London concept of only one electron occupying one atom at a time. (Yet we must retain the molecular orbital approach because we are on the “extended orbitals” side of the Mott transition as emphasized in Section I.1.) We must turn instead to the apparently naïve, disarmingly simple, but in reality much more sophisticated approximations of solid-state physics and the ideas of many-body theory, which are in fact equally applicable to molecules. Although not new,⁸⁸ they have been slow to be appreciated and we refer to a recent paper by Inglesfield⁹⁰ for the best exposition of them known to the author.

⁸⁶ In a metal the Fermi velocity becomes infinite, and the band width of sp-bonded materials considerably greater than observed. See Pines,⁸⁷ pp. 373 and 410.

⁸⁷ D. Pines, *Solid State Phys.* **1**, 367 (1955).

⁸⁸ Before the formalism was developed the main ideas were already sharp as I learned them from Mott in the mid 1950s and applied them in my thesis work.⁸⁹

⁸⁹ V. Heine, *Proc. R. Soc. London, Ser. A* **240**, 361 (1957).

⁹⁰ Section 2 of Inglesfield.⁹¹ It is evident that the approximation (8) is not part of the argument leading from (4) to (9), (10), and (12).

We content ourselves here with the merest summary and main conclusions of these ideas. Adding an electron to a system, or subtracting one, is a real physical process. To study the electrons already constituting the molecule or solid we sample them by removal, but in the following it is easier to use the language of adding an electron. Once in the system it creates around itself an exchange and correlation hole, a region more or less denuded of other electrons by the Coulomb repulsion and antisymmetry of the total wave function. This whole entity, the Landau quasi-particle, moves through the system with a certain Bloch \mathbf{k} vector in a perfect solid and more generally in an extended (molecular) orbital. Since the object has a size comparable to that of an atom, we cannot really describe it by a one-particle wave function $\psi(\mathbf{r})$. However, we can define where it is in terms of the position \mathbf{r} of the electron at its center. Near the Fermi level it is very stable, having a lifetime that tends to infinity at E_F even in a metal, and this applies equally to the hole quasi-particle states below E_F , so that it also has a well-defined quasi-particle energy E_n . Further from the Fermi level the quasi-particle tends to decay by Auger processes, represented by an imaginary part in E_n , now a complex number. Typically in aluminum the imaginary energy at the bottom of the band (the worst region) is of order 10% of the band width and is often ignored, while in some transition metals, e.g., nickel, it apparently becomes rather larger. *The complex Landau quasi-particle energy is exactly given by the one-electron-like Schrödinger-like equation (in atomic units)*²⁰

$$[-\frac{1}{2}\nabla^2 + V_N(\mathbf{r}) + V_H(\mathbf{r}) + V_{xc}]\psi_n(\mathbf{r}) = E_n \psi_n(\mathbf{r}). \quad (8.1)$$

Here V_N is the potential from the nuclei and V_H the Hartree-like potential defined by the total actual electron density $\rho(\mathbf{r})$. The exchange and correlation potential V_{xc} , also known as the self-energy Σ , is a complicated integral operator. Slater in his celebrated paper⁹² had already written the Hartree-Fock equations in this form, and shown in that case that V_{xc} is minus the electrostatic potential from the denuded exchange hole charge density ρ_{xc} , which has to be subtracted from ρ and V_H . Equations (5)–(7) of Inglesfield⁹¹ give an analogous but more complicated interpretation in the full theory. The E_n represents the excitation energy of the system for adding or subtracting one electron: multiple quasi-particle states such as an optical excitation from a “full” to an “empty” level can also be considered but become a bit more complicated if one includes their mutual “exciton” interaction. The ψ_n are as close as one can get to one-particle wave functions to describe complicated objects, and the electron density ρ and total ground state energy U can be related to them, though *not simply* by $\rho = \sum_n |\psi_n|^2$.

⁹¹ J. E. Inglesfield, *Mol. Phys.* **37**, 889 (1979).

⁹² J. C. Slater, *Phys. Rev.* **81**, 385 (1951).

All this represents a rigorous and exact basis for subsequent approximation to the V_{xc} operator. There is a world of difference between numerical approximation to a rigorously precisely defined quantity on the one hand, and lurching from one ill-defined concept to the next invalid equation on the other hand! Two possible directions present themselves. Many-body calculations have given a good picture of V_{xc} for a free-electron gas,⁹³ and for a nearly free-electron system one can modulate it to take⁸⁹

$$V_{xc} \simeq V_{xc}^{\text{free electron}}[\rho(\mathbf{r})] \quad (8.2)$$

in the spirit of Slater's⁹² $\rho^{1/3}$ approximation. In practice it becomes similar to the density functional formalism⁹⁴ (Section V,15), though different in concept.⁹¹

The second approach is the *Wigner trick*, i.e., the *Wigner–Seitz approximation to the exchange–correlation hole*.⁹⁵ In a system of monovalent atoms, if one electron is in a given atomic cell l then it is assumed that other electrons will not be. We write

$$\rho_{xc}(\mathbf{r}; \mathbf{r}') = \begin{cases} \rho(\mathbf{r}), & \text{in cell } l, \\ 0, & \text{elsewhere.} \end{cases} \quad (8.3a)$$

Here $\rho_{xc}(\mathbf{r}; \mathbf{r}')$ represents the exchange hole denudation charge density at \mathbf{r} when the electron at the center of the quasi-particle is at \mathbf{r}' . At the same time one usually assumes

$$\begin{aligned} &\text{the contribution to } V_N(\mathbf{r}) + V_H(\mathbf{r}) \text{ from cell } l \text{ is} \\ &\text{zero when } \mathbf{r} \text{ is in cell } l \neq l'. \end{aligned} \quad (8.3b)$$

The reason is that the nucleus plus ρ in cell l' constitute an electrically neutral system with an approximately spherical charge distribution giving zero potential outside itself. Approximations (8.3a) and (8.3b) have some deficiencies already recognized and to some extent corrected for by Wigner and Seitz, but it is important to see them as respectable, well-defined approximations to a rigorous and complete many-body theory. The latter, as already remarked, is equally applicable to molecules as to solids, and Anderson and Fricker⁵⁶ were the first to use (8.3) in this spirit for the alkali dimer molecules Li₂, Na₂, LiNa, K₂. The important point is that the one-particle-like Eq. (8.1) was solved to a high accuracy, without any LCAO approximation. In particular, the lopsidedness of ψ was taken into account explicitly. The same is true of subsequent work by Inglesfield^{32,33} using a different method to solve (8.1), the matching Green functions method of Section II,4. Inglesfield

⁹³ L. Hedin and S. Lundquist, *Solid State Phys.* **23**, 1 (1969); see Fig. 19, p. 89.

⁹⁴ N. D. Lang, *Solid State Phys.* **28**, 225 (1973).

⁹⁵ F. Seitz, “Modern Theory of Solids.” McGraw-Hill, New York, 1940.

was able to go further and solve the correct many-body equations for V_{xc} in an LCAO basis to establish a theoretical criterion for the validity of the Wigner–Seitz and Hartree–Fock forms.⁹¹ The conclusion is that for the alkali dimers the Wigner–Seitz approximation of course exaggerates the electron correlation (it represents complete correlation), but is nearer the truth than Hartree–Fock. Binding energies (themselves very small) are obtained correct to about 30%, equilibrium interatomic spacings to 3%, and vibration frequencies to about 20%. For H₂ Inglesfield⁹¹ finds the Hartree–Fock form for V_{xc} a much better approximation than Wigner–Seitz at the equilibrium spacing. This does not conflict with the earlier statement that the Heitler–London wave function (HL) is better than the simple molecular orbital one (MO). Because the S overlap integral is so large ($S = 0.72$) the two wave functions in fact are almost the same,⁹⁶ and Eq. (6.3a) represents rather stronger correlation than Heitler–London.

We conclude the Wigner trick has emerged from this examination with quite good marks. It is, of course, not an exact treatment of correlation and one can improve on it appreciably with the use of the best many-body calculations of $V_{xc}(\rho)$ and Eq. (8.2), but it is better than Hartree–Fock and simpler, and certainly it does not “throw out the baby with the bathwater.” The stage is therefore set for its extension to polyvalent elements.⁸¹ The idea is that the exchange–correlation hole is always confined to the atomic cell in which the moving electron finds itself so that (8.3b) applies as before. Inside the atomic cell the exchange correlation hole is assumed to have the same form as in a free atom. We conclude for polyvalent atoms

$$V_N + V_H + V_{xc} \text{ in that atomic cell} \quad (8.3c)$$

is the same as for the free atom.

(Of course, ψ satisfies different boundary conditions, wherein lies the bonding.) All this has nothing to do with the LCAO method: one has in any case to set up the potential in a solid; however, one chooses to solve the wave equation. But our conclusion shows that *the Wigner trick is peculiarly well suited to joining with the LCAO method*.

Our discussion so far has concerned itself with the quasi-particle energy levels E_n and the exchange correlation hole potential V_{xc} in (8.1). What of other quantities such as the charge density and total energy? What significance do the wave functions ψ_n in (8.1) have?

We have said earlier that the total electron density for N electrons is *not* given by

$$\rho(r) = \sum_n |\psi_n(\mathbf{r})|^2. \quad (8.4)$$

⁹⁶ The reader may verify that the overlap $\langle MO | HL \rangle$ is $(1 + S)[2(1 + S^2)]^{-1/2} = 0.98!$

The correct formula is⁹⁷

$$\rho(\mathbf{r}) = \sum_n c_n |\psi_n(\mathbf{r})|^2, \quad (8.5)$$

where the weighting factors c_n satisfy the sum rule

$$\sum c_n = N. \quad (8.6)$$

The summations in (8.4)–(8.6) are over all occupied states, i.e., over all N possible quasi-particle hole excitations. From (8.6) since the c_n average to unity, it is presumably not a bad approximation to take them as unity:

$$c_n \approx 1, \quad \text{for all } n. \quad (8.7)$$

Powerful support for this approximation comes from many-body calculations on the free-electron gas, at least for sp-bonded systems.⁹⁸ V_{xc} depends in principle on the energy E of the state being calculated in (8.1), but in practice for an electron gas of density found in most solids V_{xc} is remarkably constant (to within 10%) throughout the filled portion of the band⁹³ (and for some way above it). Now⁹¹

$$c_n^{-1} - 1 \propto \partial V_{xc}/\partial E, \quad (8.8)$$

so that if V_{xc} is taken as constant we recover (8.7) precisely. In any case because of (8.6) the errors of (8.7) must largely cancel. Similarly with the same approximation (8.7) the total energy of the system becomes⁹¹

$$U = U_{NN} + \sum_n E_n - \frac{1}{2} \int V_N(\mathbf{r}) \rho(\mathbf{r}) d^3\mathbf{r} \\ - \frac{1}{2} \iint \sum_n \psi_n^*(\mathbf{r}) V_{xc}(\mathbf{r}, \mathbf{r}', E_n) \psi_n(\mathbf{r}') d^3\mathbf{r} d^3\mathbf{r}', \quad (8.9)$$

where U_{NN} is the nuclear–nuclear repulsion, and as usual the $\sum E_n$ counts the electrostatic energy twice (both V_H and V_{xc}) so that it gets subtracted once by the subsequent terms. *We conclude that the simple independent-electron type of formulas (8.4) and (8.9) are very good approximations to the exact many-body results, at least for sp-bonded systems,* the reason lying in the sum rule (8.6) and the fact that V_{xc} in (8.8) is nearly independent of E for the electron densities usually encountered. Incidentally with approximation (8.7) our equations become more or less identical with the density functional theory⁹⁴ (Section V,15).

⁹⁷ J. E. Inglesfield,⁹¹ Eqs. (43) and (44).

⁹⁸ For a discussion of how close sp-bonded systems are to the free-electron gas, see Heine and Weaire³ and Cohen and Heine,⁶⁷ particularly pp. 252–253.

Finally with the Wigner trick one can simplify the total energy (8.9) even more, as a further step of approximation. As in the original Wigner and Seitz investigation,⁹⁵ the V_{NN} cancels the intercell part of $V_H\rho$ from (8.3b), whereas V_{xc} cancels the intracell part of $V_H\rho$ for a monovalent atom from (8.3a), so that U simplifies to

$$U \simeq \sum_n E_n. \quad (8.10a)$$

The corresponding result for polyvalent atoms is

$$U \simeq \sum_n E_n - \sum_{\text{atoms}} \left\{ \frac{1}{2} \iint_{\substack{\text{atomic} \\ \text{cell}}} \rho(\mathbf{r}) \frac{e^2}{|\mathbf{r} - \mathbf{r}'|} [\rho(\mathbf{r}') - \rho_{xc}(\mathbf{r}')] d^3\mathbf{r} d^3\mathbf{r}' \right\}, \quad (8.10b)$$

where the term in curly brackets is purely intra-atomic and hence cancels to a good approximation if one is calculating energy differences, e.g., between molecular O₂ and two adsorbed oxygen atoms, or between two crystallographic forms of an alloy. The approximation is better for close-packed solids where atomic cells are nearly spherical, than for molecules where that are not, especially at small interatomic separations. Even for close-packed solids, it may be inadequate to calculate changes in total energy with atomic displacements, such as are required in the theory of phonon frequencies. Then one must go back to (8.9), or even further to using the full many-body theory in an approximation⁹⁴ analogous to (8.2). We therefore return in Sections IV,13 and V,16 to a discussion about improving the approximation (8.10) in relation to (8.9), and also refer the reader to Bullett.⁹⁹ See also Inglesfield.¹⁰¹ However, much useful work⁹⁹ can be done even with (8.10) by exploiting various cancellations (Sections IV,13 and V,16), e.g., in comparing heats of adsorption of an atom at different sites on a solid surface.

Such then are the ideas upon which Anderson^{56,74,76} and Bullett⁹⁹ have built in order to deal with electron correlation. As we warned at the beginning, the simplicity of (8.1) and (8.9) on the one hand, and (8.3), and (8.10) as a further approximation, belie their sophistication.

9. THE MODERN ANDERSON–BULLETT SCHEME FOR THE TIGHT-BINDING METHOD

In this section we outline the scheme initiated by Anderson^{74,76} and developed over a period of years by Bullett^{81,100,102} to put the *tight-binding (LCAO) method onto a proper quantitative footing*.

⁹⁹ See Table I, Bullett,¹⁰⁰ and other references in Section 7; also Chapter II of this volume.

¹⁰⁰ D. W. Bullett, *J. Phys. C* **8**, 2707 (1975).

¹⁰¹ See Inglesfield⁹¹ and the end of Section 1 of Inglesfield.³²

¹⁰² For a complete bibliography see Bullett's chapter.

We wish to solve the Schrödinger equation

$$\mathcal{H}|\psi\rangle \equiv \left(T + \sum_{\alpha} V_{\alpha} \right) |\psi\rangle = E|\psi\rangle, \quad (9.1)$$

where T is the kinetic²⁰ energy $-\frac{1}{2}\nabla^2$ and it is convenient to use the Dirac notation. Let us consider for simplicity an s band so that we are concerned only with a single orbital per site. Thus in accordance with the abbreviated notation of Section III,7, the α and β are purely site indices here. The V_{α} are local atomiclike potentials whose further specification need not concern us here: there is no restriction against their overlapping.

The local orbitals ϕ_{α} , which we shall write as $|\alpha\rangle$, are *defined* by Anderson's equation

$$\left[T + V_{\alpha} + \sum'_{\beta} (V_{\beta} - |\beta\rangle\langle\beta| V_{\beta}) \right] |\alpha\rangle = E_{\alpha} |\alpha\rangle. \quad (9.2)$$

Let us suppose we have solved this set of equations. We can then use the $|\alpha\rangle$ as a basis

$$|\psi\rangle = \sum_{\beta} a_{\beta} |\beta\rangle \quad (9.3)$$

for attempting to solve (9.1). We substitute in (9.1) and simplify the left-hand side using the following rearrangement and relabeling of (9.2):

$$\left(T + \sum_{\alpha} V_{\alpha} \right) |\beta\rangle = E_{\beta} |\beta\rangle + \sum'_{\alpha} \langle \alpha | V_{\alpha} | \beta \rangle |\alpha\rangle, \quad (9.4)$$

where \sum' as in (9.2) excludes $\alpha = \beta$. We obtain

$$\sum_{\alpha} \left[\sum_{\beta} (E_{\beta} - E) \delta_{\alpha\beta} a_{\beta} + \sum'_{\alpha} \langle \alpha | V_{\alpha} | \beta \rangle a_{\beta} \right] |\alpha\rangle = 0. \quad (9.5)$$

The left-hand side of (9.5) will be identically zero if the coefficient of each $|\alpha\rangle$ is zero, i.e., if

$$\sum_{\beta} [(E_{\beta} - E) \delta_{\alpha\beta} + \langle \alpha | V_{\alpha} | \beta \rangle (1 - \delta_{\alpha\beta})] a_{\beta} = 0, \quad \text{for all } \alpha, \quad (9.6)$$

which leads as usual to a secular determinant

$$\det |(E_{\beta} - E) \delta_{\alpha\beta} + \langle \alpha | V_{\alpha} | \beta \rangle (1 - \delta_{\alpha\beta})| = 0, \quad (9.7)$$

where the first term gives the diagonal elements and the second term defines the off-diagonal elements (only).

Several comments on what we have done are now in order. It had been recognized for some time¹⁰³ that part of the problem with the LCAO method was the uncertainty about the selected localized orbitals being a good basis for the molecular orbitals or Bloch functions one wished to calculate. In the previous use of the LCAO method, one took what atomic orbitals were at hand and appealed to the variational principle to bolster the hope that they were adequate. By contrast we have in (9.2) constructed orbitals ϕ_α such that (9.5) and hence the Schrödinger equation (9.1) are satisfied exactly.

This is not quite as magical as it sounds. Suppose one had a band of Bloch functions. From it one could construct a set of localized Wannier⁸ functions, and if one could solve the equation for the Wannier function one would have an exact basis for the whole band. We have done a similar thing here, and the analogy illustrates another point. A Wannier function is a sum over all Bloch functions (in one Brillouin zone), and the latter like all wave functions are undetermined as regards an arbitrary multiplicative phase factor $\exp(i\gamma_k)$. By choosing different γ_k one obtains an infinite variety of bona fide Wannier functions. If one relaxes the condition that the ϕ_α be orthogonal one has an even richer choice of exact local basis orbitals, including the set defined by (9.2).

In fact the ϕ_α defined by (9.2) are the best possible local orbitals in the sense of being the most localized (see Section III,7), or very nearly so.⁷⁴ This follows from the form of the operator,

$$V_\beta = |\beta\rangle\langle\beta| V_\beta. \quad (9.8)$$

If instead of a single orbital $|\beta\rangle$ on site β we had a complete set of orbitals $|\beta_n\rangle$ there, then $\sum_n |\beta_n\rangle\langle\beta_n|$ would be the identity operator and

$$V_\beta = \sum_n |\beta_n\rangle\langle\beta_n| V_\beta \quad (9.9)$$

would be identically zero. This argument is sufficiently important that we repeat it in less sophisticated wave function form. If ψ is any wave function, then (9.8) on ψ becomes

$$V_\beta \psi = \left(\int \phi_\beta^* V_\beta \psi \, d^3\mathbf{r} \right) \phi_\beta. \quad (9.10)$$

Also

$$V_\beta \psi = \sum_n \left(\int \phi_{\beta n}^* V_\beta \psi \, d^3\mathbf{r} \right) \phi_{\beta n} \quad (9.11)$$

¹⁰³ W. H. Adams, *J. Chem. Phys.* **34**, 89 (1961); **37**, 2009 (1962); T. L. Gilbert, in "Molecular Orbitals in Chemistry, Physics and Biology" (P.-O. Löwdin and B. Pullman, eds.). Academic Press, New York, 1964.

is an expansion of $V_\beta \psi$ in terms of the complete set $\phi_{\beta n}$. In our case in (9.8) we have only one orbital ϕ_β , but we can imagine it augmented to a complete set. Since ϕ_β covers a similar region of space to V_β , the first term in (9.9), (9.11) is the dominant one and we have a high degree of cancellation in (9.8) and (9.10). This argument is familiar in pseudopotential theory,⁸² which accounts for the name "chemical pseudopotentials" given by Anderson to his method. Note, however, that in traditional pseudopotential theory the core orbitals are used for cancellation, whereas here the valence orbital(s) ϕ_β are used to cancel off most of the potential V_β on a neighboring atom β to the site α considered in (9.2). Thus the orbital ϕ_α "sees" the whole of the potential V_α more or less as in a free atom, plus some weak canceled remnants (9.8) on neighboring atoms. Actually one can go further: ϕ_β can be a pseudo-wave function in the traditional sense⁸³ and the core orbitals on atom β used to affect further cancellation, but we shall defer this and other wrinkles to the next chapter.¹⁵ The degree of cancellation is shown in Fig. 10 comparing $V_\beta |\alpha\rangle$ with

$$V_{ps,\beta} |\alpha\rangle = (V_\beta - |\beta\rangle \langle \beta| V_\beta) |\alpha\rangle. \quad (9.12a)$$

The high degree of cancellation is what makes the solution of the orbital equations (9.2) practicable. Since (9.2) contains the desired orbitals in the operator on the left-hand side, in principle the equation(s) must be solved by successive iteration until self-consistency is achieved. In practice the cancelled remnants (9.12a) are so small that (9.2) is almost the same as for

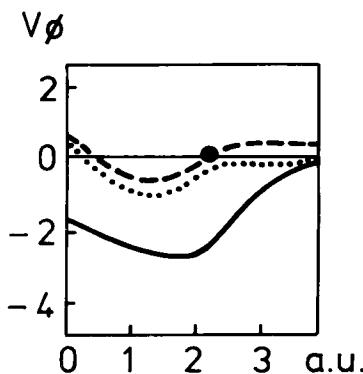


FIG. 10. Comparison of $V_\beta |\alpha\rangle$ (full line) with $V_{ps,\beta} |\alpha\rangle$ Eq. (7.12) (broken and dotted lines) for silicon in arbitrary units along two lines parallel to the bond direction from atom α to atom β and at a distance 1.0 a.u. from the line of centers. These lines have been chosen to be representative of a larger region of space than the line of centers itself. The origin is the point closest to the center of the bond and the spot closest to the center of atom β . The dotted curve refers to a line intersecting the bond from β to a further neighbor, whereas the broken curve refers to a line on the opposite side of β (after Bullett⁸¹).

an atomiclike orbital in the single spherical atomiclike potential V_α . One therefore uses the solution to the potential V_α , calculates E_α in (9.2) by first-order perturbation theory, and ignores iterating. Anderson⁷⁶ calculated the corrections to this truncated procedure and found them small for his first application of the method to the π orbitals of benzene. With the Wigner trick for the exchange correlation hole (Section III,8), V_α is equal to or close to the free-atom potential so that a further simplification is afforded by taking tabulated¹⁰⁴ atomic orbitals. Actually "the proof of the pudding is in the eating" and Bullett's considerable measure of success¹⁰² speaks for itself. The numerical accuracy in the form of the theory used at present is not comparable to APW and other standard band structure calculations on simple materials: the method is not designed for that although it can be quick and useful. The band structure of selenium¹⁰⁵ in fact compares well with the results of other methods, while the Chevrel phases¹⁰⁶ discussed by Bullett are difficult by other techniques because of the large number (14 or 15) of atoms per unit cell. The greatest uncertainty lies in the potential for situations where one most wants to use the method, in defects, at surfaces, and in amorphous or complex partially covalent, partially ionic materials. But let us emphasize that once have set up the potential we have in (9.2) a precise set of equations giving the exact solution to our problem, which can be iterated if required to any desired accuracy.¹⁰⁷ For example, in Section III,7, below Eq. (7.10), we discussed the need for orbitals, in some cases, that are a bit lopsided in the direction of a bond: Anderson¹⁰⁸ believes that such lopsidedness will be well given by the same perturbation procedure that he used⁷⁶ to investigate the adequacy of the simple atomic orbitals in benzene.

Our equations can be transformed to a *Hamiltonian-like form*. We define the matrix $H_{\alpha\beta}$ by

$$H_{\alpha\alpha} = E_\alpha, \quad H_{\alpha\beta} = \langle \alpha | V_\alpha | \beta \rangle \quad \text{for } \beta \neq \alpha. \quad (9.12b)$$

Then Eqs. (9.4), (9.6), and (9.7) become

$$\mathcal{H} | \beta \rangle = H_{\alpha\beta} | \alpha \rangle, \quad \alpha \text{ summed}, \quad (9.13a)$$

$$[H][a] = E[a], \quad (9.13b)$$

$$\det |H_{\alpha\beta} - E \delta_{\alpha\beta}| = 0, \quad (9.14)$$

¹⁰⁴ F. Herman and S. Skillman. "Atomic Structure Calculations." Prentice-Hall, Englewood Cliffs, New Jersey, 1963.

¹⁰⁵ D. W. Bullett, *Philos. Mag. [8]* **32**, 1063 (1975).

¹⁰⁶ D. W. Bullett, *Phys. Rev. Lett.* **39**, 664 (1977).

¹⁰⁷ See third sentence above Eq. (6.2)!

¹⁰⁸ P. W. Anderson, private correspondence.

where \mathcal{H} is given by (9.1) and $[a]$ is the column matrix of the a_α s, which we see are identical to Eqs. (7.4), (7.6), and (7.7). Thus we refer the reader to all the discussion in Section III,7 to the effect that $[H]$ is not the Hamiltonian matrix $[\mathcal{H}]$ in the conventional sense (7.3) but the non-Hermitian matrix $[S^{-1}][\mathcal{H}]$ because our basis functions are nonorthogonal. From (9.12b) $H_{\alpha\beta}$ represents orbital β overlapping onto the orbital and potential at site α , which clearly is different from the overlap $H_{\beta\alpha}$ from α onto the potential at β if the two are different atoms. However, $[H]$ becomes Hermitian when the ϕ_α are geometrically equivalent orbitals on different sites or sets of symmetry-related orbitals such as in d bands. The calculation of the local density of states or of band structures, etc., in the latter case proceeds in the same way as if with an orthonormal basis.¹⁰⁹ We assume that in the remainder of this chapter, leaving Haydock¹⁰⁹ to specify the calculational procedures for the general case, which are only slightly more involved. An approximate way may be to make H Hermitian by replacing $H_{\alpha\beta}$ and $H_{\beta\alpha}$ by

$$H'_{\alpha\beta} = H'_{\beta\alpha} = (H_{\alpha\beta} H_{\beta\alpha})^{1/2}, \quad (9.15)$$

which becomes exact in the case of a 2×2 secular equation (if $H_{\alpha\beta}$ and $H_{\beta\alpha}$ have the same sign!), but its general usefulness has not been tested so far. It can be convenient, for example, in an LCAO fitting procedure, to use a mixed representation intermediate between $[\mathcal{H}]$ and $[H]$. For example, Mattheiss¹¹⁰ fitted his TiO, etc., bandstructures in terms of titanium 3d orbitals and 2s2p on oxygen. There were no $S_{\alpha\beta}$ within the 5×5 d block or the 3×3 p block, but $S_{sd\sigma}$, $S_{pd\sigma}$, $S_{pd\pi}$ parameters were used in the off-diagonal elements between the d states and the s and p states, respectively. In this way the secular equation was kept Hermitian as well as the number of parameters small.

It is instructive to apply the method to the H_2 and He_2 molecules. The hydrogenic orbitals on the two atoms we designate $|1\rangle$, $|2\rangle$ and the eigenvalue $E_1 = E_2$ of (9.2) we term E_0 . Also $S = \langle 2|1\rangle$ and $h = \langle 2|V_2|1\rangle$. The $[H]$ matrix (9.12b) becomes

$$\begin{bmatrix} E_0 & h \\ h & E_0 \end{bmatrix} \quad (9.15)$$

with bonding/antibonding eigenvalues

$$E_\pm = E_0 + h. \quad (9.16)$$

Note that E_0 is not the “atomic” eigenvalue

$$E_a = \langle 1|T + V_1|1\rangle. \quad (9.17)$$

¹⁰⁹ R. Haydock, Chapter 3 of this volume, Sections 3 and 5.

¹¹⁰ L. F. Mattheiss, *Phys. Rev. B* **5**, 290 (1972).

We evaluate E_0 by taking the matrix element of

$$(T + V_1 + V_2 - |2\rangle\langle 2|V_2)|1\rangle = E_0|1\rangle, \quad (9.18)$$

which is Eq. (9.2), with $|1\rangle$. This gives

$$E_0 = E_a + \langle 1|V_2|1\rangle - Sh \quad (9.19)$$

$$E_{\pm} = E_a + \langle 1|V_2|1\rangle - Sh \pm h. \quad (9.20)$$

The reader may compare this with taking the expectation value of \mathcal{H} with respect to $|1\rangle \pm |2\rangle$, which gives a slightly different result. For E_+ we obtain the difference

$$\frac{S}{1+S}(Sh - \langle 1|V_2|1\rangle). \quad (9.21)$$

This is of order S^2 and the two terms have opposite sign, and so it will be very small. Using the Wigner trick (8.10) we have the total energies

$$U(H_2) = 2(E_a + \langle 1|V_2|1\rangle - sh + h), \quad (9.22)$$

$$U(He_2) = 4(E_a + \langle 1|V_2|1\rangle - Sh), \quad (9.23)$$

where, of course, similar symbols have different numerical values in the two cases H_2 and He_2 .

The above results on H_2 and He_2 illustrate some points of general significance. Although the secular equation $|H - E|$ (9.14) does not contain the overlap S explicitly, it is contained implicitly in E_0 (9.19), which (as already noted) is not the atomic energy E_a . Thus the bonding and antibonding energies (9.20) display their unsymmetrical values relative to E_a , which is a characteristic feature of the $|\mathcal{H} - ES|$ secular equation (7.2). In particular the term $-Sh$ in (9.23) gives the “closed-shell repulsion” since h is negative. The electrostatic overlap $\langle 1|V_2|1\rangle$ is always attractive and the nuclear-nuclear repulsion has disappeared in the Wigner trick (8.10), which is not a good approximation for a molecule when the nuclei approach close together as noted in Section III,8. However, the point we wish to make here is that the *closed-shell repulsion* associated with the overlap of Hilbert space between $|1\rangle$ and $|2\rangle$ is contained in the formalism. We see incidentally that it is also present in (9.22) for H_2 , although masked by h .

In He_2 we can obtain the sum of the bonding and antibonding energies purely from the diagonal matrix elements E_0 . Because of the invariance property of the trace of a matrix this trick can be applied to any full-band situation. For example, in silicon one can take as ϕ_α the bonding orbital between a pair of atoms and generate the complete valence band from such an orbital at every bond.⁸¹ Using the trace invariance Bullett¹⁰⁰ calculated

TABLE I. BOND LENGTHS (r_0) AND BULK MODULI (B) FOR CARBON AND SILICON^a

	Diamond	Graphite	Silicon	"Siligraphite" ^b
$r_0/\text{a.u.}$ (Wigner approx.)	2.73	2.34	4.13	3.8
$r_0/\text{a.u.}$ (corrected)	2.94	2.73	4.44	4.31
$r_0/\text{a.u.}$ (observed)	2.92	2.69	4.44	—
$B/10^{11} \text{ Nm}^{-2}$ (corrected)	4.5	—	1.0	—
$B/10^{11} \text{ Nm}^{-2}$ (observed)	4.43	—	0.98	—

^a All data from D. W. Bullett, *J. Phys. C* **8**, 2707 (1975).

^b Silicon in the graphite structure.

the interatomic interaction energy for carbon and silicon, first with the Wigner trick (8.10) and then corrected for the intercellular electrostatic energy in (8.9). The results given in Table I are very gratifying. With the same trace invariance a stacking fault energy of 62 mJm^{-2} for silicon was calculated,¹¹¹ compared with the experimental estimate¹¹² of $51 \pm 5 \text{ mJm}^{-2}$.

We have so far been rather vague about the precise definition of the V_α . It is clearly incorrect in (9.1) to add up the full atomic potentials: for hydrogen the two Coulomb potentials would give us $V(\mathbf{r})$ for H_2^+ , but not for H_2 . There is a bewildering degree of choice in the precise procedure to be adopted. First, we must decide how to set up the one-electron potential $V(\mathbf{r})$ in the whole system, including exchange and correlation hole, e.g., whether to use the Wigner trick (8.3). Second, a given $V(\mathbf{r})$ can be cut up into atomiclike contributions in an infinite number of ways if these are allowed to overlap. Third, in the cancellation term in (9.2) one can replace V_β [i.e., V_α in (9.4)] by anything one likes¹¹³ without affecting the mathematical structure of Eqs. (9.1)–(9.7). The usual choice is to write

$$V = V_\beta + (V - V_\beta), \quad (9.24)$$

where V_β is an arbitrary atomiclike potential, equal to or nearly equal to V in atomic cell β , and chosen to facilitate solution of Eq. (9.25) below for orbital ϕ_α . The V_α in (9.4) is replaced by $(V - V_\beta)$, so that in cell α it is in fact V_α

¹¹¹ D. W. Bullett, Thesis submitted to the University of Cambridge, p. 111, 1974 (unpublished).

¹¹² I. L. F. Ray and D. J. H. Cockayne, *Philos. Mag. [8]* **22**, 853 (1970); *Proc. R. Soc. London, Ser. A* **325**, 543 (1971).

¹¹³ This is a well-known feature of standard pseudopotential theory. See Heine,⁸³ Eq. (VII.3); also Anderson.^{74,76}

minus the tail of the potential V_β overlapping into cell α . Thus (9.2) and (9.4) become

$$\left\{ T + V_\beta + \left[(V - V_\beta) - \sum' |\alpha\rangle\langle\alpha|(V - V_\beta) \right] \right\} |\beta\rangle = E_\beta |\beta\rangle, \quad (9.25)$$

and the off-diagonal matrix elements in (9.7) and (9.12) become

$$H_{\alpha\beta} = \langle\alpha|(V - V_\beta)|\beta\rangle. \quad (9.26)$$

By similar arguments an alternative equation for the case of the hydrogen molecule is¹¹⁴

$$[T + V_1 + (V_2 - PV_2)]|1\rangle = E_1|1\rangle, \quad (9.27)$$

where P is the projection operator onto the function space $\{|\alpha\rangle\}$, $\alpha = 1$ and 2:

$$P = \sum_{\alpha\beta} |\alpha\rangle(S^{-1})_{\alpha\beta}\langle\beta|. \quad (9.28)$$

This gives precisely the conventional tight-binding answer for the bonding and antibonding energy levels.

It remains to give one or two examples of the application of this method. It yielded the matrix elements (including their dependence on bond angles) for the study¹¹⁵ of the electronic structure of various models of amorphous silicon by the recursion method. As for total energy, the trace invariance method above is not applicable to most systems, particularly metallic ones. The local density of states is also not so useful when there are several elements or geometrically different atoms as in the adsorption of oxygen on nickel and platinum (001) surfaces so that Bullett¹¹⁶ calculated and summed eigenvalues from a secular equation for a thin slab of material. In these systems the greatest binding is found for a site equidistant from four surface atoms in agreement with the LEED determinations for nickel, and the difference in bonding between different sites is greater than for hydrogen on the same metals due to greater directionality in the p orbitals of oxygen. Figure 11 shows a difference between the two substrates: in nickel the spectral weight of all three oxygen p orbitals is concentrated in a single peak just below the nickel d band, whereas on platinum with wider d band the P_x , P_y spectral weight is now spread throughout the band, corresponding to greater admixture with d orbitals.

¹¹⁴ D. W. Bullett, private communication.

¹¹⁵ For discussion and references see Kelly's chapter, Section 19.

¹¹⁶ D. W. Bullett and M. L. Cohen, *J. Phys. C* **10**, 2101 (1977); see also D. W. Bullett, *Surf. Sci.* **68**, 149 (1977).

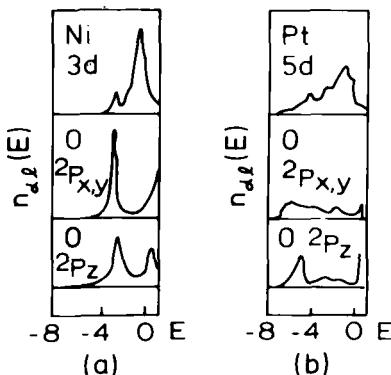


FIG. 11. Local density of states $n_{d,f}(E)$ for transition metal and oxygen at (a) oxygen on nickel (001) surface and (b) oxygen on platinum (001) surface. The energy E (eV) is measured relative to the Fermi level (after Bullett and Cohen¹¹⁶).

Finally we mention some current calculations with the method on the metal cluster molecules¹¹⁷ $[M_3(CO)_6]_n^{2-}$ with $M = Ni$ and Pt . The energy levels and matrix elements are found consistent with the red color, and with the change in color on forming dimers out of monomers, although no complete optical spectrum has yet been calculated or measured. The monomer molecule $n = 1$ is planar (Fig. 12), these stacking vertically on top of one another to produce the polymers $n = 2, 3, \dots$ (not all of which are known experimentally). Chang and Woolley¹¹⁷ calculated the energy difference between the “staggered” and “eclipsed” stacking for the dimers. For platinum and nickel the difference was almost zero in agreement with the observation of free rotation in the platinum trimer in solution. In fact, for platinum the staggered configuration had a slightly lower energy by 0.006 a.u.²⁰ after relaxation at the measured spacing, consistent with its observed shape in the solid state. However, the interaction energy as a function of distance between the two monomers for platinum had its minimum at a distance 2 a.u. larger than the observed spacing, an illustration of the common observation that variation with bond length is much more sensitive than calculation of conformational energy changes for reasons that we discuss in Sections IV,13 and V,16. In this instance, the bond length between transition metal atoms is determined by a complicated balance between the effects of the s and d electrons, which is only beginning to be understood¹¹⁸ (see Section V,16).

¹¹⁷ K. W. Chang and R. G. Woolley, *J. Phys. C* **12**, 2745 (1979); also private communication; see also R. G. Woolley, in “Transition Metal Clusters” (B. F. G. Johnson, ed.). Wiley, New York, 1980 (to be published).

¹¹⁸ D. G. Pettifor, *J. Phys. F* **8**, 219 (1978).

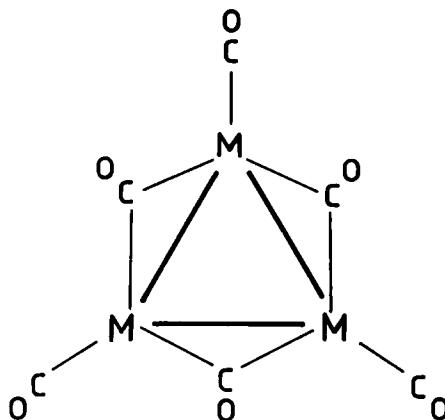


FIG. 12. Planar structure of $M_3(CO_6)_2^-$ monomer molecule.

For further examples of the use of the Anderson–Bullett scheme we refer to Bullett's chapter¹⁵ in this volume, and as the provider of matrix elements for a few applications of the recursion method in Kelly's chapter.²

10. FROM ATOMIC RESONANCES TO LCAO PARAMETERS FOR TRANSITION METALS

Probably about half of the calculations described in Section IV,11 and in the applications surveyed in Kelly's chapter² concern transition metals. For these the LCAO hopping matrix elements were obtained by a combination of the resonance theory sketched in this section together with empirical fitting.¹¹⁹ The resonance theory gives *simple analytic expressions* for the hopping parameters between d orbitals and the hybridization with sp electrons. Of course, one might feel more secure with empirical parameters directly fitted to a band structure, but often these have to be scaled to different elements or different interatomic distances, e.g., in calculating the relative energies of the three Laves structures⁴ for transition metal alloys. The resonance theory is then an indispensable guide. Incidentally, a similar theory can be developed for the p bands in compounds involving oxygen or chalcogens,¹²⁰ and one can probably produce an analytic theory for the electronic structure of a transition metal oxide, say, in terms of three parameters.

¹¹⁹ Both of these topics are discussed in more detail in Bullett's chapter.¹⁵

¹²⁰ O. K. Andersen, W. Klose, and H. Nohl, *Phys. Rev. B* **17**, 1209 (1978).

Historically, Friedel¹²¹ first applied the resonance approach to single atoms of a transition metal or p-electron element dissolved in a nearly free-electron metal. The idea was applied to the band structures of transition metals by¹²² Ziman,¹²³ Heine,⁸⁴ Pettifor,¹²⁴ and their co-workers, the latest stage of development being due to Andersen.^{120,125,126}

As is well known, the d bands of transition metals cover a relatively narrow energy range that is part of the nearly free-electron (NFE) bands (see Fig. 21 of Bullett's chapter¹⁵). Thus a tight-binding approach might appear impossible; if one integrates out the Schrödinger equation from the origin, one will get a wave function looking like an atomic d orbital

$$\phi_m(\mathbf{r}) = \phi(r) Y_{2m}(\theta, \phi) \quad (10.1)$$

inside the atom, but with an oscillating tail stretching to infinity: strictly speaking, the d states are not bound at all. Nevertheless, the d bands $E(\mathbf{k})$ manifestly have tight-binding form.^{71,72} To be more precise, the radial part of the wave equation contains a repulsive centrifugal barrier²⁰ $l(l+1)/2r^2$, which serves to confine a d electron more or less inside the atom, but it can of course tunnel out through the barrier into the NFE band outside (see Fig. 19 of Bullett's chapter¹⁵), i.e., the d states "resonate" with the plane-wave states of the same energy.¹²¹

Let us take the potential $V(\mathbf{r})$ inside a transition metal atom and integrate the radial part $u(r, E)$ of the wave function outwards to some radius, which for convenience we choose as the atomic sphere radius s . The logarithmic derivative $L_2(E) = u'(s, E)/u(s, E)$ has a very characteristic shape as a function of E (see Fig. 18 of Bullett's chapter¹⁵) from which the whole theory can be developed. But it will suit us better to surround the atom by a constant potential to infinity, which we take as equal to the energy at the bottom of the NFE band so as to represent this sea of electrons, although that is not necessary. In any case it becomes the zero of energy for²⁰ $E = \frac{1}{2}\kappa^2$. We can then turn the $L_2(E)$ into a phase shift $\eta_2(E)$ [Eq. (12.13) of Bullett's chapter¹⁵], which has the form with a characteristic resonant first term^{84,127}

$$\tan \eta_2(E) = \frac{\frac{1}{2}W(E)}{E_0(E) - E} + \tan \eta'_2(E). \quad (10.2)$$

¹²¹ J. Friedel, *Can. J. Phys.* **34**, 1190 (1956).

¹²² The following papers are only the main milestones. For further references, see Bullett.¹⁵

¹²³ J. M. Ziman, *Proc. Phys. Soc., London* **86**, 337 (1965).

¹²⁴ D. G. Pettifor, *J. Phys. C* **2**, 1051 (1969).

¹²⁵ O. K. Andersen, *Phys. Rev. B* **12**, 3060 (1975).

¹²⁶ A. R. Mackintosh and O. K. Andersen, in "Electrons at the Fermi Surface (M. Springford, ed.). Cambridge Univ. Press, London and New York, 1980 (to be published).

¹²⁷ L. D. Landau and E. M. Lifshitz, "Quantum Mechanics," p. 441. Pergamon, Oxford, 1958.

Hubbard^{128,129} has given explicit analytic expressions for all the quantities $W(E)$, $E_0(E)$, $\eta_2(E)$, and the associated wave functions. These are remarkably accurate over a wide range of E around the resonance as well as in it, even in the case of very broad resonances.¹³⁰

However, in order to introduce the main ideas we take the simplest approximation,

$$\tan \eta_2(E) = \frac{\frac{1}{2}W}{E_0 - E}, \quad (10.3)$$

where W and E_0 are constants, while emphasizing that for a more complete theory, especially for broad resonances, one must and can work with the full expression (10.2). $E_0 = \frac{1}{2}\kappa_0^2$ is the center of the resonance where the phase shift is $\pi/2$, and the width is^{84,128,129}

$$W = 2\kappa_0 \left[\int_0^s j_2(\kappa_0 r) V(r) \phi(r) r^2 dr \right]^2, \quad (10.4)$$

where $\phi(r)$ is the radial wave function $u(r, E_0)$ at the special energy E_0 , normalized such that

$$\int_0^s [\phi(r)]^2 r^2 dr = 1, \quad \phi(r) = u(r, E_0). \quad (10.5)$$

The physical interpretation¹²⁷ of the resonance width W in terms of the uncertainty principle is that \hbar/W is the escape time for an electron from the orbital ϕ_m into the plane wave states outside. The ϕ_m themselves are not exact energy eigenstates and W is the extent of their energy broadening. Such an interpretation is consistent with the form of (10.4), where $j_2(\kappa_0 r)$ is the $l = 2$ component of a plane wave, interacting via $V(r)$ with $\phi_m(\mathbf{r})$.

We now make the *small kappa approximation*,^{84,125,126} i.e., expand $j_2(\kappa_0 r)$ to lowest power in $\kappa_0 r$:

$$j_2(\kappa_0 r) \simeq (\kappa_0 r)^2 / 15. \quad (10.6)$$

One reason is that in many cases indeed the d band is not far up the NFE band so that (10.6) is quite legitimate. Another reason is that our choice for constant potential outside the atom was arbitrary: the real metal consists of atoms packed together with no space left for "outside." Third, we hardly expect the dd bandwidth and interaction to depend on the position of E_0 in the NFE band; indeed, κ_0 will drop out of our results. With this approximation (10.4) becomes

$$W = 2\kappa_0^5 M^2 / 225, \quad (10.7)$$

¹²⁸ J. Hubbard, *Proc. Phys. Soc., London* **92**, 921 (1967).

¹²⁹ See also D. G. Pettifor, *J. Phys. C* **5**, 97 (1972).

¹³⁰ J. H. Davies and V. Heine, *J. Phys. C* **11**, 4957 (1978); J. H. Davies, unpublished calculations.

where M is the reduced matrix element

$$M = \left| \int_0^s V(r) \phi(r) r^4 dr \right|. \quad (10.8)$$

The final piece of formal machinery that we need is the radial wave function at and outside the atomic sphere^{84,128,129}

$$u(r, E) = (\frac{1}{2}\kappa_0 W)^{1/2} [\cot \eta_2(E) j_2(\kappa r) - n_2(\kappa r)], \quad r > s, \quad (10.9)$$

where n_2 is the usual spherical Neumann function. We apply again the small kappa approximation [$n_2(x) \simeq -3/x^3$] and (10.3) to obtain

$$u(r, E) = M \left[\frac{E_0 - E}{M^2} r^2 + \frac{1}{5r^3} \right], \quad \text{for } r > s. \quad (10.10)$$

Note that this is independent of κ_0 as promised. At the center E_0 of the resonance, (10.10) must join smoothly onto $\phi(r)$ at the radius s :

$$u(s, E_0) = \phi(s) = M/5s^3, \quad (10.11)$$

and $\phi(r)$ is an atomiclike d orbital with a tail outside the atomic sphere given by

$$\phi(r) = M/5r^3 \quad \text{for } r > s. \quad (10.12)$$

We can now calculate the *bandwidth* Δ of the d band as our first application of the theory.⁸⁴ The bottom of the band E_{\min} corresponds to a perfect bonding relationship between the d wave functions on neighboring atoms. As in the original work of Wigner and Seitz⁵⁹ this is obtained to a good approximation when

$$u'(s) \equiv [\partial u(r, E_{\min})/\partial r]_{r=s} = 0. \quad (10.13)$$

Neighboring orbitals can then be joined up smoothly as a bonding combination (Fig. 13). Differentiating (10.10) we obtain for the condition (10.13)

$$2s \left(\frac{E_0 - E_{\min}}{M^2} \right) - \frac{3}{5s^4} = 0;$$

i.e.,

$$E_{\min} = E_0 - \frac{3}{10} \frac{M^2}{s^5}. \quad (10.14)$$

Similarly the top of the band E_{\max} represents an antibonding combination between neighboring orbitals, i.e., we require the wave function to be zero

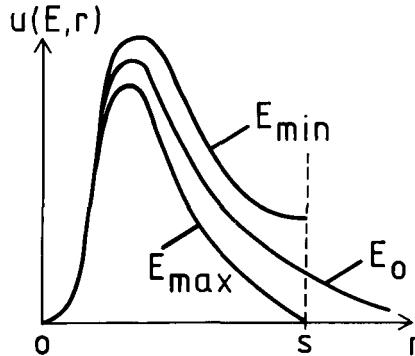


FIG. 13. Radial 3d wave function $u(E, r)$ for three energies satisfying three different boundary conditions.

half-way between atoms, which we take on average to be the distance s (Fig. 13):

$$u(s, E_{\max}) = 0. \quad (10.15)$$

From (10.10) this gives

$$E_{\max} = E_0 + \frac{1}{5} \frac{M^2}{s^5}, \quad (10.16)$$

which with (10.14) yields the *bandwidth*

$$\Delta = \frac{1}{2} M^2 / s^5. \quad (10.17)$$

This result for the bandwidth calls for several comments. We have been somewhat cavalier about the nearest-neighbor geometry and distance, so that (10.17) is an approximation for any of the densely packed structures fcc, bcc, and hexagonal close-packed, which in fact have the same bandwidth to within 10%^{131,132} and in agreement with (10.17).⁸⁴ We also note the distance dependence s^{-5} in (10.17), which has been verified as approximately correct by band structure calculations for expanded or compressed metals.¹³² However, there are higher order correction terms that become appreciable for broad resonances [see Eq. (15.23) of Bullett's chapter¹⁵]. Since the d-band structure is determined by the $dd\sigma$, $dd\pi$, and $dd\delta$ hopping integrals,¹³³ we might infer from (10.17)

$$dd\sigma/\pi/\delta \propto M^2/R^5, \quad (10.18)$$

¹³¹ See below Eq. (15.22) of Bullett's chapter.¹⁵

¹³² See, for instance, D. G. Pettifor, *J. Phys. F* 7, 613 (1977).

¹³³ If the two atoms are centered along the z axis, the $dd\sigma$ integral is the interaction between normalized $(3z^2 - r^2)$ -type functions on the two atoms, $dd\pi$ between xz (or yz) functions, and $dd\delta$ between $x^2 - y^2$ (or xy) functions.

where R is the interatomic spacing, again with an R^{-5} distance dependence. The κ_0 has cancelled out of our result (10.17) as promised: since we are dealing purely with the d-band width, where it lies relative to the NFE band is irrelevant.

However, the main surprise of result (10.17) is the paradox^{84,85} that the bandwidth, which we think of as an interatomic effect, is determined by the intra-atomic reduced matrix element M . The argument leading to it is so simple that it can hardly be wrong, at least not in its physical content as a reasonable approximation: Eq. (10.9) is standard theory, the multiplying coefficient being irrelevant in the derivation of (10.17), and (10.4) follows by taking $\eta_2 = \frac{1}{2}\pi$ in the standard integral formula¹³⁴ for $\sin \eta_l$. We return to this point shortly.

We can take the argument a step further and make from (10.17) and (10.18) an estimate of $dd\sigma$, the largest of the hopping parameters. In view of the angular dependence of the $l = 2$ spherical harmonics and the directions in which they are zero, it is clearly impossible to construct a total wave function for the solid in which each atom has σ bonds (or antibonds) to all its nearest neighbors, but it seems reasonable that we can form about four such per atom. Thus we can write

$$\Delta \simeq 8|dd\sigma|, \quad (10.19)$$

which is intended to be only a rough approximation, say within a factor of 2, because $dd\pi$ and $dd\delta$ integrals also contribute to Δ , although the numerical factor 8 just happens to give the correct answer to 1%, as we shall see. From (10.17), and (10.19) we have

$$|dd\sigma| \simeq (1/16)M^2/s^5. \quad (10.20)$$

In order to insert explicitly the R^{-5} distance dependence for arbitrary R , we note that for the nearest neighbors in the fcc structure we have

$$s \gtrsim \frac{1}{2}R = 0.5526R,$$

i.e.,

$$s/0.5526R = 1. \quad (10.21)$$

From (10.21) and noting that $dd\sigma$ is negative, we can rewrite (10.20) as

$$dd\sigma \simeq -\frac{1}{16} \frac{M^2}{s^5} \left(\frac{s}{0.5526R} \right)^5 = -1.2 \cdots (M^2/R^5). \quad (10.22)$$

Aside from doubts about the numerical coefficient, all this is quite respectable.

¹³⁴ A. Messiah, "Quantum Mechanics," p. 405. North-Holland Publ., Amsterdam, 1961. But note the different normalization used in this reference.

We now journey to understand these hopping integrals $dd\sigma$, $dd\pi$, and $dd\delta$ in greater inwardness, starting first across firm ground but then taking off in a flight of heuristic exploration to arrive at numerically the correct result. The hopping integral between orbitals $\phi_m(\mathbf{r} - \mathbf{R}_A)$ and $\phi_m(\mathbf{r} - \mathbf{R}_B)$ on atoms centered at \mathbf{R}_A , \mathbf{R}_B is

$$\int \phi_m^*(\mathbf{r} - \mathbf{R}_A) V(\mathbf{r} - \mathbf{R}_B) \phi_m(\mathbf{r} - \mathbf{R}_B) d^3\mathbf{r}. \quad (10.23)$$

As already mentioned in connection with Eq. (7.11), the major part of this integral does *not* come from the region halfway between \mathbf{R}_A and \mathbf{R}_B as one envisages (correctly) for sp bonds. Because for d orbitals ϕ_m peak at such a small radius (0.6 a.u. in copper) *the dominant part of the hopping integrand in (10.23) lies deep inside atom B, which explains qualitatively the reason for intra-atomic matrix elements determining Δ in (10.17) and (10.18)*. In fact, (10.23) becomes

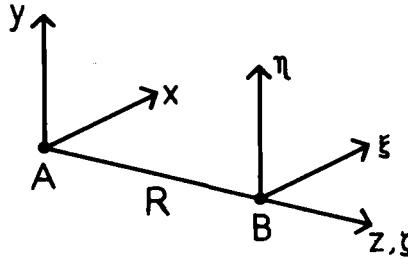
$$\int (\text{tail from } A \text{ spreading onto } B) V(\mathbf{r} - \mathbf{R}_B) \phi_m(\mathbf{r} - \mathbf{R}_B) d^3\mathbf{r}, \quad (10.24)$$

where from (10.9) and (10.10) the amplitude of the tail is proportional to M , which gives the first factor of M in (10.17) and (10.18). Because of the spherical harmonic in $\phi_m(\mathbf{r} - \mathbf{R}_B)$ we expand the tail from atom A also in spherical harmonics about \mathbf{R}_B and the remaining integral turns out identical to (10.8), thus giving the other factor of M in (10.17) and (10.18). So far the argument is quite general and remains valid in essence whether or not one makes any of the approximations in (10.6) etc.⁸⁵ However, the right way (or a right way) to obtain good numerical results from (10.24) is to use for ϕ_m the radial orbital (10.12) with energy E_0 and the small kappa expansion. Certainly this is the only way of obtaining a monotonically decreasing orbital out of (10.9), which makes sense for a tight-binding theory: anything else from (10.9) would give oscillating interactions decreasing only slowly with distance. However, there must be fuller mathematical reasons hidden in Andersen's theory,^{125,126} in particular one should be able to prove that this choice of orbital also gives good convergence of the hybridization interaction (see below) to higher plane waves since the latter would otherwise also contribute significantly to the d-band width by second-order perturbation theory.^{124,135,136} These points remain to be explained.

With such justification then, we insert the radial orbital $\phi(r)$ of (10.5), (10.11), and (10.12) in Eq. (10.23) to evaluate $dd\sigma$, $dd\pi$, and $dd\delta$. Take \mathbf{R}_A at the origin and \mathbf{R}_B a distance R along the z axis (Fig. 14) and define ξ, η, ζ

¹³⁵ R. L. Jacobs, *J. Phys. C* **1**, 492 (1968).

¹³⁶ See Section 13 of Bullett's chapter,¹⁵ from below Eq. (13.7).

FIG. 14. Coordinates for evaluation of hopping integrals $dd\sigma/\pi/\delta$.

axes parallel to x, y, z , but with origin at \mathbf{R}_B . For simplicity let us calculate first $dd\delta$, for which we take

$$\phi_m(\mathbf{r} - \mathbf{R}_B) = C(\xi\eta/\rho^2)\phi(\rho), \quad (10.25)$$

with

$$\rho^2 = \xi^2 + \eta^2 + \zeta^2. \quad (10.26)$$

In (10.25) C is an angular normalizing factor [ϕ is normalized by (10.5)] and $\xi\eta/\rho^2$ is the angular factor $Y_{2m}(\theta, \phi)$ for an xy -type d orbital. Similarly we have from (10.12) for the tail of $\phi_m(\mathbf{r} - \mathbf{R}_A)$ extending onto atom B ,

$$\text{tail of } \phi_m(\mathbf{r} - \mathbf{R}_A) = C(xy/r^2)(M/5r^3). \quad (10.27)$$

Since \mathbf{R}_A and \mathbf{R}_B are separated a distance R along the z axis, we have $x = \xi$ and $y = \eta$, and r in the region of atom B is effectively R . Thus (10.27) becomes

$$\text{tail of } \phi_m(\mathbf{r} - \mathbf{R}_A) = C(\xi\eta/\rho^2)\rho^2 M/5R^5. \quad (10.28)$$

When we substitute (10.25) and (10.28) into (10.24), the angular function $C^2(\xi\eta/\rho^2)^2$ integrates to unity by normalization, and we are left with

$$\int (\rho^2 M/5R^5) V(\rho) \phi(\rho) \rho^2 d\rho, \quad (10.29)$$

which reduces to the M integral (10.8). Remembering that V is negative we obtain finally

$$dd\delta = -M^2/5R^5, \quad (10.30)$$

consistent with (10.18). We can similarly use the orbitals

$$C(xz/r^2)\phi(r) \quad (10.31)$$

to obtain the $dd\pi$ hopping integral. The tail from A is now

$$\text{tail from } A = C(xz/r^2)(M/5r^3), \quad (10.32)$$

which we must expand properly this time as a Taylor series about the point \mathbf{R}_B , picking out the term

$$(2\xi\zeta/2!)\left[\frac{\partial^2}{\partial x \partial z} (\text{tail from } A)\right]_{\mathbf{r}=\mathbf{R}_B}, \quad (10.33)$$

which by the orthogonality of spherical harmonics is the only term giving a nonzero contribution to the integral with $\phi(\mathbf{r} - \mathbf{R}_B)$. We can cut a corner by noting that near \mathbf{R}_B we have r effectively equal to z , and thus write (10.32) as

$$\text{tail from } A = CM/5z^4, \quad (10.34)$$

from which the differentiation in (10.33) gives

$$(\xi\zeta)(-4CM/5z^5)_{z=R} = C(\xi\zeta/\rho^2)(-4M/5R^5)\rho^2. \quad (10.35)$$

Substitution of (10.35) and

$$\phi_m(\mathbf{r} - \mathbf{R}_B) = C(\xi\zeta/\rho^2)\phi(\rho) \quad (10.36)$$

into (10.24) yield the same angular and radial integrals as before, giving the result

$$dd\pi = +4M^2/5R^5. \quad (10.37)$$

Finally, $dd\sigma$ can be evaluated similarly from the $(3z^2 - r^2)$ -type of orbital, in this case the corresponding term in the Taylor series expansion being

$$(1/6) \frac{[\partial^2/\partial\xi^2 - \partial^2/\partial\zeta^2]}{[3\xi^2 - \rho^2]}. \quad (10.38)$$

The result is

$$dd\sigma = -6M^2/5R^5 \quad (10.39)$$

in almost exact agreement with (10.22). We have therefore calculated all three hopping integrals $dd\sigma$, $dd\pi$, and $dd\delta$.

We now wish to relate our results to the notation and formalism of Andersen's theory^{125,126} (see Section 15 of Bullett's chapter¹⁵) in view of its importance and the fact that we have been partly using its ideas here. We have in (10.7) related our parameter M to the resonance width, but we can equally well from (10.9) and (10.10) take it as a measure of the amplitude of the tail of the wave function. We obtain the same physical interpretation because the rate of an electron inside the atom escaping will be proportional to the square of the tail wave function. With the aid of (10.11) we can therefore express all our results in terms of Andersen's parameter [see Eq. (15.24) of Bullett's chapter¹⁵]

$$(\mu_d s^2)^{-1} = \frac{1}{2}s[u(s, E_0)]^2 = M^2/50s^5. \quad (10.40)$$

The bandwidth (10.17) becomes¹³²

$$\Delta = 25(\mu_d s^2)^{-1} \quad (10.41)$$

in agreement with Eq. (15.23) and the statement below Eq. (15.22) in Bullett's chapter,¹⁵ while the hopping integrals (10.30), (10.37), and (10.39) are seen to be identical to Eq. (15.25) of Bullett's chapter, first obtained by Pettifor¹³² and by Anderson *et al.*¹²⁰

We now turn to the interaction or *hybridization* of the d orbitals with the NFE band, which gives rise to the resonance width W as discussed below Eq. (10.5). In the band structure of a transition metal it manifests itself in the way shown in Fig. 21 of Bullett's chapter¹⁵: the NFE band and the d band of the same symmetry cannot cross (shown as dashed lines) but are split apart by the interaction (full curves).

Let us calculate the hybridization matrix element

$$h_{mk} = \langle \phi_m | V | PWk \rangle \quad (10.42)$$

between one orbital and the plane wave $|PWk\rangle$ having the same energy $E_0 = \frac{1}{2}k^2$. It is interesting to note why the potential V occurs in (10.4), (10.8), and (10.42) and not the full Hamiltonian $\mathcal{H} = T + V$, where T is the kinetic energy operator. The d orbitals and the plane waves are not orthogonal, having a nonzero overlap integral S between them. According to the theory of Eq. (7.2) the interaction for nonorthogonal orbitals should be given by

$$\langle \phi_m | T + V - E | PWk \rangle, \quad (10.43)$$

but T operating on $|PWk\rangle$ just gives its energy E , thus canceling the overlap term ES . We thus obtain (10.42). At least this argument is valid at the energy E_0 of the resonance.

We return to evaluate (10.42). The $l = 2$ component of the plane wave \mathbf{k} with normalization volume Ω is

$$\Omega^{-1/2} 4\pi i^2 j_2(kr) \sum_m Y_{2m}^*(\hat{\mathbf{k}}) Y_{2m}(\theta, \phi), \quad (10.44)$$

where $\hat{\mathbf{k}}$ designates the direction of \mathbf{k} . Substitution in (10.42) and the small κ expansion (10.6) yield

$$h_{mk} = -4\pi\Omega^{-1/2} \left[\int \phi(r) V(r) j_2(kr) r^2 dr \right] Y_{2m}^*(\hat{\mathbf{k}}), \quad (10.45a)$$

$$= (4\pi/15\Omega^{1/2}) M k^2 Y_{2m}^*(\hat{\mathbf{k}}). \quad (10.45b)$$

Let us take the $m = 0$ orbital with normalized spherical harmonic

$$Y_{2m}(\theta, \phi) = (5/16\pi)^{1/2} (3 \cos^2 \theta - 1), \quad (10.46)$$

and \mathbf{k} in the z direction. Then (10.45b) with $\theta_k = 0$ becomes the hybridization matrix element

$$h_{m=0, k \text{ in } z \text{ direction}} = (4\pi/45\Omega)^{1/2} k^2 M. \quad (10.47)$$

Here Ω should be the volume per atom. To be precise, in a crystal with N atoms we should consider the interaction of the Bloch tight-binding function

$$\phi_{mk} = N^{-1/2} \sum_l \exp(i\mathbf{k} \cdot \mathbf{R}_l) \phi_m(\mathbf{r} - \mathbf{R}_l) \quad (10.48)$$

summed over atomic sites \mathbf{R}_l , but the various factors of N cancel to give (10.47). Thus the gap opened up between the two interacting bands in Fig. 21 of Bullett's chapter¹⁵ is twice the interaction (10.47). The formula gives good answers.⁸⁴

So much for the hybridization at energy E_0 , but what form does the interaction of (10.48) with $|\text{PWk}\rangle$ take for arbitrarily high k ? In other words, what is the correct expression to insert in the d-PW off-diagonal block in the model Hamiltonian Eq. (13.8) of Bullett's chapter?¹⁵ The current analytic theory of this is complicated^{84,124,135} (see Section 13 of Bullett's chapter).¹³⁶ For low k the k^2 dependence of h_{mk} is correctly given by (10.45), but for larger k , a bit larger than the radius of the Brillouin zone, the hybridization cuts off rapidly to zero.⁷² As already stated, it should be possible to short-circuit the older complicated theory of Lorentzian and Gaussian splits¹³⁶ and derive the form of the hybridization from Andersen's^{125,126} theory and the orbitals $\phi(r)$ defined by (10.5) and (10.12), but this has not yet been done.

In any case we have demonstrated our main point: from the logarithmic derivative of the wave function for an atom, characterized by a small number of parameters (mainly M , also κ_0 , and the high- k cutoff of h_{mk} in our simplest version of the theory) *it is possible to develop analytically and quantitatively the dd hopping integrals and the d-PW hybridization needed for describing the electronic structure of a transition metal.*

One point remains to be tidied up: we have derived the hybridization interaction in the form of matrix elements between d orbitals and plane wave states. How is this *hybridization to be described in LCAO terms*? The lower nearly free-electron sp bands are quite well describable in terms of atomic s and p states.⁵⁹ Since we are more concerned with illustrating the line of the argument than with giving an exhaustive treatment, let us confine ourselves to s orbitals ϕ_s to write

$$|\text{PWk}\rangle \simeq \sum_l \exp(i\mathbf{k} \cdot \mathbf{R}_l) \phi_s(\mathbf{r} - \mathbf{R}_l) \quad (10.49)$$

normalized to unity per atom, the same as (10.44). There is no hybridization between the d orbital ϕ_m at the origin and ϕ_s on the same atom because of their different symmetries. However the $\phi_s(\mathbf{r} - \mathbf{R}_l)$ on neighboring atomic

centers \mathbf{R}_l overlap onto the central atom and have an $l = 2$ component in their tail, which does hybridize with ϕ_m . As in calculating the dd hopping integrals, we do a Taylor series expansion of the tail of $\phi_s(\mathbf{r} - \mathbf{R}_l)$ about the origin. Since ϕ_s is spherically symmetric the only $l = 2$ term is

$$\frac{1}{6}Ar^2(3\cos^2\theta_{rl} - 1) = (4\pi/15)Ar^2 \sum_{m'} Y_{2m'}^*(\hat{l})Y_{2m}(\theta, \phi), \quad (10.50)$$

where θ_{rl} is the angle which the vector $\mathbf{r} = (x, y, z)$ centered at 0 makes with the unit vector \hat{l} , which specifies the direction of \mathbf{R}_l . Also A is given by

$$A = \left[\frac{d^2\phi_s}{dr^2} - \frac{1}{r} \frac{d\phi_s}{dr} \right]_R. \quad (10.51)$$

Now the integral of (10.50) with $\langle \phi_m | V$ will pick out the $m' = m$ component of (10.50). The radial integral becomes

$$\int \phi_d(r)V(r)r^2 \cdot r^2 dr = -M \quad (10.52)$$

as before, giving

$$\langle \phi_m(0) | V(0) | \phi_s(\mathbf{R}_l) \rangle = -(4\pi/15)AMY_{2m}^*(\mathbf{l}). \quad (10.53)$$

We can express this result in terms of the standard $s d\sigma$ hopping integral obtained by taking $m = 0$ and \hat{l} in the z direction:

$$sd\sigma = -\frac{2}{3}(\pi/5)^{1/2}AM. \quad (10.54)$$

We return now to the summation in (10.49) and for simplicity confine ourselves to a bulk element in the fcc structure with \mathbf{R}_l running over the 12 nearest neighbors only at a distance R . The exponential factor in (10.49) can also be expanded in spherical harmonics and to a first approximation the cubic environment, like a spherically symmetric once, picks out the $l = 2$ term analogous to (10.44). We obtain

$$\begin{aligned} h_{mk} &= \langle \phi_m(0) | V(0) | \sum_l \exp(i\mathbf{k} \cdot \mathbf{R}_l) \phi_s(\mathbf{R}_l) \rangle \\ &= (16\pi^2/15)AM \sum_l Y_{2m}^*(\hat{l}) j_2(kR) \sum_{m'} Y_{2m'}^*(\hat{\mathbf{k}}) Y_{2m}(\hat{l}). \end{aligned} \quad (10.55)$$

Summing over 12 nearest neighbors is similar to integrating over the whole solid angle of \hat{l} so that \sum_l picks out the $m' = m$ term from the m' sum, with a numerical coefficient C that is not quite unity but $15/8\pi \simeq 0.6$. This coefficient will differ for other structures. We therefore obtain for the hybridization matrix element

$$h_{mk} = (16\pi^2/15)AMC j_2(kR) Y_{2m}^*(\hat{\mathbf{k}}) \quad (10.56a)$$

$$= (64\pi^3/5)^{1/2} C (sd\sigma) j_2(kR) Y_{2m}^*(\hat{\mathbf{k}}). \quad (10.56b)$$

We note this result has the same angular dependence as the previous expression (10.45b), the same proportionality to M , and to k^2 for small k . In fact we recover (10.45b) precisely if we write

$$\phi_s(r) = \Omega^{-1/2} f(r/R), \quad A = (\text{const}) \Omega^{-1/2} R^{-2}, \quad (10.57)$$

and take the constant as 2. Finally we note (10.56) cuts off at k beyond the “radius” of the Brillouin zone as we said below (10.48) that h_{mk} should: in fact, in comparison with Mueller⁷² it cuts off a little sooner. Of course “the proof of the pudding is in the eating” and one must carry out a fitting¹³⁷ to a known band structure, but our theoretical discussion bids us fair success even with one parameter $s\delta\sigma$ whose value we can already estimate from a simple look at the magnitude of the hybridization band splitting in a symmetry direction of $E(\mathbf{k})$. If we wish to improve the situation further, an extension of the analysis will help to answer such questions as: Is it better to include $s\delta\sigma$ to second neighbors or $p\delta\sigma$ and $p\delta\pi$ to nearest neighbors? If the latter, are $p\delta\sigma$ and $p\delta\pi$ independent quantities or related similar to $d\delta\sigma/\pi/\delta$ in (10.30), (10.37), and (10.39)? If so, what is their ratio?

In conclusion, therefore, we have come a long way in describing quantitatively the d bands and the d-PW hybridization of a transition metal in LCAO form, all more or less in terms of a single reduced matrix element M related to the resonance width. This is possible because the interaction with a transition atom occurs deep inside the atom where the d orbital and $V(\mathbf{r})$ are both large: we have, for example, neglected any change in $V(\mathbf{r})$ with crystal structure near the edge of the atom. The theory defines a standard transition metal that can be used for scaling between different elements, crystal structures, and interatomic spacings. The theory can obviously be extended to alloys of two or more transition metals, and to compounds with p bands of tight-binding form. As presented here the theory is only valid to lowest order in M but higher order effects can be and are calculated [see Eq. (15.23) of Bullett's chapter¹⁵]. Some of these corrections are crystal field terms that are zero in the lowest-order theory, thus explaining why Mueller⁷² and Hodges *et al.*⁷¹ found them so small. All in all the theory provides a first-order framework to which more detailed calculations must conform.

IV. Calculations with Atomic Orbitals, Atomic Vibrations, and Other Localized Basis Sets

11. WORKING WITH CLUSTERS, MOMENTS, AND CONTINUED FRACTIONS

In Sections II,4-II,6, we discussed the Green function matching method for calculating the local electronic structure, which has mostly been applied

¹³⁷ See, for example, W. E. Pickett and P. B. Allen, *Phys. Lett. A* **48**, 91 (1974).

to nearly free-electron types of situation including semiconductors, involving s and p electrons. We turn now to the *recursion method*^{1,2,138,139,140} specifically designed for the case where the physics is described in terms of a *local basis set* ϕ_{al} in the sense of Eq. (1.8). Usually the ϕ_{al} will be atomiclike orbitals of type α (s, p, d, etc.) on site \mathbf{R}_l calculated by the methods of Part III, but they may be *bond orbitals* between pairs of atoms, for instance, or with small changes of notation the ϕ_{al} may be unit atomic displacements to describe *lattice vibrations*.

The recursion method can calculate any quantity expressible as a *diagonal element of the Green function*

$$G_{\chi\chi}(E) = \langle \chi | (E + iO - H)^{-1} | \chi \rangle, \quad (11.1)$$

where χ is defined in terms of the ϕ_{al} . Commonest is just the local density of states

$$n_{al}(E) = -\pi^{-1} \operatorname{Im} \langle \alpha l | (E + iO - H)^{-1} | \alpha l \rangle \quad (11.2)$$

as discussed at the end of Section I,1.

The mechanics of calculating (11.1) and (11.2) is described in the next section, but we survey here some of the main ideas connected with the method and the kind of applications that can be made. The procedure hinges around defining a new *orthonormal basis set* u_n using the recurrence relation

$$b_{n+1} |u_{n+1}\rangle = H |u_n\rangle - a_n |u_n\rangle - b_n |u_{n-1}\rangle \quad (11.3)$$

starting with the first one $|u_0\rangle$ set equal to $|\chi\rangle$ in (11.1) or $|\alpha l\rangle$ in (11.2). The a_n, b_n are numerical coefficients. Let us suppose $|u_0\rangle$ is $|\alpha l\rangle$ localized on the central atom shown in Fig. 15. Then the Hamiltonian H in (11.3) is a matrix connecting $|u_0\rangle \equiv |\alpha l\rangle$ with the $|\alpha' l'\rangle$ on nearby atoms, nearest and next nearest neighbors, for example. In that case $H|u_0\rangle$ in (11.3) generates an extended orbital $|u_1\rangle$ spreading over these neighbors. Similarly, $H|u_1\rangle$ and hence $|u_2\rangle$ in (11.3) extend over the neighbors of the neighbors included in $|u_1\rangle$. Thus $|u_1\rangle, |u_2\rangle, |u_3\rangle$, etc. (Fig. 15) spread out further and further away from the central $|u_0\rangle$ whose local density of states (11.2) we desire. We therefore have a suitable basis set for describing the local electronic structure on the central atom, with the higher $|u_n\rangle$ and higher a_n, b_n contributing progressively less to the local density of states there.

The method clearly does not require lattice periodicity, only a finite number (typically of order 10,000) of orbitals $|\alpha l\rangle$ on some finite cluster of atoms, with the Hamiltonian matrix elements connecting them. Harking back to the discussion of Section I,1, we see that the method is suited to amorphous

¹³⁸ R. Haydock, V. Heine, and M. J. Kelly, *J. Phys. C* **5**, 2845 (1972).

¹³⁹ R. Haydock, V. Heine, and M. J. Kelly, *J. Phys. C* **8**, 2591 (1975).

¹⁴⁰ P. E. Meek, *Philos. Mag. [8]* **33**, 897 (1976).

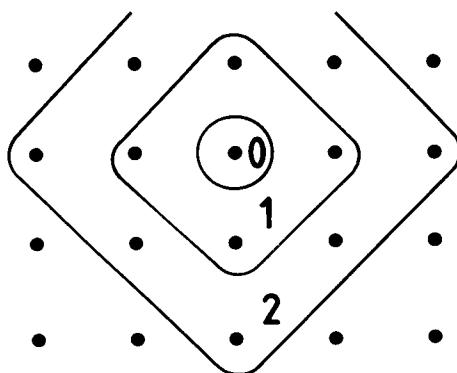


FIG. 15. The regions 0, 1, 2, etc., covered by the outward spreading basis orbitals u_0, u_1, u_2 , etc.

materials, surfaces, etc. enumerated there. We calculate the local density of states on an atom near the center of the cluster, and then appeal to the invariance theorem of Sections I,2 and II,5 to say the result is near enough to what would have been obtained with an indefinitely large cluster, i.e., the boundary corrections (Section II,5) arising from the finite cutting-off of the cluster are acceptably small.

Indeed, in the case of amorphous materials one can turn the argument backward. One has available carefully constructed and relaxed models comprising only a few hundred atoms, which sounds like a large number, but in fact with 512 atoms, seven-eighths of them lie within two atomic diameters of the cluster surface and are therefore strongly affected by surface states. Thus it is only from the *local* density of states near the center of the cluster that one can obtain results characteristic of the bulk material, not from the total density of states of the whole cluster. The recursion method is ideally suited to this purpose, and Figs. 7 and 10 of Kelly's chapter² show calculations of the electronic density of states for amorphous silicon and amorphous arsenic derived from different competing structural models for these materials. In the case of silicon the aim was to relate features of the electronic structure with the presence or absence of odd-numbered rings in the atomic structure, for comparison with experiment. Since in an amorphous material each atom has a slightly different environment, it is necessary to calculate and average over a few atoms, five to ten usually being more than adequate. Similarly in our Fig. 16¹⁴¹ we show analogous results for the density of vibrational modes $n(\omega)$ for various models of germanium, which

¹⁴¹ P. E. Meek, in "Proceedings of the IVth International Conference on the Physics of Non-Crystalline Solids at Clausthal-Zellerfeld" (G. H. Frischat, ed.), p. 586. Trans. Tech. Publ., Aedermannsdorf, Switzerland, 1977.

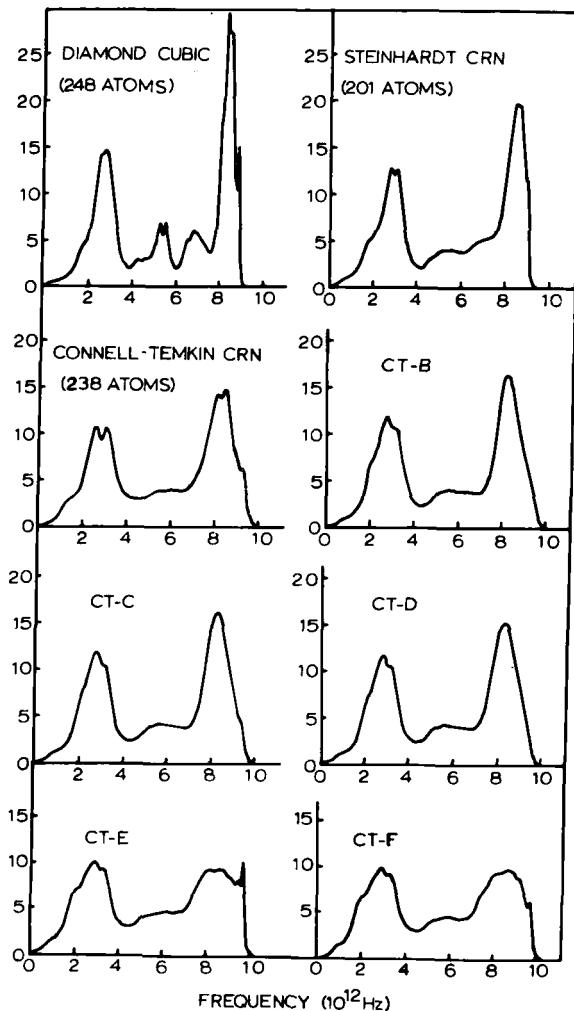


FIG. 16. Density of vibrational modes in crystalline and several models of amorphous germanium, calculated with a bond charge model (after Meek¹⁴¹).

were found to correlate not with the topological connectivity of the models but with their spread in the distribution of bond angles. (We do not labor the point here, but it is well known from the study of a simple linear chain that the calculation of the frequencies squared $\omega^2(\mathbf{k})$ of the phonons is algebraically the same as a tight-binding calculation of an electronic band structure $E(\mathbf{k})$, a result that is generally true for the dynamical matrix.¹⁴²)

¹⁴² J. D. Joannopoulos and W. B. Pollard, *Solid State Commun.* **20**, 947 (1976).

Some *local densities of electron states* computed with the recursion method have already been shown in Fig. 2 for the ferromagnetic alloy Fe₃Al. Figure 17 shows the results for the d band at two different surfaces¹⁴³ of the bcc structure, compared with that in the bulk. These early calculations considered the d band only, although more recent ones include the nearly free-electron sp band (see, for example, Fig. 5 of Kelly's chapter²). The results have been summed over the five d orbitals $m = -2$ to $+2$ on atom "zero" at the surface. On the (001) surface the xz and yz orbitals are clearly equivalent under the fourfold symmetry and give the same $n_{m0}(E)$ so that only one of them needs to be computed, but the $x^2 - y^2$, $2xy$, and $3z^2 - r^2$ orbitals all have different $n_{m0}(E)$. Figure 17 was in fact obtained by computing them separately and summing, but it has since been realized¹⁴⁴ that with a more sophisticated use of *symmetry* a single calculation with the starting state

$$|\chi\rangle = 5^{-1/2} \sum_m |m0\rangle \quad (11.4)$$

would suffice. Inserted into (11.1) it gives

$$\frac{1}{5} \sum_m \sum_{m'} G_{m0m'0}(E), \quad (11.5)$$

but all the cross terms $m' \neq m$ are zero since all the $m0$ have different symmetry under the $4mm$ rotational symmetry group at a surface atom. Thus (11.4) gives

$$\frac{1}{5} \sum_m G_{m0m0}(E), \quad (11.6)$$

and hence from (11.2) the mean surface density of states of the five orbitals. See Section 14 of Haydock's chapter¹ for a precise statement of the symmetry theorem.

Just as there is no restriction to nearest-neighbor interaction in the recursion method, so there is no restriction to orthogonal orbitals. For non-orthogonal orbitals the recursion relation (11.3) remains unaltered, with H being the matrix defined in Eqs. (7.4) and (7.8), but the formulas¹⁴⁵ for evaluating the coefficients a_n, b_n involve the overlap matrix S as detailed in Sections 3 and 5 of Haydock's chapter.¹

Figure 18 will serve as a peg on which to hang a discussion of the *accuracy* of our procedure. Of course, in computational physics there is no such thing as exactness, only varying degrees of convergence. The figure shows the local density of modes $n_l(\omega)$ at the center of a finite crystallite of germanium,

¹⁴³ R. Haydock and M. J. Kelly, *Surf. Sci.* **38**, 139 (1973).

¹⁴⁴ J. M. Gallagher, Thesis submitted to the University of Cambridge, 1978.

¹⁴⁵ An earlier alternative procedure described in the following reference suffers from some numerical instabilities and is not now recommended [R. Haydock, and M. J. Kelly, *J. Phys. C* **8**, 1290 (1975)].

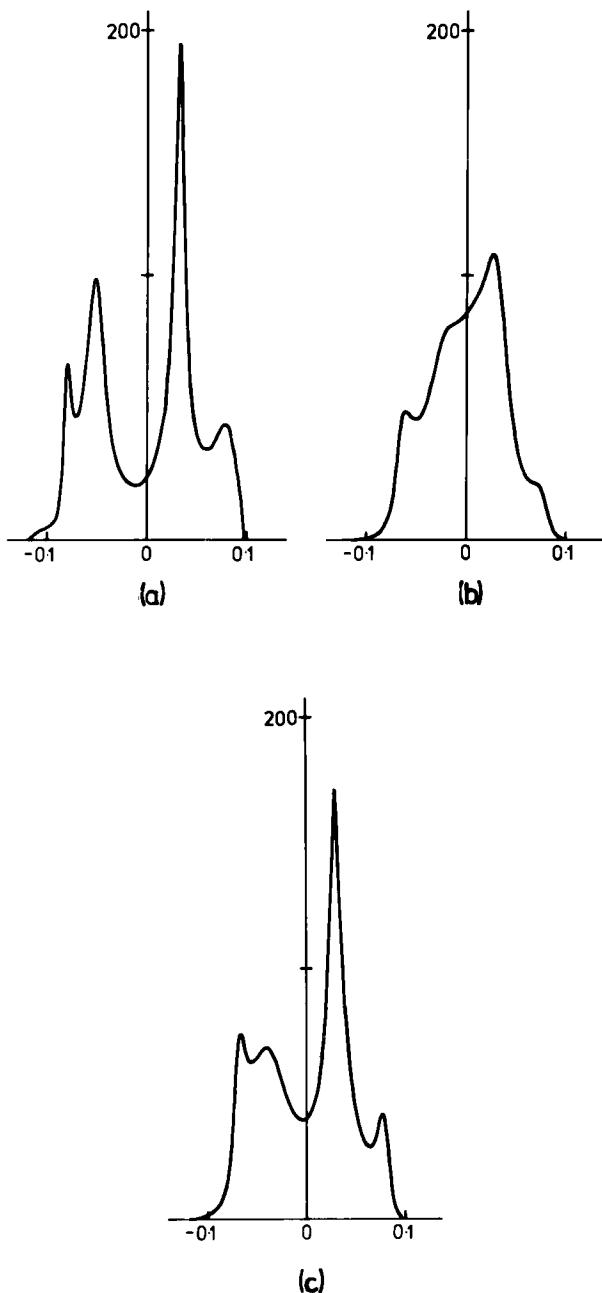


FIG. 17. Local density of states for a d band in a bcc structure: (a) in the bulk, and at a surface atom in a (111) surface (b) and a (110) surface (c), respectively (Haydock and Kelly¹⁴³).

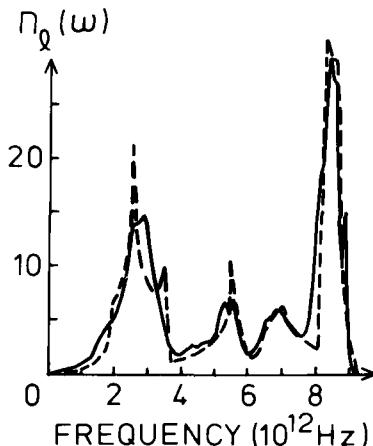


FIG. 18. Density of vibrational modes in crystalline germanium according to a bond charge model. Full line: calculated with the recursion method on a 248 atom cluster with free bonds at the surface. Broken line: calculated by Brillouin zone integration from the phonon dispersion relations for bulk germanium.

compared with the more exact result for the bulk diamond structure obtained from the phonon dispersion $\omega(\mathbf{k})$ by integrating over the Brillouin zone. For a similar comparison for the electron d bands in the fcc structure, see Fig. 2 of Kelly's chapter.² The two main points we wish to make relate to any use of finite clusters and not to the recurrence procedure as such. First, we see in Fig. 18 that the cluster calculation does not reproduce well the sharp Van Hove singularities⁸ in the density of modes, which is hardly surprising since their shape is intimately connected with \mathbf{k} space. To be more precise, we saw in Section II,5 that the cancellation argument about the effects of the cluster boundary breaks down as $\mathbf{k} \rightarrow 0$, i.e., as we approach the Van Hove singularity. Here \mathbf{k} is really to be interpreted as $\mathbf{k} - \mathbf{k}_0$, where \mathbf{k}_0 refers to the singularity. Stated physically, the perturbation caused by the cut-off at the edge of the cluster propagates inwards with infinite wavelength at the energy (frequency) of the Van Hove singularity.

Turning aside from the Van Hove singularities specifically, we can make some further comments about the cluster boundary corrections from the general theory of Section II,5. A spherical cluster seems best avoided, or any shape such as the octahedron for a simple cubic s band, which matches the shapes of the outward growing $|u_n\rangle$,¹⁴⁶ for in such cases the corrections

¹⁴⁶ C. M. M. Nex, unpublished calculations. These covered a variety of shapes from nearly spherical to re-entrant starlike shapes. A cube seemed a good compromise, and there seemed no advantage in computing the local density of states at an off-center point. The boundary effects were stronger for an s band than for a d band with several orbitals per site, as one might expect.

from all parts of the boundary add up coherently in phase at the center instead of interfering as envisaged in Section II,5. One might also wonder whether to chop off the cluster abruptly, leaving any severed bonds dangling, or whether to immerse the cluster effectively in a similar medium, or to simulate such a medium by applying periodic boundary conditions at the cluster surface or for amorphous clusters joining up surface atoms in some way, e.g., from opposite sides of the cluster. All these have been tried and it seems that^{139,147} with clusters of 1000 atoms or so the boundary corrections are sufficiently small for it hopefully not to matter which method one adopts. For smaller (crystalline) clusters one might expect that periodic boundary conditions are the best, eliminating errors to one higher order in $1/L$, where L is the diameter of the cluster,¹⁴⁸ but it is found in practice that this can be overshadowed by much larger deleterious side-effects.¹⁴⁷

The second major point about clusters is that the local (or total) *density of states is not a stable convergent quantity*.¹⁴⁹ Before elaborating on this statement, let the reader be reassured that almost all physical quantities are *integrals* over all the states of the system and that such integrals are satisfactorily convergent. However, the mentioned lack of stability does have some practical consequences seen often in cluster calculations, which have to be taken properly into account. Let us revert to the language of electron states rather than phonon modes. For any finite cluster, the energy levels from a discrete spectrum of eigenvalues E_n instead of a continuous band, and $n_{al}(E)$ is given by a corresponding set of delta functions

$$n_{al}(E) = \sum_n \omega_n \delta(E - E_n) \quad (11.7)$$

with weights ω_n given by Eqs. (1.10) and (1.13). Thus $n_{al}(E)$ for any given E is, strictly speaking, either zero or infinite, and it can easily change from one to the other if, by enlarging the cluster and changing the E_n , we move onto or off one of the delta functions. In that mathematical sense, (11.7) for any given E does not converge as the size of the cluster is increased. However, the physically significant "form" of $n_{al}(E)$ of course converges, the number of delta functions increasing with the cluster size and the weight ω_n of one eigenfunction on a given atom decreasing correspondingly. Thus mathe-

¹⁴⁷ M. J. Kelly and C. M. M. Nex, unpublished calculations.

¹⁴⁸ If for a finite one-dimensional chain of N atoms the total energy U has the form

$$U = Nu \times \gamma + O(1/N),$$

then u is interpreted physically as the bulk energy per atom and γ as the termination energy analogous to surface tension. Thus γ must vanish for periodic boundary conditions. The argument can be generalized and justified mathematically.

matically an integral over $n_{\alpha l}(E)$ converges,¹⁴⁹ in particular, the simplest integral, namely, the total number of electrons in orbital αl up to energy E :

$$N_{\alpha l}(E) = \int_{-\infty}^E n_{\alpha l}(E') dE'. \quad (11.8)$$

Similarly the total energy in this orbital converges with increasing cluster size:

$$U_d = \int_{-\infty}^{E_F} E' n_{\alpha l}(E') dE', \quad (11.9)$$

where E_F is the Fermi level. In fact, the convergence is good enough that the method has been employed for calculating quite small energy differences¹⁵⁰ for crystal structures,⁴ lattice distortions,^{151,152} and ferromagnetism.⁷ But strictly the local density of states for any finite cluster is a mess of spikes, and one must not complain if that is what the calculations give. We discuss presently how one can best smooth it into a more acceptable curve.

Apart from the effect of the cluster size, there is usually also a *truncation error* in the use of the recursion method itself. The computation of the $|u_n\rangle, a_n, b_n$ in (11.3) is the most time-consuming part of the calculation, which must be stopped at some value $n = N$ say, typically 15–25 recursions. It has already been stated that the higher $|u_n\rangle$ affect the answer less and less. In fact truncating the recursion has very much the same effect as using a finite cluster instead of an infinite medium. In the recursion method, the diagonal element of the Green function in (11.1) and (11.2) is given by a *continued fraction*

$$G_{xx}(E) = \cfrac{1}{E - a_0 - \cfrac{b_1^2}{E - a_1 - \cfrac{b_2^2}{E - a_2 - \dots}}} \quad (11.10)$$

This object may be unfamiliar to those schooled in the mathematics mostly used in quantum mechanics, but we hope no reader will have difficulty following the simple derivation in Section IV.12. The wider field of mathematics related to (11.3) and (11.10) is developed in Haydock's chapter.¹ Now (11.10) truncated at level N can be simplified algebraically to a sum of rational fractions

$$\sum_s w_s'(E - E_s'), \quad (11.11)$$

¹⁴⁹ See Section 18 of Haydock's chapter.¹

¹⁵⁰ See Sections 25–29 of Kelly's chapter² and references given there.

¹⁵¹ See, for example, S. J. Gale and D. G. Pettifor, *Solid State Commun.* **24**, 175 (1977).

¹⁵² A. M. Papon, M. C. Desjonquères, J. P. Simon, and P. Guyot, *Philos. Mag. B* **39**, 301 (1979).

and when we interpret E as a complex number (Section I,3) $E + iO$ we obtain by Eq. (3.4) the imaginary part of (11.10) as a sum of delta functions

$$n_{\alpha l} = \sum_s w'_s \delta(E - E'_s) \quad (11.12)$$

similar to (11.7). There are several devices for turning this into a smooth continuous function (see Section 8 of Kelly's chapter²). The simplest is to evaluate (11.10) with E having a small but nonzero imaginary part, a little larger than the spacing between the E'_s in (11.11): the effect is to broaden each delta function into a little Lorentzian. The second method is to continue the fraction (11.10) to infinity with all a_n, b_n set equal to suitably extrapolated constant values a, b : this tail can then be summed algebraically. Third, the most sophisticated approach¹⁵³ is Nex's subroutine Supersmooth included in the Cambridge Recursion Method Library¹⁵⁴ of computer programs. It depends on the important convergence property¹⁵³ of the integral $N_{\alpha l}(E)$ of (11.8); for any given E and n , there are upper and lower bounds beyond which $N_{\alpha l}(E)$ cannot stray no matter what one does to the continued fraction (11.10) for levels $n' > n$. Supersmooth takes the average of the two bounds and redifferentiates to give the best estimate of $n_{\alpha l}(E)$ possible from the available information, i.e., from the a, b , down to level n .

Since there are two independent sources of error, what compromise between the finite cluster size and truncation of the continued fraction deploys the computing resources to best advantage? A simple empirical rule of thumb¹⁴⁷ is to make the cluster as big as possible subject to *carrying the recursion to a level $n \simeq 2n_c$ where n_c is the minimum number of nearest-neighbor steps from the central αl needed to get outside the cluster*. The reason is as follows. From (11.3) we can think of $|u_n\rangle$ as essentially $H^n|u_0\rangle$, which in turn can be pictured as the sum of all "paths" of n "steps" from $|\alpha l\rangle = |u_0\rangle$. A step is a hop from one orbital $|\alpha'l'\rangle$ to another $|\alpha''l''\rangle$ connected by a nonzero hopping matrix element in H , and we assume nearest-neighbor steps are the most important. The last two terms of (11.3), the orthogonalizing terms, serve to eliminate certain paths from $|u_n\rangle$, e.g., any that return to $|\alpha l\rangle$ on the way.¹³⁸ For $n > n_c$ some paths would finish outside the cluster if one added further atoms to it, and so the a_n, b_n are affected by the boundary corrections. However for $n \lesssim 2n_c$ most paths will still remain wholly inside the cluster, unaware of its surface, because like in a random walk most of the paths wind around near the origin. For $n \lesssim 2n_c$ one starts to pick up more of the discreteness of $n_{\alpha l}(E)$ for the finite cluster so that $n_{\alpha l}(E)$

¹⁵³ See Section 22 of Haydock's chapter.¹

¹⁵⁴ C. M. M. Nex, Cambridge Recursion Method Library of computer programs in FORTRAN, available from the Secretary, T.C.M. Group, Cavendish Laboratory, Madingley Road, Cambridge, CB3 0HE, England, at a price of £5, \$10, or the equivalent to cover handling costs. The library continues to be improved, the latest version (November 1978) having been standardized to the same notation as in the present volume.

develops spurious spikes¹⁴⁷: after all, if for the cluster one carries on the recursion until b_n is zero for some very high value of n , then $n_{\alpha l}(E)$ will (correctly) be a set of delta functions even with Supersmooth (and, of course, for molecules one may wish to do this). Note that the energy levels of a cluster remain discrete even with periodic or other such boundary conditions. One cannot obtain detail in $n_{\alpha l}(E)$ for the solid, on a finer energy scale than the mean spacing ΔE_n between consecutive energy levels of the cluster. However, it must always be remembered that the errors in integrated quantities such as $N_{\alpha l}(E)$ (11.8) and $U_{\alpha l}$ (11.9) are much smaller than in $n_{\alpha l}(E)$ itself. For further discussion see Section 8 of Kelly's chapter² and Sections 8, 22, and 24 of Haydock's chapter¹; see also our Section IV,14a for how one can in principle estimate the effect on $n_{\alpha l}(E)$ of boundary errors and truncation of the a_n, b_n . The rule to carry the recursion to $n \approx 2n_c$ must be treated as no more than a rule of thumb applicable to many but not all situations. For example, the results in Figs. 16 and 18 were calculated to much higher values of n for the following reason.^{140,141} Recursion to order n achieves an energy/frequency resolution of order W/n , where W is the total bandwidth. In Meek's bond charge model, the bond charges are given a small nonzero mass for calculational purposes, resulting in a spurious band of very high-frequency modes not shown in Figs. 16 and 18 and thus an unusually large W .

Incidentally, to compute an *integrated quantity* such as (11.8) or (11.9), the $n_{\alpha l}(E)$ is best represented as a set of delta functions (11.11), the integral then becoming a simple sum over the E'_s . This "Gaussian quadrature" method¹⁵⁵ makes best use of the information in the a_n, b_n both for definite and indefinite integrals and is computationally faster than numerical integration routines on the continued fraction (11.10). The levels E'_s and weights w'_s are obtained by diagonalizing H from its already tridiagonal form (12.7). These and other wrinkles are incorporated in the Cambridge Recursion Method Library.¹⁵⁴

Figure 19 shows the local density of states for an s band (one s orbital on each atom) near the (001) surface of a simple cubic structure.¹⁴³ The hopping matrix elements are $|h| = 0.06$ units between nearest neighbors, and orbital overlap is neglected. The figure shows well how the local density of states¹⁵⁶ $n_l(E)$ "heals" to its bulk value as we move inwards from the surface. The Friedel oscillations⁸ (Sections I,2 and II,5) can be seen if we focus our attention at some particular E , say $E = 0$ or $E = 0.3$; the $n_l(E)$ is below the bulk value on the surface atom, greater than the bulk value on the next layer, and low again on the third layer. See Sections IV,14a and IV,14c for an outline of the theory of Friedel oscillations in terms of the continued fraction method.

Also in Fig. 19, the absolute bandwidth is the same at the surface as in the bulk because there is nothing to stop any bulk Bloch state reaching the surface: the only question concerns its amplitude there [see Eq. (13.1) of Kelly's

¹⁵⁵ See Section 18 of Haydock's chapter.¹

¹⁵⁶ For an s band with one orbital per site we drop the subscript α in the orbital designation αl .

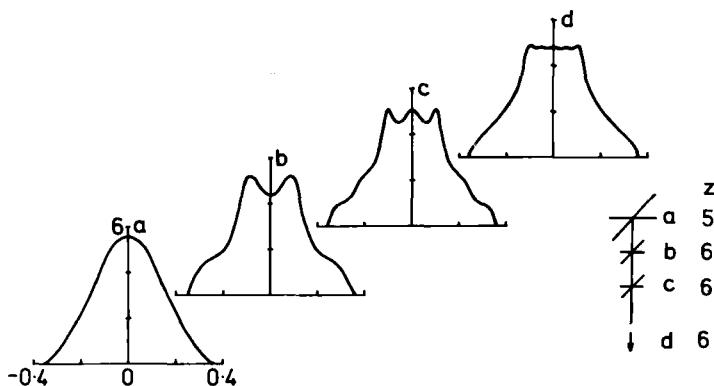


FIG. 19. Local density of states for an s band on a simple cubic structure: (a) for a surface atom, (b) in one and (c) in two layers below the surface, (d) in the bulk (after Haydock and Kelly¹⁴³).

chapter²]. The latter is reflected in the fact that $n_i(E)$ on a surface atom has a smaller root-mean-square width than in the bulk, as seen both in Figs. 17 and 19. Qualitatively this can be regarded as due to a reduction in the number of neighbors (since it is the hopping to neighbors that generates the bandwidth in a tight-binding description), and can be observed experimentally.¹⁵⁷

We can describe these features more precisely in terms of the p th moments¹⁵⁸ $\mu_{\alpha l}^p$ of $n_{\alpha l}(E)$:

$$\mu_{\alpha l}^p = \int_{-\infty}^{\infty} E^p n_{\alpha l}(E) dE. \quad (11.13)$$

In terms of the discrete eigenstates ψ_n (eigenvalues E_n) of a large finite system and their representation Eq. (1.8) in local orbital terms, from Eqs. (1.10) and (1.13) the $\mu_{\alpha l}^p$ becomes

$$\begin{aligned} \mu_{\alpha l}^p &= \sum_n |a_{n, \alpha l}|^2 \int E^p \delta(E - E_n) dE \\ &= \sum_n |a_{n, \alpha l}|^2 E_n^p \\ &= \sum_n \langle \alpha l | n \rangle E_n^p \langle n | \alpha l \rangle \\ &= \langle \alpha l | \left(\sum_n |n\rangle E_n^p \langle n| \right) | \alpha l \rangle \\ &= \langle \alpha l | H^p | \alpha l \rangle, \end{aligned} \quad (11.14)$$

¹⁵⁷ M. Mehta and C. S. Fadley, *Phys. Rev. Lett.* **24**, 1569 (1977); *Phys. Rev. B* **20**, 2280 (1979).

¹⁵⁸ J. Friedel, in "The Physics of Metals: I. Electrons" (J. M. Ziman, ed.), p. 340. Cambridge Univ. Press, London and New York, 1969.

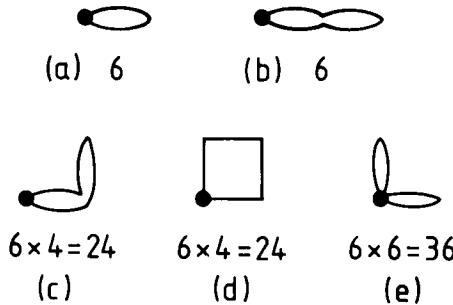


FIG. 20. Closed paths on a simple cubic structure, (a) of length two steps and (b)–(e) four types of length four steps, with the number of distinct paths of each. The last diagram is intended to include three subtypes that go twice out and back, each out and back in each of the six directions. Thus in the bulk there are a total of 90 paths of length four steps, giving $\mu_{al}^4 = 90h^4$ for an s band.

where the last step follows by recognizing that the operator in square brackets is H^p expressed in the eigenstate representation. We can interpret (11.14) in terms of the paths and steps defined earlier in this section. Eq. (11.14) can be evaluated by multiplying out p times the matrix H in the orbital representation:

$$\mu_{al}^p = \sum_{\alpha'l'} \sum_{\alpha''l''} \cdots \langle \alpha l | H | \alpha' l' \rangle \langle \alpha' l' | H | \alpha'' l'' \rangle \langle \alpha'' l'' | H \cdots H | \alpha l \rangle. \quad (11.15)$$

Any term in (11.14) is nonzero only if each factor is nonzero, i.e., represented by a step. Thus μ_{al}^p is a sum (11.15) over all closed paths of p linked steps starting and finishing at αl .¹⁵⁸ Figure 20 shows paths contributing to μ_{al}^2 and μ_{al}^4 in the simple cubic structure. Indeed on any structure for an s band we have¹⁵⁶

$$\mu_l^2 = z_l h^2, \quad \text{s band}, \quad (11.16)$$

and for the root-mean-square (rms) bandwidth

$$W_{\text{rms}, l} = z_l^{1/2} |h|, \quad \text{s band}, \quad (11.17)$$

where z_l is the number of nearest neighbors. The analogous result for a d band (in the notation of Section III,10) is^{85,158a}

$$5(W_{\text{rms}, l})^2 = \sum_m \mu_{ml}^2 = \sum_i z_i [(dd\sigma)^2 + 2(dd\pi)^2 + 2(dd\delta)^2], \quad (11.18)$$

where z_i is the number of neighbors in shell i . The expression in square brackets follows immediately if we take the axis of m quantization in the

^{158a} An earlier reference is F. Ducastelle and F. Cyrot-Lackmann, *J. Phys. Chem. Solids* **31**, 1295 (1970).

direction of a neighboring atom, but since it relates to the band as a whole it must be invariant under choice of axes and hence applicable to all neighbors equally.

Earlier in this section we identified $|u_n\rangle$ and hence a_n , b_n with some open paths of length n from $|\alpha l\rangle$, and we can restate this in terms of closed paths out and back to $|\alpha l\rangle$ of length $2n$, but some paths are excluded as mentioned earlier. The precise relationship is that a_n , b_n are expressible in terms of all moments up to order $2n + 1$ and $2n$, respectively, and conversely (see Section 5 of Kelly's chapter² and references there for details). Indeed the French group¹³² compute the a_n , b_n via the moments and (11.15).

We can now interpret the results in Fig. 19 in terms of moments.¹⁴³ A surface atom has only five nearest neighbors so that its $W_{\text{rms},l}$ is reduced from the bulk value by a factor $(5/6)^{1/2}$. One layer below the surface μ_l^2 and $W_{\text{rms},l}$ already have their bulk value but μ_l^4 is 1/90th less than the bulk value because one closed path of type Fig. 20b does not exist. More generally we can re-express the invariance theorem of Sections I,2 and II,5 as follows: $n_{\alpha l}(E)$ near the surface has all moments up to order $2n_c$ the same as the bulk density of states, where n_c is the minimum number of steps needed to take one from αl up to a surface atom.

We obtain an interesting view of *metallic bonding* from (11.17). The strongest binding clearly occurs when the band is half full, i.e., we have one electron per atom in the s band, counting both spins, which occupies all bonding states and no antibonding ones. Each occupied state has on average an energy of about $W_{\text{rms},l}$ below the center of the band, i.e., below the atomic level. Thus the total bonding energy of a solid is roughly

$$U \simeq - \sum_l z_l^{1/2} |h|. \quad (11.19)$$

We only have one electron per atom so that we cannot form z saturated bonds to the z nearest neighbors, and (11.19) is therefore not proportional to z . Rather we have enough electrons for one bond per atom of energy $-|h|$ per electron, which resonates between z possible positions, and it is this resonance that gives the extra factor $z^{1/2}$ in (11.19). The factor $z^{1/2}$ renders it very difficult to make any general decomposition of metallic binding into a sum of nearest-neighbor, etc., interactions, in contrast to the case of saturated hydrocarbon molecules.

Now consider an ad atom (of the same type) on a (001) surface of the simple cubic structure. The ad atom has $z_l = 1$ neighbor, whereas the underlying surface atom has changed from $z_l = 5$ to 6 neighbors. The binding energy accordingly is

$$B_{001} = [(\sqrt{6} - \sqrt{5}) + \sqrt{1}] |h| = 1.21 |h|. \quad (11.20a)$$

On the other hand, attached in its regular position to a (110) surface, it binds to two surface atoms that otherwise would have four neighbors each. We obtain

$$B_{110} = [2(\sqrt{5} - \sqrt{4}) + \sqrt{2}]|h| = 1.89|h|. \quad (11.20b)$$

Similarly an ad atom is threefold coordinated on the (111) surface, giving

$$B_{111} = [3(\sqrt{4} - \sqrt{3}) + \sqrt{3}]|h| = 2.54|h|. \quad (11.20c)$$

Here again we see that B does not increase as rapidly as the surface coordination of the ad atom. Moreover, the broad generality of the argument—from Eq. (11.15) μ_{al}^2 is always the sum of $|\langle \alpha l | H | \alpha' l' \rangle|^2$ over nearest neighbors as in (11.18)—gives it a wider validity as a general feature of metallic binding not related specifically to s bands. The result is in agreement with experiment on transition metals.¹⁵⁹ Finally, consider a dimer molecule of the metal: it has a dissociation energy of $2|h|$: but placed against a (010) step on a (001) surface, the dissociation energy is only

$$D = 2(\sqrt{3} - \sqrt{2})|h| = 0.64|h|, \quad (11.21)$$

which is a very large reduction. Indeed it is well known that many molecules dissociate on metal surfaces.

Of course these sums based on (11.19) must not be taken too literally: rather they indicate qualitatively the features one might expect from proper calculations. Thus they illustrate the following general feature of metallic covalent bonds. In an LCAO description, the energy of a “molecular” orbital, e.g., the band structure $E(\mathbf{k})$, is given in terms of the bonding or anti-bonding relationship between neighboring atomic orbitals, exactly as with the hydrogen molecule H_2 or the tetrahedral carbon atom. The difference in close-packed metals is that one has many more nearest neighbors than the number of complete bonds that can be formed from the available electrons, and so the bonds resonate between the various sites and one loses the directional and saturation effects associated with forming the closed shell in the tetrahedral carbon atom, etc. For example, the dissociation energy D (11.21) is so low because in a sense one has not destroyed any bond: the two ad atoms when separated are still bound to the step, but in breaking the direct link between them one is merely reducing the number of positions that the bonding electrons can resonate between. Incidentally it should be possible to deduce the factor $z^{1/2}$ in (11.19) from such a combinatorial point of view but this has so far not been done.

The expression for U in (11.9) of course only sums the occupied one-electron energies. It omits other electrostatic terms, as we discuss in Section

¹⁵⁹ E. W. Plummer and T. N. Rhodin, *J. Chem. Phys.* **49**, 3479 (1968).

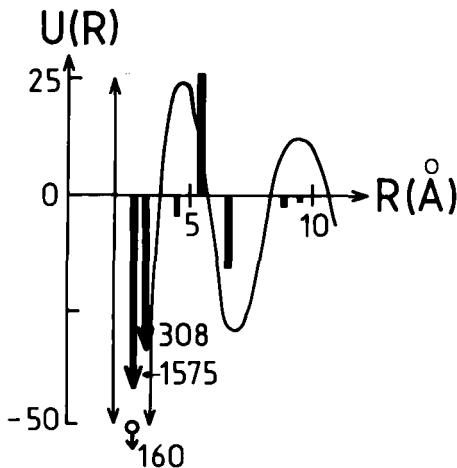


FIG. 21. Variation with distance R of the interaction energy $U(R)$ for two rhenium atoms on a tungsten 110 surface. Vertical bars: calculated results taken from Burke.¹⁶¹ These first calculations without self-consistency and sp electrons should be taken as illustrative of what can be done rather than as definitive. Continuous curve: deduced from experiment, after Tsong,¹⁶² but note that there is some uncertainty about the reproducibility and interpretation of the experimental results.^{161,162}

IV,13, connected with the electron distribution between the atoms, but can be used successfully in comparing *the energies of two similar situations* where the electrostatic terms will be nearly enough equal to effectively cancel out. Such is the case in comparing the fcc and hcp structures of transition metals¹⁶⁰ or the three Laves structures for alloys,⁴ where in each case the nearest-neighbor coordination is nearly identical (see Sections 25 and 26 of Kelly's chapter² and figures there). The same is true of a recent calculation of the intrinsic and extrinsic stacking fault energies and twinning energies of close-packed transition metals.¹⁵² Figure 21^{161,162} shows a similar calculation¹⁶¹ (carried out with the technique described in Section IV,14c) for the interaction energy of two rhenium atoms on a (110) surface of tungsten. At the closest the two atoms occupy next-nearest-neighbor positions with some direct interaction, but the purpose of the calculation lay in the interaction at larger separations. These early results for a d band only give remarkably small energies falling off rapidly with distance, so that one may conclude this

¹⁶⁰ D. G. Pettifor, *J. Phys. C* **3**, 367 (1970).

¹⁶¹ N. R. Burke, *Surf. Sci.* **58**, 349 (1976); private communication; also thesis presented to the University of Cambridge, 1976.

¹⁶² T. T. Tsong, *Phys. Rev. Lett.* **31**, 1207 (1973); but see also D. W. Bassett, *Surf. Sci.* **53**, 74 (1975).

is presumably not the dominant mechanism. However, note the expected Friedel-type oscillations (see Section 14c).

The point was emphasized in Section I,1 that the local point of view requires that *any desired physical property be expressed in terms of the Green function*. In the recursion method we can in Eq. (11.1) calculate the local density of states projected onto any desired orbital $|\chi\rangle$. An example is a “surface group orbital,” which is the linear combination of atomic orbitals in the surface with the right symmetry to bond to a given atomic orbital on an ad atom adsorbed on a given site (see Section 16 of Kelly’s chapter²). As already mentioned, the recursion method is not limited to electronic structure: it has, for instance, been used to calculate spin waves in a dilute disordered magnetic alloy¹⁶ with the interaction between spins described by a Heisenberg model. These and other applications are covered in Kelly’s chapter.²

It may be that an *off-diagonal* part of the Green function $G_{ab}(E)$ is required, where $|a\rangle$ and $|b\rangle$ are arbitrary state vectors defined in terms of the $|\alpha l\rangle$:

$$|a\rangle = \sum_{\alpha l} a_{\alpha l} |\alpha l\rangle, \quad |b\rangle = \sum_{\alpha l} b_{\alpha l} |\alpha l\rangle. \quad (11.22)$$

The recursion procedure is then not directly applicable.^{14,5} We must first define combined state vectors (see also Section 17 of Haydock’s chapter¹)

$$\begin{aligned} |\chi_1\rangle &= 2^{-1/2}[|a\rangle + |b\rangle], & |\chi_2\rangle &= 2^{-1/2}[|a\rangle - |b\rangle], \\ |\chi_3\rangle &= 2^{-1/2}[|a\rangle + i|b\rangle], & |\chi_4\rangle &= 2^{-1/2}[|a\rangle - i|b\rangle], \end{aligned} \quad (11.23)$$

and then have

$$G_{ab}(E) + G_{ba}(E) = \langle \chi_1 | (E - H)^{-1} | \chi_1 \rangle - \langle \chi_2 | (E - H)^{-1} | \chi_2 \rangle, \quad (11.24a)$$

$$i[G_{ab}(E) - G_{ba}(E)] = \langle \chi_3 | (E - H)^{-1} | \chi_3 \rangle - \langle \chi_4 | (E - H)^{-1} | \chi_4 \rangle, \quad (11.24b)$$

where each term can be calculated by the recursion method as per (11.1) and (11.10). In many cases (11.24a) suffices for what is required, e.g., if the orbitals are real. An example is the calculation of bond orders, i.e., the degree of bonding, between neighboring atoms, which is proportional to the imaginary part of $G_{ab}(E)$ (see Section 3 of Kelly’s chapter²).

As a final more sophisticated example we discuss how optical absorption and photoemission^{13,14,16³}

^{16³ V. Heine, *Jpn. J. Appl. Phys., Suppl.* **2**, Pt. 2, 679 (1974).}

final state $|f\rangle$ of energy E , picked out (say) by angular resolved photo-emission, is proportional to

$$\sum_i \langle f | \mathbf{A} \cdot \mathbf{p} | i \rangle \langle i | \mathbf{A} \cdot \mathbf{p} | f \rangle \delta(E - h\nu - E_i), \quad (11.25)$$

where for simplicity of presentation we have dropped the niceties about the difference between $\mathbf{A} \cdot \mathbf{p}$ and $\mathbf{p} \cdot \mathbf{A}$ and with it the specifically surface absorption term.¹⁶⁴ Here $|i\rangle$ is an occupied initial eigenstate of energy E_i and assumed expressible in LCAO form in terms of the atomic orbitals $|\alpha l\rangle$. We wish to express (11.25) in Green function form. The unit operator

$$\sum_{\alpha l} |\alpha l\rangle \langle \alpha l| \quad (11.26)$$

is introduced twice into (11.25), which becomes

$$\text{Im} \sum_i \sum_{\alpha l} \sum_{\alpha' l'} \langle f | \mathbf{A} \cdot \mathbf{p} | \alpha' l' \rangle \langle \alpha' l' | \left[\frac{|i\rangle \langle i|}{E - h\nu - E_i} \right] |\alpha l\rangle \langle \alpha l | \mathbf{A} \cdot \mathbf{p} | f \rangle. \quad (11.27)$$

Here the quantity in square brackets is the operator

$$(E - h\nu - H)^{-1} \quad (11.28)$$

where E as usual is interpreted as $E + iO$ (Section I,3). The final state is pictured as a single plane wave $|\mathbf{k}\rangle = \exp(i\mathbf{k} \cdot \mathbf{r})$, or rather the pseudowave function corresponding to it is: but there is no difficulty including some final-state scattering^{13,14} among a few other plane waves. In the last factor of (11.27), \mathbf{A} is constant in the dipole approximation, and \mathbf{p} operating on a plane wave produces simply $\hbar\mathbf{k}$, so that we have

$$\langle \alpha l | \mathbf{A} \cdot \mathbf{p} | f \rangle = \hbar \mathbf{A} \cdot \mathbf{k} \langle \alpha l | \mathbf{k} \rangle = \hbar \mathbf{A} \cdot \mathbf{k} f_{\alpha}(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{R}_l). \quad (11.29)$$

Here f_{α} depends on the nature of the orbital $\phi_{\alpha l}$ at \mathbf{R}_l and there is no difficulty including s, p, d orbitals simultaneously, e.g., for a transition metal, with perhaps an adsorbed overlayer.¹⁴ With final-state scattering included in $|f\rangle$ we obtain the sum of several terms like (11.29), and can even add an imaginary part to \mathbf{k} corresponding to a finite mean free path or escape length.¹³ In any case we can write

$$\langle \alpha l | \mathbf{A} \cdot \mathbf{p} | f \rangle = a_{\alpha l}, \quad (11.30)$$

and now define

$$|\chi\rangle = \sum_{\alpha l} a_{\alpha l} |\alpha l\rangle. \quad (11.31)$$

¹⁶⁴ See, for example, McClean and Haydock;¹³ W. L. Schaich and N. W. Ashcroft, *Phys. Rev. B* **3**, 2452 (1971).

With (11.28), (11.30), and (11.31), our expression (11.27) becomes

$$\text{Im}\langle\chi|(E - h\nu - H)^{-1}|\chi\rangle, \quad (11.32)$$

so that we take $|\chi\rangle$ (11.31) as the starting vector in the recursion method and proceed as usual. The hole in the initial state also has a finite lifetime τ , so that we take E as having a finite imaginary part instead of the usual infinitesimal quantity. The effect is to wash out fine structure in $G_{xx}(E - h\nu)$ corresponding to deeper levels a_n, b_n in the continued fraction (11.10), which in turn arise from the more distant basis functions $|n\rangle$. Thus the optical properties are seen as sampling the local electronic structure to some degree^{13,163} although the most elementary calculation of momentum matrix elements for Bloch states involves the whole crystal.

In conclusion, therefore, we have shown how *physical quantities* may be expressed as *diagonal matrix elements of the Green function*. These may be calculated by the *recursion method* on a *finite cluster* of atoms with no need for symmetry. If there are sufficient atoms in the cluster surrounding the central ones of prime interest, the finiteness of the cluster should not cause significant errors.

12. THE HAYDOCK RECURSION SCHEME

We now describe the recursion method^{1,2,138,139} for calculating the local density of states $n_{\alpha l}(E)$ in relation to a set of atomic orbitals $\phi_{\alpha l}(\mathbf{r})$ or atomic vibrational displacements or other local basis sets for the problem at hand. The method involves setting up a new orthonormal basis set $u_n, n = 0, 1, 2, \dots$, the first of which u_0 we choose as the particular orbital $\phi_{\alpha l}$ or linear combination of orbitals for which we want the local density of states for (Section IV,11). We handle it as a column matrix (column vector) of coefficients and indeed express all the u_n in this way:

$$u_n = \sum_{\alpha' l'} u_{n, \alpha' l'} \phi_{\alpha' l'}(\mathbf{r}). \quad (12.1)$$

Thus u_0 would often be the vector

$$\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftarrow \alpha l, \quad (12.2)$$

representing one orbital on one atom, but, as emphasized in Section IV, 11, any starting vector u_0 can be chosen in accordance with the purpose at hand. The Hamiltonian similarly is expressed as the matrix $[H_{\alpha l, \alpha' l'}]$ of hopping matrix elements, etc., between the $\phi_{\alpha l}(\mathbf{r})$.

Having chosen u_0 , we define u_1 as

$$b_1 u_1 = H u_0 - a_0 u_0 \quad (12.3)$$

and similarly generate the whole set u_n by the recurrence relation

$$b_{n+1} u_{n+1} = H u_n - a_n u_n - b_n u_{n-1}, \quad (12.4)$$

of which (12.3) is the starter. Here a_n , b_n are coefficients to orthogonalize $H u_n$ to the preceding vectors u_n , u_{n-1} , and b_{n+1} is the coefficient to normalize u_{n+1} to unity. Thus aside from the orthogonalization, *the new basis set u_n is defined by successive operation with H* . If u_0 is localized on $\phi_{\alpha l}$, then the process can be pictured perhaps like a radial outward integration of the Schrödinger equation although no specific energy E is assumed. Let us suppose that H only involves hopping to nearest neighbors, although there is no such restriction in the method. Then the first application of H in (12.3) creates an orbital u_1 located on the first shell of neighbors of $u_0 = \phi_{\alpha l}$, and so on through (12.4) so that u_n reaches out to the n 'th shell of neighbors, as already shown in Fig. 15. By orthogonality all the u_n ($n > 0$) have zero weight on the initial orbital $u_0 = \phi_{\alpha l}$ and in the case of a nearest-neighbor s-band model also on the nearest neighbors of the central site, but they are not otherwise generally zero inside their domain. In any case u_n represents a basis for solutions of the Schrödinger equation in regions increasingly remote from αl , yielding coefficients a_n , b_n becoming less and less significant in determining the local density of states $n_{\alpha l}(E)$ as we go further down the continued fraction (11.10).

Incidentally to iterate (12.4) one only needs to store the previous two vectors u_n , u_{n-1} , all earlier ones being discarded. With typically of the order of 10^4 orbitals $\phi_{\alpha l}$ (and hence 10^4 elements in each column vector $u_{r, \alpha' l'}$) and iteration to $n = 30$ in the continued fraction, this is not a trivial point in the amount of computer store required. Timewise the rate-determining step of the calculation lies in the subprogram for computing $H u_n$. Of course, one does not store H as a $10^4 \times 10^4$ matrix consisting of about 99.9% zeros, since H operating on one orbital only produces a result on nearest-neighbor orbitals (and next nearest neighbors, etc., as desired). For an amorphous cluster (and often for other cases too) it may be advantageous to number all atoms and store a list ("map") of which ones are the neighbors of each. This is one method included in the Cambridge Recursion Method Library.¹⁵⁴ If one stores the coordinates of the neighbors, the matrix elements of H may be calculated and recalculated from a simple algorithm as they are used. On

the other hand, in dealing with a crystallite one may store all matrix elements for one unit cell and then apply them for each cell in the cluster. It would save time in the initial stages to use row rather than column multiplication with a switch so that only nonzero elements in u_n are multiplied by H explicitly.

The essential information needed from the recursion process to calculate the local density of states is *only the set of coefficients a_n , b_n* . Taking the matrix element of (12.4) with u_n , u_{n-1} , etc., we obtain their evaluation:

$$a_n = \langle u_n | H | u_n \rangle, \quad (12.5a)$$

$$b_b = \langle u_{n-1} | H | u_n \rangle. \quad (12.5b)$$

The best numerical algorithms for evaluating a_n , b_n , i.e., the most stable and least susceptible to rounding errors, are discussed by Haydock (Section 7)¹ and contained in the Cambridge Recursion Method Library.¹⁵⁴ Formulas (12.5) apply only for an orthonormal basis set of ϕ_{al} (to which we restrict ourselves in Sections IV,11–14) and must be modified in the more general case of nonorthogonal orbitals.^{145,165} In that case the recursion (12.4) proceeds as before, using the matrix H in the sense of Eqs. (7.4) and (7.8) and Section 9, not the \mathcal{H} matrix. No matrix inversion is involved because the methods of Sections III,9 and IV,10 give H directly, and the modified formulas¹⁶⁵ for a_n , b_n involve only H and S , not S^{-1} .

In the new basis u_n the Hamiltonian H takes on a particularly simple form. From (12.5) and (12.4) we have the matrix elements

$$\langle u_n | H | u_n \rangle = a_n, \quad (12.6a)$$

$$\langle u_{n-1} | H | u_n \rangle = \langle u_n | H | u_{n-1} \rangle = b_n, \quad (12.6b)$$

$$\langle u_n | H | u_m \rangle = 0 \text{ otherwise.} \quad (12.6c)$$

Thus H has the *tridiagonal form*

$$\begin{bmatrix} a_0 & b_1 & & & & & & & \\ b_1 & a_1 & b_2 & & & & & & \\ & b_2 & a_2 & b_3 & & & & & \\ & & b_3 & a_3 & b_4 & & & & \\ & & & - & - & - & - & & \\ & & & & - & - & - & - & \end{bmatrix} \quad (12.7)$$

where the matrix is zero apart from the diagonal and the line of elements immediately to either side of it. It is now easy to calculate

$$G_{al,al}(E) = \langle u_0 | [E - H]^{-1} | u_0 \rangle, \quad (12.8)$$

¹⁶⁵ See Section 5 of Haydock's chapter,¹ particularly Eqs. (5.10) and (5.12).

which we need for $n_{ai}(E)$, (1.11), from the matrix

$$[E - H] = \begin{vmatrix} E - a_0 & -b_1 & \cdot & \cdot & D_0 \\ -b_1 & \begin{vmatrix} E - a_1 & -b_2 & \cdot & D_1 \\ -b_2 & \begin{vmatrix} E - a_2 & \cdot & -b_3 & D_2 \\ \cdot & \cdot & \ddots & \end{vmatrix} & \end{vmatrix} & (12.9) \\ \cdot & \cdot & \ddots & \end{vmatrix}$$

Here E is a general complex number, which in (1.11) and (12.8) we later take with a small positive imaginary part. Note that we do not want the whole of the inverse matrix $[E - H]^{-1}$, only the one element (12.8), and on this the whole method depends. In physics we do not usually inquire about the whole of life, but about one particular matrix element (at a time) and we have chosen u_0 so that (12.8) gives us what we want (see Section IV,11). The leading element (12.8) of $[E - H]^{-1}$ is given in the usual way as “cofactor divided by determinant,” i.e.,

$$\langle u_0 | [E - H]^{-1} | u_0 \rangle = \det|D_1|/\det|D_0|, \quad (12.10a)$$

which we write as

$$\langle u_0 | [E - H]^{-1} | u_0 \rangle = \frac{1}{\det|D_0|/\det|D_1|}, \quad (12.10b)$$

where D_0 is the whole determinant of $E - H$ and D_1 similarly with the first row and column removed as shown in (12.9). From the Cauchy expansion¹⁶⁶ of a determinant we have

$$\det|D_0| = (E - a_0) \det|D_1| - b_1^2 \det|D_2|, \quad (12.11)$$

so that (12.10) can be written

$$\langle u_0 | [E - H]^{-1} | u_0 \rangle = \frac{1}{E - a_0 - (b_1^2/\det|D_1|/\det|D_2|)}. \quad (12.12)$$

The analogous relation to (12.11) holds for $|D_n|$, $|D_{n+1}|$, and $|D_{n+2}|$ for all n , so that we have

$$\frac{\det|D_n|}{\det|D_{n+1}|} = E - a_n - \frac{b_{n+1}^2}{\det|D_{n+1}|/\det|D_{n+2}|}, \quad (12.13)$$

¹⁶⁶ A. C. Aitken, “Determinants and Matrices,” 6th ed., Chapter IV. Oliver & Boyd, Edinburgh, 1949.

¹⁶⁷ L. M. Mattheiss, *Phys. Rev.* **133**, A1399 (1964).

and the continued fraction (11.10) for the Green function follows immediately from (12.12) and (12.13) by iteration, which completes the main purpose of this section. Thus once the a_n , b_n have been determined, the continued fraction (11.10) can be evaluated very rapidly by iteration of (12.13) for any number of values of E required (see Sections 17, 18, and 22 of Haydock's chapter¹).

It remains to tidy up some mathematical details about Eq. (12.4) that we have so far skirted over. The a_n in (12.4) is computed from (12.5a), but it is not obvious that the coefficient b_n in (12.4) should be the same as the normalizing factor in the preceding relation

$$b_n u_n = H u_{n-1} - a_{n-1} u_{n-1} - b_{n-1} u_{n-2}. \quad (12.14)$$

Incidentally, the normalization process defining the bs only determines $|b_n|^2$, and this is all that enters the continued fraction (11.10), so that there is no loss of generality in assuming b_n real (and positive), as already assumed in (12.5b). Taking the matrix element of (12.14) with $\langle u_n |$ we obtain

$$b_n = \langle u_n | H | u_{n-1} \rangle, \quad (12.15a)$$

whereas orthogonality to $\langle u_{n-1} |$ in (12.4) requires

$$b_n = \langle u_{n-1} | H | u_n \rangle, \quad (12.15b)$$

which is the same as (12.15a) in view of the reality of b_n and hermiticity of H .

The other point is to prove that u_{n+1} defined by (12.4) is automatically orthogonal to u_{n-2} , u_{n-3} , and all earlier u_{n-r} , ($r > 2$) without the need to include corresponding further orthogonalizing terms in the recursion relation (12.4). It follows from the way the u_n are defined. Suppose we formally included orthogonalizing terms to all previous u_{n-r} in (12.4): the coefficients would be equal to $\langle u_{n-r} | H | u_n \rangle$. This is the same as $\langle u_n | H | u_{n-r} \rangle$, which is the expansion coefficient of u_n when $H u_{n-r}$ is expanded in terms of the u_m . But u_{n-r+1} is so defined that it "absorbs" all of $H u_{n-r}$, that is orthogonal to the earlier vectors, and hence $H u_{n-r}$ does not involve higher vectors, in particular u_n , i.e., $\langle u_n | H | u_{n-r} \rangle$ is zero, which completes the proof. Note that no restriction of $H_{\alpha l, \alpha' l'}$ to nearest neighbors or otherwise (except hermiticity) is involved. For the case of nonorthogonal orbitals and consequently nonhermitian H (7.8), see footnote 165 as aforementioned.

Do the new basis functions u_n form a complete set? It can be shown¹ that they do as far as anything involving the starting orbital u_0 is concerned, i.e., $\phi_{\alpha l}$ in our example. However, it is also clear from the way the u_n are constructed by successive operation with H that all the u_n have the same symmetry as u_0 , i.e., they transform according to the same base vector of the same irreducible representation of the symmetry group of the complete H about the site of u_0 .

Thus the set $\{u_n\}$ does not contain functions of different symmetry, which must be constructed from different initial orbitals.

13. TOTAL ENERGY CALCULATIONS

In a chapter primarily devoted to electronic structure from a local point of view, the discussion of total energy has unfortunately become dismembered. Points concerning the exchange and correlation energy have already been made in Section III,8. This section develops the conventional wisdom about the total energy of a system, while Section V,16 is devoted to a genuine local formulation of the pressure–volume relation, which can of course be integrated to give the total energy or (in some cases) differentiated to give elastic and other force constants. In particular, we shall see in Section V,16c that the local force theorem gives an extra degree of justification for the use of the one-electron energy U_1 . See also Section IV,14c for a formulation of coupling energies in the recursion method.

For the present, the main issue is the separation of the total energy U of the system into the band structure or *one-electron energy* U_1 (sum of the one-electron eigenvalues E_n), and the so-called *electrostatic term* U_{es} :

$$U = U_1 + U_{es} \quad (13.1)$$

$$U_1 = \sum_n E_n \quad (13.2)$$

$$U_{es} = \frac{1}{2} \sum_{i \neq i'} \sum_{l l'} \frac{Z_i Z_{i'}}{R_{ll'}} - \frac{1}{2} \left[\iiint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3\mathbf{r} d^3\mathbf{r}' + U_{xc} \right]. \quad (13.3)$$

Here the first term in U_{es} is the Coulomb repulsion²⁰ of the nuclei or ion cores of the atoms, where as usual the factor $\frac{1}{2}$ corrects for the fact that $\sum_l \sum_{l'}$ counts each pair of atoms twice. Similarly, the second term of (13.3) is the Coulomb energy of the electron charge density $\rho(\mathbf{r})$, while the exchange and correlation energy U_{xc} has already been discussed in connection with Eq. (8.9). The derivation of (13.1)–(13.3) varies according to the treatment of the many-body aspect of the electron gas by the Hartree or Hartree–Fock equations, density functional theory, Landau quasi-particle formulation, etc., but the form of the result is always the same.

The name *electrostatic energy* for (13.3) is in one sense a misnomer. It contains the ion–ion Coulomb energy, but has the electron–electron energy with a minus sign, while the ion–electron Coulomb attraction goes into the self-consistent field $V(\mathbf{r})$ in the one-electron Schrödinger equation and hence is included in the eigenvalues E_n and thus in U_1 . The electron–electron Coulomb energy between pairs of electrons i and j is counted twice in U_1 , once

through the $V(\mathbf{r}_i)$ felt by electron i and once through $V(\mathbf{r}_j)$ felt by electron j , which is why it has to be subtracted once in U_{es} . Thus U_{es} is *not* the electrostatic interaction of the two charge clouds of two atoms (ions plus valence electrons).

Let us review the qualitative forms of U_1 and U_{es} , starting with the latter. U_{es} is essentially a short-range repulsive interaction. Consider just a pair of atoms l, l' and let us assume for a moment the total electron density ρ is just the superposition of those for isolated free atoms ρ_a :

$$\rho = \rho_{al} + \rho_{al'}, \quad (13.4)$$

as is commonly assumed in the Mattheiss¹⁶⁷ prescription for calculating the total potential in a solid. Then the middle term of (13.3) can be split up into

$$\rho\rho = \rho_{al}\rho_{al'} + \frac{1}{2}\rho_{al}\rho_{al} + \frac{1}{2}\rho_{al'}\rho_{al'}, \quad (13.5)$$

of which the first part gives an interatomic electrostatic energy and the remainder is purely intra-atomic, i.e., a constant independent of the separation R between l and l' . The U_{xc} makes only a small interatomic contribution, which we neglect: see also Section III,8. Thus the interatomic part of U_{es} is the Coulomb interaction between the two positive ions, minus the Coulomb interaction between the free-atom valence charge clouds. The two exactly cancel at large R because nonoverlapping spherical charge distributions can be represented by point charges. At smaller R the ion-ion energy will always outweigh the valence-valence one because the latter charge clouds are more spread out. We therefore reach our main conclusion: U_{es} gives a monotonic repulsive interatomic interaction.¹⁶⁸ At short distances the valence-valence energy tends to a constant of order $Z_l Z_{l'}/R_a$, where R_a is an atomic radius, whereas the ion-ion part diverges as $Z_l Z_{l'}/R$. For a diatomic molecule one can easily see that the latter is the dominant term in the total energy at small R . For example, if we bring two hydrogen atoms together to form a helium atom, all contributions to the electronic energy remain quite finite while the proton-proton repulsion diverges.

We believe this formulation of U_{es} as a short-range repulsion is correct and appropriate for considering elastic shear constants, phonons, defects, etc. However, a different analysis may be more appropriate in one situation, viz., the uniform compression of a close-packed solid, where atomic overlap is already large. The solid is divided into Wigner-Seitz polyhedra, nonoverlapping and to a good approximation spherical. Then the ion-ion and valence-valence parts of the interatomic U_{es} cancel exactly, leaving the intraatomic valence-valence Coulomb energy and U_{xc} . These are of order

¹⁶⁸ B. Gyorffy and W. Pickett, in "Superconductivity in d and f Band Metals: Report of Rochester Conference" (D. Douglas, ed). Am. Inst. Phys., New York, 1977.

Z_i^2/R_{ws} and Z_i/R_{xc} , respectively, where R_{ws} is the radius of the Wigner-Seitz sphere and R_{xc} that of the exchange and correlation hole. For a polyvalent element the former dominates so that U_{es} behaves like Z_i^2/R_{ws} at high compression, similar to our previous analysis but considerably weaker because the previous $Z_i Z_i / R$ applied to each pair of neighbors. As already discussed in Section III,8, for the case of monovalent atoms the intra-atomic valence-valence energy to a good approximation cancels U_{xc} , leaving U_{es} in this approximation totally zero. We conclude that *for uniform compression* of a close-packed solid, the ion-ion repulsion in U_{es} is much more effectively screened by the valence electrons since the latter are confined to the Wigner-Seitz polyhedron unable to spill out freely as around a diatomic molecule. Bullett⁸¹ has extended this analysis for monovalent metals to the tetrahedrally directed bonds in the diamond structure, as mentioned in Sections III,8 and III,9. The carbon, silicon, etc., atom is divided into four more or less nonoverlapping and independent valence bond orbitals, each containing one electron so that the same logic as before can be applied to it, at least as a zero-order approximation to which Bullett then added a small correction,¹⁰⁰ as seen in Table I.

It is, of course, always good to arrange calculations so that the quantities least well known cancel to a large degree. In fact, one very often is interested physically in an *energy difference* or can set up the calculation that way. For example, in calculating, say, the heat of adsorption of an oxygen atom onto a transition metal surface, there would be uncertainty about the intra-atomic Coulomb energy and U_{xc} , and their change as a result of bonding to the surface. Thus Bullett¹¹⁶ considered two oxygen atoms, first bonded to one another and then bonded to the metal. The aforementioned terms then cancel out of this energy difference to some degree of approximation, probably to a greater degree than one can individually calculate them without very elaborate and long computations. If one finally wants the adsorption energy relative to free-oxygen atoms, one has only to add the measured dissociation energy of the free-oxygen molecule, which is a matter of experimental molecular chemistry and not of surface science. A similar logic applies to some of the other energy calculations already mentioned in Section IV,11. The charge densities in the fcc and hcp structures will be very similar because of the same number of nearest neighbors arranged in nearly the same way at the same distance. Thus it seemed a good approximation to take U_{es} as equal in the fcc and hcp structures in calculating which was the preferred form for the transition elements.¹⁶⁰ The same applies to the three Laves phase structures for alloys,⁴ and to moving two atoms around on a surface to calculate their interaction.¹⁶¹ However, in comparing the fcc and bcc structures there is a change of interatomic distance, and a correction for U_{es} had to be applied semi-empirically¹⁶⁰ (see Fig. 23 of Bullett's chapter¹⁵). The latter is a

guide to the importance of the effects. We conclude that *we can ignore U_{es} when calculating on monovalent or quasi-monovalent systems or when comparing energies in two structures with the same interatomic spacing.* See also Section V,16c.

Unfortunately there does not appear to be any thorough discussion in the literature of the reliable calculation of the interatomic part of U_{es} when one needs it explicitly, particularly the calculation of the interatomic valence-valence Coulomb energy. There is no difficulty if one adopts the approximation (13.5) of overlapping spherical charge distributions: one first calculates the potential due to ρ_i and then its interaction with $\rho_{l'}$ by a two-dimensional integral in spheroidal coordinates. If, however, one wants the interaction between odd-shaped Wigner-Seitz cells, the mind turns to multipole expansions.¹⁶⁰ It would not appear to be valid simply to replace each cell by its usual multipole moments, because a multipole expansion only represents the potential outside a sphere centered at the expansion point and wholly enclosing the replaced charge distribution.¹⁶⁹ It is not clear how large the error would be (it might be negligible), but in any case one might be able to do the following. Replace ρ_i by a multipole expansion not about the ionic position \mathbf{R}_i but about a point \mathbf{R}_∞ essentially at infinity. Since hopefully the plane boundary between the cells can be considered a sphere centered at \mathbf{R}_∞ , the multipole expansion produces the correct potential at all points in cell i' . The energy of this system can then be calculated by replacing $\rho_{l'}$ by a multipole expansion at $\mathbf{R}_{l'}$, since there is no difficulty enclosing $\rho_{l'}$ by a sphere excluding the multipoles at \mathbf{R}_∞ . Finally the multipoles at \mathbf{R}_∞ can be represented by the second type of multipole expansion¹⁶⁹ valid inside a sphere centered at $\mathbf{R}_{l'}$ wholly excluding the multipoles at \mathbf{R}_∞ but with radius large enough to be valid at $\mathbf{R}_{l'}$. It is hardly surprising that in studies of phonon spectra of transition metals the U_{es} is most often represented by an empirically adjusted (repulsive) interaction.¹⁷⁰

Another viewpoint on U_{es} is of interest in developing a theory of elastic shear constants or phonons.¹⁷³ Suppose the valence charge density $\rho(\mathbf{r})$ in a transition metal may be idealized as spherically symmetric around each atomic nucleus but tending to a constant ρ_0 at the edge of the cell. Suppose

¹⁶⁹ W. K. H. Panofsky and M. Phillips, "Classical Electricity and Magnetism," Eq. (5.21). Addison-Wesley, Reading, Massachusetts, 1955.

¹⁷⁰ See, for example, Varma and Weber¹⁷¹ and Gale *et al.*,¹⁷² also Sections 25-29 of Kelly's Chapter.²

¹⁷¹ C. M. Varma and W. Weber, *Phys. Rev. Lett.* **39**, 1094 (1977).

¹⁷² S. J. Gale, V. Heine, D. G. Pettifor, D. M. Roy, and A. M. B. G. de Vallera, in "Transition Metals: Proceedings of the Toronto Conference" (M. J. G. Lee, J. M. Perz, and E. Fawcett, eds.), Inst. Phys. Conf. Ser., Vol. 39, p. 161. Inst. Phys. Press, Bristol, 1978.

¹⁷³ I am indebted to Prof. O. K. Andersen (private communication) for this discussion.

further that in a deformation of the metal the spherically symmetric part of $\rho(\mathbf{r})$ moves rigidly with the atomic nucleus and that over the cell boundary $\rho(\mathbf{r})$ remains equal to ρ_0 as before. The change in U_{es} is very similar to the change in the Ewald energy of positive point charges in a uniform negatively charged medium, though it is not clear whether the Ewald technique can be adapted to U_{es} since the latter is not the electrostatic energy of any charge system. Anyway the Ewald-like energy would relate to charges of magnitude $\Omega\rho_0$, where Ω is the atomic volume, because the intercell energy of the main spherically symmetric part of $\rho(\mathbf{r})$ would largely cancel the interaction of the ionic point charges Z . By contrast, in nearly free-electron metals³ one starts with bare phonons corresponding to the Ewald energy with charges Z (which are then screened), and a similar approach in transition metals with Z up to 10 would give very high and unphysical bare frequencies. It is therefore gratifying to see our analysis leading to bare phonons corresponding to charges $\Omega\rho_0$, which is of the order of one or two electronic charges¹⁷³ throughout the transition metal series. See also Section V,16.

We now turn to discuss U_1 in (13.2). U_1 contains (almost) all the interesting structure-dependent part of the energy, as demonstrated most dramatically by Varma and Weber's calculation¹⁷¹ of the phonon spectrum of niobium with all its anomalies. But the point had been evident earlier from the theory of the crystal structures of the transition metals¹⁷⁴ and the electron concentration effects in Laves phase alloys,⁴ as well as the success¹⁷⁵ of the Hückel theory⁶⁹ for conformational energy differences in organic molecules.

In the terms of Section IV,11, U_1 is calculated as

$$U_1 = \sum_l U_{1l}, \quad (13.6)$$

$$U_{1l} = \sum_\alpha \int_{-\infty}^{E_F} E n_{\alpha l}(E) dE, \quad (13.7)$$

where

$$Z_l = \sum_\alpha \int_{-\infty}^{E_F} n_{\alpha l}(E) dE \quad (13.8)$$

determines the Fermi level E_F . If all atoms are equivalent the calculation of (13.6)–(13.8) is straightforward enough by the methods of Section IV,11. However, if all atoms are structurally or chemically not equivalent then there can be *charge transfers* between them, setting up additional potentials that have to be calculated self-consistently. Let their effect be extra diagonal energy terms ΔE_l on the atoms. Now because one atomic unit²⁰ of energy

¹⁷⁴ R. A. Deegan, *J. Phys. C* 1, 763 (1968); Pettifor.¹⁶⁰

¹⁷⁵ I am indebted to Dr. J. C. Phillips for this information.

(27.2 eV) is rather large compared with most bandwidths, a correspondingly small amount of charge transfer sets up quite substantial potentials ΔE_l . Hence in (13.8) we have taken ΔE_l to be determined by the self-consistency condition such that charge transfers are effectively zero to a good approximation, i.e., ΔE_l has to be adjusted to keep atom l electrically neutral.

A further transformation is often useful when calculating *energy differences between two structures A and B*, for example, the same transition metal in the fcc and hcp structures. Let us drop superfluous subscripts αl and write

$$U_A = \int_{-\infty}^{E_{FA}} En_A(E) dE, \quad (13.9a)$$

$$U_B = \int_{-\infty}^{E_{FB}} En_B(E) dE, \quad (13.9b)$$

where

$$Z = \int_{-\infty}^{E_{FA}} n_A(E) dE = \int_{-\infty}^{E_{FB}} n_B(E) dE. \quad (13.10)$$

Different energies U_A , U_B will result from different densities of states $n_A(E)$, $n_B(E)$, which will also give Fermi levels E_{FA} , E_{FB} whose difference is of comparable order of magnitude and cannot be glossed over. Multiplying (13.10) by E_{FA} and subtracting from (13.9), we obtain

$$U_A - ZE_{FA} = \int_{-\infty}^{E_{FA}} (E - E_{FA})n_A(E) dE, \quad (13.11)$$

$$\begin{aligned} U_B - ZE_{FA} &= \int_{-\infty}^{E_{FA}} (E - E_{FA})n_B(E) dE \\ &\quad + \int_{E_{FA}}^{E_{FB}} (E - E_{FA})n_B(E) dE. \end{aligned} \quad (13.12)$$

In (13.12) we can take $n_B(E)$ as a constant over the small interval of integration in the last term, which then becomes after integration

$$\frac{1}{2}(E_{FB} - E_{FA})^2 n_B(E_{FA}). \quad (13.13)$$

This quantity is of the second order of smallness, which we neglect. Then subtracting (13.12) from (13.11) gives

$$U_A - U_B = \int_{-\infty}^{E_{FA}} (E - E_{FA})[n_A(E) - n_B(E)] dE. \quad (13.14)$$

Clearly by symmetry we can obtain an equivalent expression to (13.14) with E_{FB} replacing E_{FA} , and it is not difficult to show that it would differ from

(13.14) by another second-order quantity like (13.13). We therefore drop the subscript A on E_{FA} and write

$$U_A - U_B = \int_{-\infty}^{E_F} (E - E_F)[n_A(E) - n_B(E)] dE, \quad (13.15)$$

which completes the transformation. Stated in words, by introducing the $-E_F$ into the integral, we have transformed to an integrand that is zero at E_F so that the difference between E_{FA} and E_{FB} may be dropped.

We shall not catalog here the applications that have been made of these results with the recursion method: they are well reviewed in Sections 25–29 of Kelly's chapter.² Suffice it to say that they include¹⁷⁶ stability of crystal structures,^{4,174} phonons,¹⁷² and ferromagnetism.⁷ See also our Section IV,14c for the interaction energy of adsorbed atoms at a surface and Section V,16 for the pressure–volume relation. Similarly, Bullett in his chapter¹⁵ recites calculations that have been made with his technique relating to¹⁷⁶ cohesion,¹⁰⁰ structure,¹⁰⁰ elastic constants,¹⁰⁰ stacking fault energy,¹¹¹ adsorption dissociation, and migration at surfaces.¹¹⁶ Almost all of these applications have involved no explicit calculation of U_{es} , exceptions being Bullett's calculation of the bulk modules of the diamond-type semiconductors¹⁰⁰ and a few other such. Of course there have been numerous calculations of total energy since the 1930s not from the local point of view and hence outside the scope of the present volume.

Finally it is necessary to return to U_{es} (13.3) to raise and dispose of an important question. As we have described them so far, U_{es} relates to rather atomiclike charge densities and U_1 contains the bonding effects that we see in the band structure: the more bonding the occupied electron states, the lower the band energy U_1 . Now the degree of bonding of the wave functions also affects the valence charge density $\rho(\mathbf{r})$: indeed for a nearest-neighbor s band the effects on U_1 and ρ can be shown to be directly proportional. A bonding wave function heaps up charge between the atoms whereas antibonding functions subtract it. Let us denote by $\Delta\rho_{\text{bond}}(\mathbf{r})$ the contribution to the valence charge density $\rho(\mathbf{r})$ due to the bonding or antibonding nature of the wave functions. Clearly $\Delta\rho_{\text{bond}}$ enters through ρ into U_{es} . Thus it cannot be true that *all* the multiautom structure-dependent effects are contained in U_1 , as alleged earlier, with none in U_{es} . For transition metals Pettifor and Duthie¹⁷⁷ have calculated the contribution of $\Delta\rho_{\text{bond}}$ to the electrostatic energy and found it to be of the order of 10% of the corresponding effects in U_1 . A similar conclusion must be drawn from comparing Varma and Weber's calculation¹⁷¹ of the niobium phonon spectrum with experiment because

¹⁷⁶ The following references are illustrative and not exhaustive. See Kelly's² and Bullett's¹⁵ chapters.

¹⁷⁷ D. G. Pettifor and J. C. Duthie, private communication.

the magnitude of the anomalies is reproduced well without taking $\Delta\rho_{\text{bond}}$ in U_{es} into account. In nearly free-electron metals the effects can be calculated explicitly. The contribution of a reciprocal lattice vector \mathbf{g} to the structure-dependent part of the total energy is³

$$[S(g)v_{\text{ion}}(g)/\epsilon(g)]^2 \chi(g) \epsilon(g), \quad (13.16)$$

where S is the structure factor, v_{ion} bare ion pseudopotential, ϵ the dielectric screening factor of the electron gas, and χ the "perturbation characteristic." Without the factor $\epsilon(g)$ at the end, (13.16) is the contribution to U_1 : the inclusion of $\Delta\rho_{\text{bond}}$ gives the extra factor of $\epsilon(g)$. Now at the first two reciprocal lattice vectors in a typical metal such as aluminum, $\epsilon(g)$ is of the order of 1·1 or 1·2, i.e., the $\Delta\rho_{\text{bond}}$ effects in U_{es} are about 10%–20% of those in U_1 and otherwise identical in form. Whereas in a metal we have perfect screening at long wavelengths and $\epsilon(q) \rightarrow \infty$ as $q \rightarrow 0$, the screening is already rather ineffective typically at the first reciprocal lattice vectors of a close-packed structure: however, the same might not be true for covalent open structures or long wavelength phonons.

In conclusion, we find that to 80% accuracy the important structure-sensitive parts of the energy occur in U_1 , with U_{es} being a rather featureless monotonic short-range repulsion. By one means or another, including the calculation of energy differences, the effect of U_{es} can be rendered small or even negligible in some calculations but by no means in all. In this connection see also Section V,16c. One might say that until recently there has been considerable ignorance about the form of U_{es} , a situation not yet entirely in the past as a reading of this section shows, so that calculations were selected where its effects were hopefully small.

14. VARIOUS TYPES OF PERTURBATION

Every theoretical calculus requires its corresponding perturbation theory. Where would physics be without the concept of successive approximation? In the present instance we introduce in turn three quite distinct types of perturbation calculation, leaving to Sections 20–28 of Haydock's chapter¹ their further elaboration and variation according to whether one has a continuous band or discrete levels, whether one wants only changes in a local density of states or a more elaborate expression involving the Green function, etc.

a. Comparing Different Environments

The simplest question is what change is induced in the continued fraction (11.10) and corresponding local density of states by small changes δa_n , δb_n

in the coefficients? The changes $\delta a_n, \delta b_n$ may be generated by a perturbation in the Hamiltonian as in Section V,14b, but a much wider class of applications lies in comparing physically similar but not identical environments. Suppose we wish to compare the fcc and hcp forms of a transition metal (Section IV,13). In each, an atom has 12 nearest neighbors arranged similarly in triangles with only an orientational difference. From this we conclude (Section IV,11) that the second moments of the densities of states $n_l(E)$ are the same, and we can show that the third moments are too.¹⁷⁸ The first difference lies in the fourth moment, related to a_2, b_2 and higher coefficients (Section IV,11). The consequent difference in $n_l(E)$ controls (see Section IV,13), which elements in the transition series are stable in which structure^{160,178} (see Section 14 of Bullett's chapter¹⁵ and Figs. 2 and 13 of Kelly's²). A similar situation arises if we wish to compare, say, a magnetic atom at a surface and in the bulk.

If $n_l(E)$ is the local density of states corresponding to the continued fraction (11.10), then *the change $\delta n_l(E)$ to first order induced by changes $\delta a_n, \delta b_n$ in the coefficients is*

$$\delta n_l(E) = n_l(E) \sum_n [f_n(E) \delta a_n + g_n(E) \delta b_n], \quad (14.1)$$

where the functions f_n, g_n are detailed by Haydock's¹ Eq. (20.23). In specific calculations the easiest route may well be to calculate the two densities of states and subtract, since this is so fast, rather than to use (14.1), but the perturbation theory is important for a proper theoretical understanding of the relation of the density of states to local atomic structure.

Some points of interest arise, which we describe in relation to a single continuous band of states. The functions f_n, g_n have $2n + 1$ and $2n$ zeros inside the energy range of the band, in addition to the zeros in $\delta n_l(E)$ at the endpoints of the band due to the factor $n_l(E)$ in (14.1). Suppose we have the first few a_n, b_n identical for two situations and δa_n is the first non-zero term in (14.1), giving a $\delta n_l(E)$ with $2r + 1$ internal zeros. Now in almost any conceivable physical case, the higher a_n, b_n will also differ, but it can be proved¹⁷⁹ that the corresponding higher terms in (14.1) can add further zeros in $\delta n_l(E)$ but *cannot reduce the number of zeros below $2r + 1$.* That can also be proved from the point of view of moments, because of the connection (Section IV,11) between moments of $n_l(E)$ and the a_n, b_n . In our example, μ_{2r+1} and all higher moments will differ, whereas the lower ones are identical. Look at it this way: to conserve the total number of states (the zeroth moment), any change in $n_l(E)$ has to be positive for some energies and negative for others to compensate, i.e., $\delta n_l(E)$ has to have at least one zero where it changes from positive

¹⁷⁸ F. Ducastelle and F. Cyrot-Lackmann, *J. Phys. Chem. Solids* **32**, 285 (1971).

¹⁷⁹ R. Haydock, private communication.

to negative. Similarly, to conserve the mean μ_1 , one must move some states up in energy and others correspondingly down, i.e., the simplest $\delta n_i(E)$ would be positive at the top and bottom of the band and negative in the middle, i.e., it would have at least two zeros. Clearly, to conserve higher moments even more subtle changes of the spectral weight $\delta n_i(E)$ are required, with at least $2r + 1$ zeros to conserve the first $2r$ moments as shown by Ducastelle and Cyrot-Lackmann.¹⁷⁸ See Section IV,14.c for the application of these ideas to Friedel oscillations.

b. Changes in the Hamiltonian

The second situation is the conventional question of *what happens if we add the perturbation λV to an unperturbed Hamiltonian H_0* , where λ as usual is an expansion parameter that defines the various orders of perturbation but is set equal to unity at the end.

Two distinct approaches to this question are apparent. The more obvious one is to consider the recurrence relation (12.4) defining the a_n , b_n , and u_n . Each of these quantities can be expanded in powers of λ and the coefficients determined from (12.4). The formulas are tedious but there is no difficulty of principle (Section 26 of Haydock's chapter¹). This approach can then be used in tandem with the method of Section IV,14.a to obtain the changes in the local density of states consequent upon the changes in the a_n , b_n .

The second possibility lies in the usual perturbation expansion of the Green function

$$G = G_0 + G_0 \lambda V G_0 + G_0 \lambda V G_0 \lambda V G_0 + \dots \quad (14.2)$$

Here G and G_0 are the perturbed and unperturbed Green functions

$$G = [E - H_0 - \lambda V]^{-1}, \quad G_0 = [E - H_0]^{-1}, \quad (14.3)$$

which satisfy the Dyson equation

$$G = G_0 + G_0 \lambda V G, \quad (14.4)$$

as can easily be verified by multiplying (14.4) by $(E - H_0)$ from the left and by $(E - H_0 - \lambda V)$ from the right. The first correction term in (14.2) follows by substituting the approximation $G \approx G_0$ in the right-hand side of (14.4) and similarly by iteration the higher terms of (14.2). It is not clear from what we have said so far that quantities like $G_0 \lambda V G_0$ can be calculated, but Haydock (Section 27 of his chapter¹) has shown that it is "almost as simple as densities of states." This should allow calculation of transition rates produced by perturbations and response functions due to applied "forces," as well as

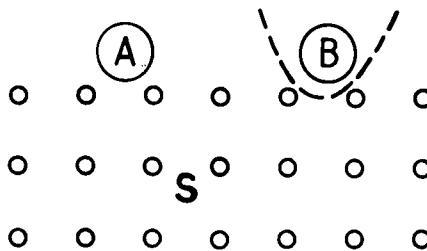


FIG. 22. Interaction between two atoms *A* and *B* on the surface of a substrate *S*.

changes in energy bands due to variation of the Hamiltonian, e.g., for self-consistency in potential or for changes in interatomic spacing due to lattice displacements.

c. Coupling Energies

Our third concern is with a special but important type of perturbation or rather difference calculation, namely, *switching on or off a coupling interaction between two subsystems*, say between an adsorbed atom and the adsorbate surface. We would like to calculate the coupling energy.

Figure 22 shows a slightly more elaborate application of the method to calculate the interaction between two adsorbed atoms *A* and *B* on a substrate *S* as a function of their separation along the surface.¹⁶¹ *B* is one subsystem, *A* and *S* the other, and turning off the coupling between them is represented by the dashed line in Fig. 22. Clearly when *B* is decoupled we can move it around the surface at will and recalculate the coupling at a new site. In this way the distance dependence of the interaction between *A* and *B* is obtained. Moreover, the electrostatic corrections to the energy discussed in Section IV,13 only depend on the immediate neighborhood of *B* and thus will remain constant when *B* is moved to a different equivalent site. These corrections therefore cancel out of the interaction of *A* and *B*, and we only require the sum of the energy eigenvalues, i.e., the integral (13.15) of the *total* density of states.

The recursion method normally gives a *local* density of states $n_{\alpha l}(E)$, and in a situation of low symmetry such as near a surface it would appear necessary to obtain a *total* density of states or total energy for the whole system by summing over a large number of symmetry-unrelated atoms. It is clearly preferable to calculate (if one can) the change due to the coupling directly, rather than as a numerical difference between two large quantities. It turns out somewhat by accident that we can apply^{161,180} the mathematics of

¹⁸⁰ See Section 25 of Haydock's chapter.¹

Section 12 to calculate directly the change in the total density of states, i.e., the formalism takes the difference produced by switching on the coupling and *sums over all atoms automatically*. Strictly speaking the method is not a perturbation theory because the result is exact: there is no perturbation expansion in the sense of Section IV,14b. To demonstrate the ideas we restrict ourselves to the simplest case of one subsystem containing only a single orbital ϕ_0 , e.g., an adsorbed atom with an s state, but the method can be extended^{161,180} to a finite number of orbitals.

We first require a general formula for the total density of states $n(E)$ of an arbitrary system with Hamiltonian H . From Eq. (3.4) we have in terms of the eigenvalues E_n of H

$$\begin{aligned} n(E) &= \sum_n \delta(E - E_n) \\ &= -\pi^{-1} \operatorname{Im} \sum_n 1/(E + iO - E_n) \\ &= -\pi^{-1} \operatorname{Im} \sum_n (\partial/\partial E) \log(E + iO - E_n) \\ &= -\pi^{-1} \operatorname{Im}(\partial/\partial E) \log \Pi_n(E + iO - E_n) \\ &= -\pi^{-1} \operatorname{Im}(\partial/\partial E) \log \det|E + iO - H|. \end{aligned} \quad (14.5)$$

Here Π_n is the product of all terms n , and the last step is a consequence of $\det|E - H|$ being invariant under a unitary transformation, in particular to the diagonal form of the previous line in terms of the eigenstates ψ_n . Thus if H_0 is an unperturbed Hamiltonian corresponding to H the change in $n(E)$ due to switching on the perturbation is

$$\begin{aligned} n(E) &= n(E) - n_0(E) \\ &= \pi^{-1} \operatorname{Im}(\partial/\partial E) \log[\det|E + iO - H_0|/\det|E + iO - H|]. \end{aligned} \quad (14.6)$$

The ratio of determinants in (14.6) is reminiscent of Eq. (12.10), and can in our case be handled in analogous fashion. We take the orbital ϕ_0 of the B atom in Fig. 22 (i.e., of the subsystem to be decoupled) as the starting vector u_0 for the recursion procedure of Section IV,12. Then u_1 is a linear combination of atomic orbitals on the surface, a “surface orbital” or “group orbital”¹⁸¹ that ϕ_0 directly bonds to, and by the orthogonality construction all u_n for $n > 1$ do not involve ϕ_0 . Thus the matrix $E - H$ in the form (12.9) is cleanly separated into different parts: $E - a_0$ relates to the surface atom, the part denoted by D_1 in (12.9) describes only the substrate, while b_1 is the coupling between the two. The matrix $E - H_0$ is simply obtained by setting

¹⁸¹ For a definition see Section 16 of Kelly's chapter.²

b_1 equal to zero to decouple the adsorbed atom from the substrate, and we have, analogously to (12.11),

$$\det|E - H| = (E - a_0) \det|D_1| - b_1^2 \det|D_2|, \quad (14.7)$$

$$\det|E - H_0| = (E - a_0) \det|D_1|, \quad (14.8)$$

$$\frac{\det|E - H_0|}{\det|E - H|} = \frac{E - a_0}{E - a_0 - b_1^2 \det|D_2|/\det|D_1|}, \quad (14.9)$$

where the last expression can be developed as a continued fraction exactly as in (12.12) and (12.13). Here E is a general complex variable, which we take with small positive imaginary part in substituting in (14.6). This then gives $\delta n(E)$, which has to be integrated as in Section IV,11 and IV,13 to obtain the interaction energy.

Returning to the example of two adsorbed atoms on a surface (Fig. 22), their interaction energy *oscillates* with the separation between them; at least $\delta n(E)$ oscillates, and so will the integrated energy (Fig. 21). The a_n , b_n coefficients at atom B will only “know about” the presence of atom A (Fig. 22) for $n > r$, where r is the number of steps in the recursion (12.4) needed for the u_n to reach A in the sense of Fig. 15. For a nearest-neighbor Hamiltonian, r is the number of nearest-neighbor steps between A and B , and the $\delta n(E)$ will be related to the $f_r(E)$ and $g_r(E)$ of Eq. (14.1). Now from Eqs. (11.12) and (20.33) of Haydock’s chapter¹ we have

$$\sum_n f_n(E) = 0, \quad (14.10)$$

i.e., at a fixed E , in order to sum to zero, $f_n(E)$ must oscillate with increasing n , i.e., the $\delta n(E)$ oscillates with increasing separation r . This oscillating behavior can also be seen more generally from the oscillating property of the polynomials $P_n(E)$ and related functions in terms of which f_n is defined (see Section 20 of Haydock’s chapter¹), with an ever-increasing number of zeros as n increases. We may think of n as a continuous variable, generating more and more oscillations from, say, the bottom of the band, which then spread upwards past any fixed E into the upper part of the band. Since the zeros of $P_{n+1}(E)$ always interleave with those of $P_n(E)$, this is a valid description. It is an example of how the Friedel oscillations (see Sections I,2 and II,5) manifest themselves in a discrete space, or in Section 10 of Haydock’s chapter¹ the oscillations of the wave function.

When the subsystem has *several orbitals*, then any of different symmetry do not interact (see Section IV,12) and can be treated separately. But one must use symmetry rigorously: for example the system shown in Fig. 22 has only one mirror plane about the B atom, so that several orbitals will have

the same symmetry (even or odd with respect to this plane) and must be considered together. As already mentioned, the method can be extended to decouple them one at a time in turn.^{161,180}

V. Miscellaneous Properties from the Local Point of View

15. A LOCAL COMPUTATIONAL SCHEME FOR THE DENSITY FUNCTIONAL FORMALISM

The purpose of introducing the density functional formalism¹⁸² briefly at this stage is twofold. First, we require it to prove Andersen's theorem¹⁸³ in the next section. Second, Ying¹⁸⁴ and Bohnen¹⁸⁵ have recently developed a computational scheme that goes some way toward exploiting the local nature of the density functional formalism. The density functional formalism is conceptually local¹⁸⁶ in the sense of being expressed in terms of the local electron density $\rho(\mathbf{r})$. However, to be applied in practice¹⁸² it usually has to be converted to a one-electron-like Schrödinger-like equation like (8.1) whose eigenfunctions spread throughout the whole solid or molecule so that it is difficult to extract local information as discussed in Section I,1. In contrast, the scheme of Ying and Bohnen goes back to the original local functional and its variational property.

We summarize the basic results¹⁸² of the density functional formalism without proof. We consider a system in its ground state of stationary nuclei with an electron density $\rho(\mathbf{r})$. The total energy is given by a functional F of the electron density:

$$\begin{aligned} F[\rho(\mathbf{r})] = & U_{\text{NN}} + T_s[\rho] + \int V_N(\mathbf{r})\rho(\mathbf{r}) d^3\mathbf{r} \\ & + \frac{1}{2} \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3\mathbf{r} d^3\mathbf{r}' + U_{\text{xc}}[\rho], \end{aligned} \quad (15.1)$$

where the notation follows that of Sections III,8 and IV,13. By a functional we mean that, given $\rho(\mathbf{r})$ everywhere, it is in principle possible to calculate

¹⁸² For a review of the density functional formalism with detailed references, see Lang.⁹⁴

¹⁸³ See Section 3.1 of Machkintosh and Andersen.¹²⁶

¹⁸⁴ S. C. Ying, *J. Phys. (Paris)* **38**, C4 (1977).

¹⁸⁵ K. P. Bohnen and S. C. Ying, to be published.

¹⁸⁶ Of course, the density functional is not a local operator in the mathematical sense, but its degree of nonlocality is of the order of atomic dimensions, not the size of the whole solid, so that it qualifies as "local" within the meaning of this volume.

F by some set of operations (which may be very complicated): the electrostatic energy of the electron gas and the interaction with the potential V_N of the nuclei are very simple terms written explicitly in (15.1), but the kinetic energy T_s and exchange-correlation energy U_{xc} are more complicated functionals of which the latter is only known in various approximations, e.g., if $\rho(\mathbf{r})$ is nearly constant. U_{NN} is the nuclear-nuclear Coulomb energy.

The functional F has three important properties. First, existence: given $\rho(\mathbf{r})$ there is only one set of nuclei and one set of positions for them from which $\rho(\mathbf{r})$ could have arisen, and hence a unique energy for the system. This uniqueness and indeed the whole formalism is only true for the ground-state charge density. Second, the ground-state property: (15.1) correctly gives the ground-state energy U_G of the system when one substitutes the ground-state charge density $\rho_G(\mathbf{r})$ into (15.1):

$$U_G = F[\rho_G(\mathbf{r})]. \quad (15.2)$$

Third, the variational theorem: F is a minimum for the correct ground-state charge density ρ_G if one varies ρ in (15.1) keeping the nuclear positions constant. We may write this as

$$\delta F[\rho(\mathbf{r})] = 0, \quad \text{for } \delta\rho = \rho(\mathbf{r}) - \rho_G(\mathbf{r}), \quad (15.3)$$

remembering that it is only true to first order in $\delta\rho$. Also the variation has to conserve the total number N of electrons

$$\int \rho(\mathbf{r}) d^3\mathbf{r} = N. \quad (15.4)$$

As a point of notation, since the whole formalism is concerned with the ground state we drop the suffix G in what follows, except in special cases where we wish to emphasize it as in (15.2) or where we explicitly consider deviations from ρ_G as in (15.3).

The kinetic energy functional is usually evaluated by transforming to a one-electron-like representation with wave functions ψ_n , where

$$\rho(\mathbf{r}) = \sum_n |\psi_n(\mathbf{r})|^2 \quad (15.5)$$

(summed over occupied states). The ψ_n are determined by a one-electron-like Schrödinger equation²⁰

$$(-\frac{1}{2}\nabla^2 + V_N + V_H + V_{xc})\psi_n = E_n\psi_n \quad (15.6)$$

analogous to (8.1), with V_N and V_H having the same meaning as before, i.e., V_N is the potential due to the nuclei and V_H is the Hartree-like potential due to the charge density (15.5)

$$V_H(\mathbf{r}) = \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3\mathbf{r}'. \quad (15.7)$$

Equation (15.6) has been so chosen that the kinetic energy T_s in (15.1) is given by

$$T_s = \sum_n \langle \psi_n | -\frac{1}{2} \nabla^2 | \psi_n \rangle \quad (15.8)$$

$$= \sum_n E_n - \int (V_N + V_H + V_{xc}) \rho(\mathbf{r}) d^3\mathbf{r}, \quad (15.9)$$

where the last transformation follows from (15.6) and (15.5).

The most common and simplest form for $U_{xc}[\rho]$ is the local density approximation

$$U_{xc}[\rho(\mathbf{r})] = \int u_{xc}[\rho(\mathbf{r})] \rho(\mathbf{r}) d^3\mathbf{r}, \quad (15.10)$$

where $u_{xc}(\rho)$ is the exchange–correlation energy per electron of a uniform free-electron gas of density ρ . Approximation (15.10) is therefore in spirit like the Thomas–Fermi approximation, treating each element of volume $d^3\mathbf{r}$ as a macroscopic piece of electron gas of density $\rho(\mathbf{r})$. With (15.10) the V_{xc} potential in (15.6) turns out to be

$$V_{xc}(\mathbf{r}) = d[\rho u_{xc}(\rho)]/d\rho = \mu_{xc}[\rho(\mathbf{r})], \quad (15.11)$$

where μ_{xc} is the exchange–correlation component of the chemical potential μ of the free-electron gas.

It must be emphasized that the density functional theory is purely concerned with the ground state, that the eigenfunctions ψ_n are only a device for calculating the kinetic energy, and that the energy values E_n do not have any direct physical significance in contrast to the Landau quasi-particle energies in Section III.8, which are genuine excitation energies of the system. That is why we eschewed the density functional formalism in the discussion of Section III.8. In some areas of physics one is purely concerned with the ground state, but much of physics is concerned with probing a solid to produce excitations. Incidentally, the Fermi surface can be formulated as a ground-state property and is correctly given by both approaches. The difference between the two can be seen, for example, in the fact that the V_{xc} of (15.6) is the same for all the “electrons” ψ_n , whereas the self-energy operator V_{xc} of Eq. (8.1) certainly varies through the band. However, this comparison also shows that the two formalisms are not very different, if at all, in practice. We have already seen⁹⁸ in Eqs. (8.7) and (8.8) that for a free-electron gas of densities common in solids (at least for the valence electrons) the self-energy varies only by the order of 10% through the band, and if we take it as equal to its value μ_{xc} at the Fermi level we recover the approximation (15.11) of the present section. In any case, (15.6) and (8.1) are identical at the Fermi level.

Substituting (15.7)–(15.11) in (15.1) and (15.2) yields the total energy in the form similar to (8.9) and Section IV,13:

$$U = U_{\text{NN}} + \sum_n E_n - \frac{1}{2} \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3\mathbf{r} d^3\mathbf{r}' \\ - \int \{\mu_{xc}[\rho(\mathbf{r})] - u_{xc}[\rho(\mathbf{r})]\}\rho(\mathbf{r}) d^3\mathbf{r}. \quad (15.12)$$

Such then is the usual development of the density functional theory.

We turn now to the recent alternative development of density functional theory by Ying¹⁸⁴ and Bohnen,¹⁸⁵ which seeks to avoid the Schrödinger equation (15.6) and to evaluate the local form of the kinetic energy T_s in (15.1) directly. Let us express the ψ_n in terms of a local basis set $\phi_\alpha(\mathbf{r})$ as in (1.8) and (7.1), where α is a composite site and orbital-type index:

$$\psi_n(\mathbf{r}) = \sum_\alpha a_{n\alpha} \phi_\alpha(\mathbf{r}). \quad (15.13)$$

The charge density is easily expressed in terms of the ϕ_α ,

$$\rho(\mathbf{r}) = \sum_{\alpha\beta} \sum_n a_{n\alpha}^* a_{n\beta} \phi_\alpha(\mathbf{r})^* \phi_\beta(\mathbf{r}), \quad (15.14)$$

where the summation \sum_n is over occupied states only. Thus the charge density is characterized completely by the density matrix $[\rho]$:

$$\rho_{\alpha\beta} = \sum_n a_{n\alpha}^* a_{n\beta}, \quad (15.15)$$

which can therefore be used as the variable in the variational theorem (15.3). The kinetic energy from (15.8) can be expressed in terms of it,

$$T_s = \sum_{\alpha\beta} \rho_{\alpha\beta} \langle \phi_\alpha | -\frac{1}{2} \nabla^2 | \phi_\beta \rangle, \quad (15.16)$$

which is a local formulation of T_s in the sense of the present volume in that $\rho_{\alpha\beta}$ gives the kinetic energy localized between the sites α and β . (It tends to zero when α, β are more than a few sites apart.)

There is one problem about varying $[\rho_{\alpha\beta}]$: it cannot be varied arbitrarily but the matrix must satisfy the idempotency property

$$[\rho]^2 = [\rho] \quad (15.17)$$

as follows from (15.15) using the orthogonality of the ψ_n :

$$\langle \psi_n | \psi_m \rangle = \sum_\alpha a_{n\alpha}^* a_{m\alpha} = \delta_{nm}. \quad (15.18)$$

If one applies the constraint (15.17) via a Lagrange multiplier, or rather a whole matrix of Lagrange multipliers, then as usual one arrives back at eigenvalue equations not very different from (15.6). Ying¹⁸⁴ and Bohnen¹⁸⁵

therefore used an alternative device. Let us suppose we are considering a defect in a solid or a solid surface. Then far from the defect or surface $[\rho]$ is equal to its value $[\rho_0]$ in the bulk material, and let us assume this is known from some conventional band structure calculation or otherwise. In calculations on potassium,¹⁸⁵ the bulk $[\rho_0]$ was just determined from the pseudopotential by linear screening theory,⁶⁷ no band structure calculation being necessary. The $[\rho]$ in the presence of the surface or defect is then written in the form

$$[\rho] = (\rho_0 + \rho_0^c \Delta \rho_0)(1 + \rho_0 \Delta \rho_0^c \Delta \rho_0)^{-1}(\rho_0 + \rho_0 \Delta \rho_0^c), \quad (15.19)$$

where $[\rho_0^c] \equiv [1] - [\rho_0]$ and $[\Delta]$ is an *arbitrary* matrix. The form (15.19) automatically satisfies¹⁸⁷ the idempotency condition so that $[\Delta]$ can be regarded as a matrix of constraint-free variational parameters $\Delta_{\alpha\beta}$. At some distance from the surface or defect we set $\Delta_{\alpha\beta} = 0$ and then $[\rho]$ (15.19) reduces to $[\rho_0]$. Thus the nonzero part of $[\Delta]$ is confined to near the surface or defect.

So far, so good. However, there is a matrix inversion in the middle factor of (15.19), inverting a matrix that at first sight spans the whole space covered by the ϕ_α , i.e., its dimension as a matrix is equal to the number of orbitals! Now such a matrix inversion would be equivalent to solving a Schrödinger equation for the whole system. But in the present case $[\rho]$ and $[\rho_0]$ can be cut off a little distance from the defect or surface where $[\rho]$ becomes equal to its bulk value, thus reducing $[\rho]$ and the matrix inversion to a more limited size. In the case of a point defect it would involve only a finite number of orbitals on the near neighbors. In Bohnen and Ying's¹⁸⁵ study of a surface of potassium, this trick reduced $[\rho]$ to a few layers perpendicular to the surface, but it still was of infinite extent parallel to the surface. To deal with that they had to expand everything in two-dimensional \mathbf{k}_\parallel space parallel to the surface, so that their method is not truly local in the sense of the present volume. However, numerically they found Δ to be nearly independent of \mathbf{k}_\parallel . Perhaps further developments will overcome this problem, possibly by using the periodicity of $[\rho]$ in a different way. It may also be of advantage to avoid explicit matrix inversion by expanding the middle term of (15.19) as a power series that can be evaluated to a dozen or so terms by matrix multiplication. Let us simply note that in the first application of the method to a potassium surface¹⁸⁵ with quite a limited basis set, a good value of the work function was obtained but a somewhat high surface energy.

For a different way of exploiting the local form of the density functional formalism, with application to hydrogen and helium in solution and in vacancies in metals, see, Stott and Zaremba.^{187a}

¹⁸⁷ R. McWeeney, *Rev. Mod. Phys.* **32**, 35 (1960).

^{187a} M. J. Stott and E. Zaremba, to be published.

16. PRESSURE, FORCE, AND ENERGY

a. Pressure–Volume and Energy–Volume Relations

In the kinetic theory of gases one visualizes the pressure arising from molecules hitting a boundary wall and having their momentum changed there, leading to the well-known formula

$$PV = \frac{1}{3} Nmc\overline{c^2} \quad (16.1)$$

for the pressure P of a perfect gas. Even in the interior of the gas, one can imagine an element of surface dS traced out, consider the molecules passing through it, and calculate the change in momentum that they would suffer if the element dS were made solid. One can further take into account the attractive force between the molecules by introducing a correction to P as in Van der Waals' equation.

Remarkably enough, an analogous *local formulation of the pressure* can be carried out in quantum mechanics. We can draw any surface S in a solid or a molecule and inquire about the force exerted by the electrons across it. In particular, we may take for S the surface of the Wigner–Seitz cell and obtain an expression due to Liberman¹⁸⁸ for the pressure in the system as²⁰

$$3PV = \int \left\{ \frac{1}{4} \sum_n [(\nabla\psi_n^*)(\mathbf{r} \cdot \nabla\psi_n) - \psi_n^* \nabla(\mathbf{r} \cdot \nabla\psi_n) + \text{c.c.}] + \frac{1}{3} \rho u_{xc} \mathbf{r} \right\} \cdot d\mathbf{S}, \quad (16.2)$$

where c.c. denotes complex conjugate and u_{xc} is the exchange–correlation energy density per electron, evaluated for the electron density found on the surface. Note that the momentum operator occurs twice, like the square of the velocity in (16.1), and there is an interaction term as already remarked in connection with Van der Waals' equation. Pettifor¹⁸⁹ has tested this formulation on the hydrogen molecular ion.

After approximating to the Wigner–Seitz cell by an atomic (Wigner–Seitz) sphere of radius s , Pettifor¹⁹⁰ transformed (16.2) using the fact that ψ_n satisfies the Schrödinger equation, to the form

$$\begin{aligned} 4\pi s^2 P = \sum_l \int^{E_F} dE n_l(E) u_l^2(s, E) \{[E - V_{xc}(s)]s^2 \\ + (D_l - l)D_l + l + 1) + \frac{1}{3}u_{xc}(s)s^2\}, \end{aligned} \quad (16.3)$$

¹⁸⁸ D. A. Liberman, *Phys. Rev. B* **3**, 2081 (1971).

¹⁸⁹ D. G. Pettifor, *J. Chem. Phys.* **69**, 2930 (1978).

¹⁹⁰ D. G. Pettifor, *Commun. Phys.* **1**, 141 (1976).

also obtained by Niemenen and Hodges¹⁹¹ in a different way. Here V_{xc} is the exchange–correlation potential defined in Sections III,8 or V,15, $u_l(r)$ the radial wave function normalized to unity in the sphere as in Section III,10, and D_l the logarithmic derivative¹⁹²

$$D_l(E) = \frac{s}{u_l(s, E)} \left. \frac{\partial u_l(r, E)}{\partial r} \right|_{r=s}. \quad (16.4)$$

Also $n_l(E)$ is the partial density of states of angular momentum l (not the local density of states at lattice site as elsewhere in this chapter; here all sites are assumed equivalent):

$$n_l(E) = \sum_{nm} |a_{nlm}|^2 \delta(E - E_n), \quad (16.5)$$

where the eigenstates ψ_n have been decomposed into angular momentum components

$$\psi_n(\mathbf{r}, E) = \sum_{lm} a_{nlm} u_l(r, E) Y_{lm}(\theta, \phi). \quad (16.6)$$

Calculations of the P–V relation have been made from (16.3) after various further transformations^{126,190} and simplifications with the atomic sphere approximation.^{120,125,126} Glötzel *et al.*¹⁹³ investigated the 4d and 5d transition metal series near equilibrium, while Pettifor¹¹⁸ considered up to $\pm 50\%$ changes in volume for the 4d series. The calculations employ the method of Section V,17 for the self-consistency of the potential, and the results for the 5d metals are shown in Fig. 23. In these the equilibrium Wigner–Seitz radius s and bulk modulus B were evaluated from the P – V relations, and the results must be regarded as very gratifying.

An interesting feature of Eq. (16.3) is its decomposition into angular momentum components l , thus allowing an identification of the respective roles of the d electrons and sp electrons in the cohesion. Figure 24 shows the cancellation, at equilibrium, of *an inward pressure from the d electrons balanced by an outward pressure from the sp electrons*. We have seen in Section III,10 that the d bandwidth is proportional to s^{-5} , so that it widens rapidly on compression. Thus if it is partially filled, the energy of the occupied d states will be lowered by the compression, resulting in an inward (negative) pressure. The positive pressure of the sp electrons results partly from the change in the bottom of the sp band on compression and partly from their large free-electron kinetic energy, i.e., the Fermi pressure, as argued by Kollar and

¹⁹¹ R. M. Niemenen and C. H. Hodges, *J. Phys. F* **6**, 573 (1976).

¹⁹² Note that D_l differs from L_l defined in Section III,10 by the factor s .

¹⁹³ Y. Glötzel, D. Glötzel, and O. K. Andersen, to be published.

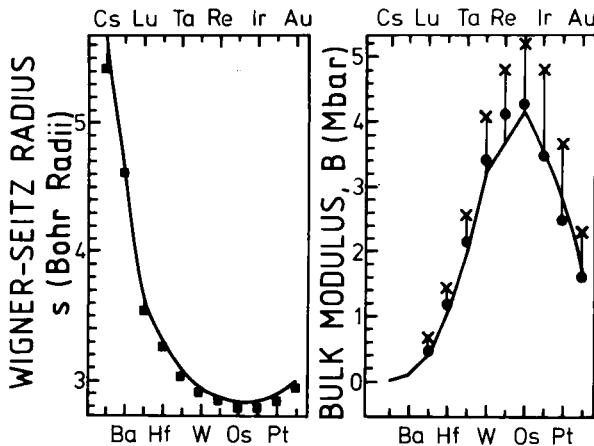


FIG. 23. Calculated and observed equilibrium atomic radii s and bulk moduli B of the 5d transition series. Full line: experiment; squares: calculated s ; dots: B calculated at the observed s ; crosses: B calculated at the calculated s (after Mackintosh and Andersen¹²⁶).

Solt.¹⁹⁴ From this physical picture we may speculate that in a dimer molecule of the transition element or in cluster molecules or at the surface of the metal, the sp electrons have an opportunity to spill outward to relieve their pressure, leaving the d electrons to produce a contraction of the interatomic spacing. Such a contraction is found in calculations on the dimer molecules,¹⁹⁵ and in observations on some metal cluster molecules¹¹⁷ (particularly where halogens extract the s electrons,¹⁹⁶ e.g., Re_2Cl_8) and probably on the molybdenum surface.^{190,197}

Pettifor^{118,190} has further shown that with some approximation the formulas for the partial pressures in (16.3) may be integrated analytically from infinite separation to a finite s to give an expression U_l for the contributions to the *total energy* U . Thus the d electrons contribute an energy (for N_d electrons per atom, relative to the free atom)

$$U_d = N_d \left[\frac{E_d - E_d^{\text{atom}}}{4\mu_d} - \frac{u_{xc} - \frac{4}{3}E_d^{\text{atom}}}{2\mu_d} + \bar{E} \right], \quad (16.7)$$

where $\mu_d \approx 4$ is the effective mass for the d band defined in (10.40) and (10.41), and \bar{E} is the mean energy of the occupied states of the band relative

¹⁹⁴ J. Kollar and G. Solt, *J. Phys. Chem. Solids* **33**, 651 (1972).

¹⁹⁵ R. O. Jones and R. Harris, *J. Chem. Phys.* (to be published). Note that exceptions to our statement occur when an atomic d^5 6S configuration behaves essentially like a closed shell.

¹⁹⁶ R. G. Woolley, private communication.

¹⁹⁷ A. Ignatiev, F. Jona, H. D. Smith, D. W. Jepson, and P. M. Marcus, *Phys. Rev. B* **11**, 4787 (1975), but see also G. P. Kerker, K. M. Ho, and M. L. Cohen, *Phys. Rev. Lett.* **40**, 1593 (1978).

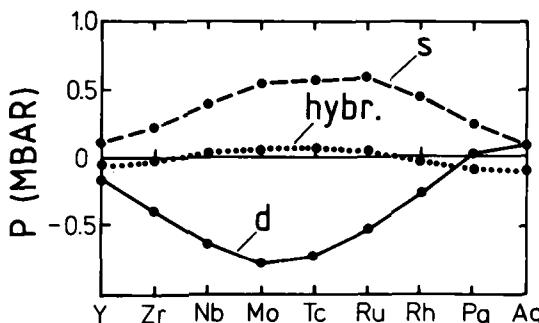


FIG. 24. Contributions at equilibrium to the pressure P for the 4d transition series. Full line: from the d electrons only; dashed line: from the sp electrons only; dotted line: extra contribution due to the hybridisation between the d bands and sp band. The three do not add exactly to zero because they were evaluated at the observed atomic volume, not the calculated equilibrium (after Pettifor¹¹⁸).

to the center of the band E_d . The latter is the same as E_0 in Section III,10 and it varies with s because the whole potential varies with compression and because of overlap as in (9.19). In fact, E_d rises as s decreases,¹³² and for a full d shell the first term of (16.7) is the dominant term, the closed-shell repulsion, because \bar{E} is then zero. With¹⁹⁰ $(E_d - E_d^{\text{atom}}) \propto s^{-5}$ and $\mu_d \propto s^{-3}$ from (10.40), the closed-shell repulsion interaction is proportional to s^{-8} , but this will be different for different l . Note that the term $(E_d - E_d^{\text{atom}})$ is renormalized in (16.7)¹⁹⁰ by a large reduction factor $(4\mu_d)^{-1} \approx 0.06$.¹³² The one-electron energy U_1 of Section IV,13 is simply $E_d - E_d^{\text{atom}}$ per electron for a full band, without any reduction factor, but most of this is cancelled by the intra-atomic electrostatic energy U_{es} (Section IV,13) as already found numerically by Watson *et al.*¹⁹⁸ in their renormalized atom approach. In the present theory such cancellation is taken into account analytically, leading to the renormalization factor. Of course, \bar{E} dominates in (16.7) for a partially filled band as already discussed in connection with the d contribution to the pressure.

A further point of interest emerges from (16.7), namely, that *the closed-shell interaction* at large separations is attractive, as has been noted by Kollar and Solt¹⁹⁹ in their calculations on copper. At large distance the $(4/3)E_d^{\text{atom}}$ term in (16.7) dominates [renormalized by $(2\mu_d)^{-1}$ as before] giving a negative interaction energy. It appears to be the electrostatic effect of the electron density from the edge of one atom penetrating into the attractive potential (attractive for electrons) inside the other atom, i.e., the middle term of (9.23).

¹⁹⁸ R. E. Watson, H. Ehrenreich, and L. Hodges, *Phys Rev. Lett.* **24**, 829 (1970).

¹⁹⁹ J. Kollar and G. Solt, *J. Phys. Chem. Solids* **35**, 1121 (1974).

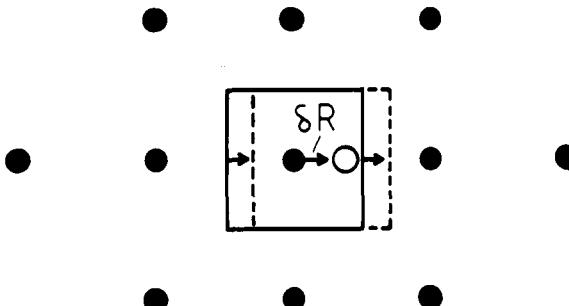


FIG. 25. Displacement of the atomic cell boundary when one atom in a solid is moved a distance δR .

This presumably is the reason why Tricky *et al.*²⁰⁰ found bonding in the rare gas solids even though the Van der Waals interaction had not been included explicitly in their calculations. The changeover from long-range attraction to short-range expulsion is more clearly seen in the pressure formula before integration¹⁹⁰:

$$P_d \propto E_d - u_{xc}, \quad (16.8)$$

i.e., it occurs when E_d rises above u_{xc} as the atoms are brought together. In silver the calculations²⁰¹ show this to happen at a separation somewhat greater than the equilibrium one, i.e., at equilibrium the d shells are somewhat under compression as shown in Fig. 24. Indeed, metallurgical evidence suggests that the situation applies to both silver and gold (copper does not appear to have been studied) because a small but noticeable contraction of the atomic volume occurs in simple alloys with nearly free-electron metals where the d shells are kept out of contact by the other element present.²⁰²

b. The Local Force Theorem

Andersen¹⁸³ has proved a remarkable theorem relating the force due to arbitrary atomic displacements to a certain change in one-electron energy $U_1 = \sum_n E_n$ (Section IV,13) only. Suppose we move a single atom in a solid (Fig. 25); we can draw an atomic cell around it, and consider the force on it as acting across the cell boundary, which is displaced with the atom as shown

²⁰⁰ S. B. Tricky, F. R. Green, and F. W. Averill, *Phys. Rev. B* **8**, 4822 (1973).

²⁰¹ See Fig. 3c of Pettifor¹³² and note that ϕ is missing from the label $\phi(s)$. In the Slater⁹² $\rho^{1/3}$ approximation used by Pettifor, our $u_{xc} = (3/4)[\text{his } \phi(s)]$. Also his C_d is our E_d .

²⁰² M. V. Nevitt and subsequent Discussion, in "Phase Stability of Metals and Alloys" (P. S. Rudman, J. Stringer, and R. I. Jaffee, eds.), p. 281. McGraw-Hill, New York, 1967.

in Fig. 25. In a uniform expansion the cells are pulled apart: in a shear they are slid over each other. Thus any deformation can be treated. Following Andersen we prove the theorem from the density functional form of the total energy (Section V.15) but there seems no doubt that the result is equivalent to the pressure formulation (16.2) because they give identical results for a uniform compression, as we shall see.

The proof¹⁸³ of the theorem is algebraically somewhat tedious. Consider the solid cut into two pieces a and b along some surface S usually chosen along the boundaries of Wigner–Seitz cells but in principle anywhere. Piece b is displaced by a small amount δR , keeping region a fixed in position and opening up a small region c along the cut. The force between the two pieces is to be calculated from $-\delta U/\delta R$, and the purpose is to calculate the energy change δU from Eq. (15.1). The trial electron density $\rho(\mathbf{r})$ is defined in a slightly indirect way as $\sum_n |\psi_n|^2$, where the ψ_n are eigenfunctions of the one-electron Schrödinger equation with some potential $V(\mathbf{r})$ to be specified below. The kinetic energy is obtained analogously to (15.9) as

$$T_s = \sum_n E_n - \int V(\mathbf{r})\rho(\mathbf{r}) d^3\mathbf{r}, \quad (16.9)$$

and with the local density approximation (15.10) the total energy (15.1) becomes

$$\begin{aligned} U = U_{NN} + \sum_n E_n - \int V\rho d^3\mathbf{r} + \int V_N\rho d^3\mathbf{r} \\ + \frac{1}{2} \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3\mathbf{r} d^3\mathbf{r}' + \int u_{xc}\rho d^3\mathbf{r}. \end{aligned} \quad (16.10)$$

The δU on making the displacement will be the sum of six terms δU_1 to δU_6 , corresponding to the six contributions on the right-hand side of (16.10). The first two are trivial:

$$\delta U_1 + \delta U_2 = U_e(N_a; N_b) + \delta \left(\sum_n E_n \right). \quad (16.11)$$

Here δU_{NN} is written in a new notation, where

$$U_e(X; Y) = \text{mutual electrostatic energy of charge systems } X \text{ and } Y. \quad (16.12)$$

Since the nuclei in region b are displaced rigidly from region a , there is no change in the Coulomb interaction of the nuclei of region a among themselves or of region b among themselves.

It now becomes necessary to specify $V(\mathbf{r})$. In the undisplaced system, V is the self-consistent field for that system:

$$V = V_{\text{SCF}} = V_N + V_H + \mu_{xc} \quad (16.13)$$

in the notation of (15.6) and (15.11). The V in the displaced system is defined in terms of the V_{SCF} (16.13) of the *undisplaced* system: in region a the V is taken as V_{SCF} unchanged; in region b it is taken as V_{SCF} displaced rigidly with the nuclei; and in region c we leave V still unspecified for the present. The important point is that the genuine self-consistent potential of the displaced system need never be calculated. Our V differs from it by something of order δR , resulting in our trial ρ for the displaced system having errors of order δR , but the variational property (15.3) means that δU only has errors of order $(\delta R)^2$ and hence our calculation gives the force correctly.

Let us designate the perturbed charge density as $\rho_a + \delta\rho_a$ in region a , as $\rho_b^d + \delta\rho_b$ in region b , and ρ_c in region c , where ρ_b^d is the rigidly displaced unperturbed charge density, and the suffix a , etc., is an abbreviation for the variable \mathbf{r}_a , etc. Clearly, we have

$$\int V_b^d \rho_b^d d\mathbf{r}_b = \int V_b \rho_b d\mathbf{r}_b \quad (16.14)$$

since the displacement is just a change of origin that does not affect the value of the integral. Note also we have abbreviated $d^3\mathbf{r}$ to $d\mathbf{r}$, and have added a subscript b to indicate the integration $d\mathbf{r}_b$ is taken over region b only, as is also evident from the suffix on ρ_b . Thus from (16.10) δU_3 becomes

$$\begin{aligned} \delta U_3 &= - \int V_a \delta\rho_a d\mathbf{r}_a - \int V_b \delta\rho_b d\mathbf{r}_b - \int V_c \rho_c d\mathbf{r}_c \\ &= - \int (V_N + V_H + \mu_{xc}) \delta\rho_a d\mathbf{r}_a \\ &\quad - \int (V_N + V_H + \mu_{xc}) \delta\rho_b d\mathbf{r}_b - \int V_c \rho_c d\mathbf{r}_c. \end{aligned} \quad (16.15)$$

Similarly, δU_4 becomes

$$\begin{aligned} \delta U_4 &= \int V_N \delta\rho_a d\mathbf{r}_a + \int V_N \delta\rho_b d\mathbf{r}_b + \int V_N \rho_c d\mathbf{r}_c \\ &\quad + \int \delta V_N \rho_a d\mathbf{r}_a + \int \delta V_N \rho_b^d d\mathbf{r}_b^d. \end{aligned} \quad (16.16)$$

The last two terms can be expressed in the notation of (16.12). The δV_N in

region a is only due to the nuclei N_b of region b being displaced, and $\delta V_N(\mathbf{r}_b^d)$ arises from the nuclei N_a in region a :

$$\begin{aligned}\delta U_4 = & \int V_N \delta\rho_a d\mathbf{r}_a + \int V_N \delta\rho_b d\mathbf{r}_b + \int V_N \rho_c d\mathbf{r}_c \\ & + \delta U_e(\rho_a; N_b) + \delta U_e(N_a; \rho_b).\end{aligned}\quad (16.17)$$

We turn to the fifth term. In the displaced system ρ is the sum of the five parts $\rho_a, \delta\rho_a, \rho_b, \delta\rho_b, \rho_c$. The product $\rho(\mathbf{r})\rho(\mathbf{r}')$ therefore contains 25 terms. Of these, nine are quadratic of order $(\delta R)^2$, e.g., $(\delta\rho_a)(\delta\rho_b)$ and $(\delta\rho_a)\rho_c$, and we drop them because we are only interested in the linear term for calculating the force, and because we shall use the variational theorem (15.3), which applies only to the first order. One term $\rho_b^d\rho_b^d$ in the displaced system cancels $\rho_b\rho_b$ undisplaced for the same reason as in (16.14), and $\rho_a\rho_a$ cancels directly. The remaining $25 - 9 = 14$ terms in $\rho\rho$ can all be grouped into seven pairs such as $\rho_a\rho_b^d$ and $\rho_b^d\rho_a$, thus canceling the factor $\frac{1}{2}$ in (16.10). Note that $\rho_a\rho_b^d$ does not equal exactly the undisplaced $\rho_a\rho_b$: the difference between the two terms is the work done in pulling these two charge distributions apart. We are left with the following contribution from the $\rho\rho$ term in (16.10):

$$\begin{aligned}\delta U_5 = & \int \frac{\rho_a(\mathbf{r})[\rho_b^d(\mathbf{r}') - \rho_b(\mathbf{r}')]}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' \\ & + \int \frac{(\rho_a + \rho_b^d)(\mathbf{r}') \delta\rho_a}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}_a d\mathbf{r}' \\ & + \int \frac{(\rho_a + \rho_b^d)(\mathbf{r}') \delta\rho_b}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}_b^d d\mathbf{r}' \\ & + \int \frac{(\rho_a + \rho_b^d)(\mathbf{r}') \rho_c}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}_c d\mathbf{r}'.\end{aligned}\quad (16.18)$$

The first term can be expressed in the notation of (16.12). In the remaining terms the difference between ρ_b and ρ_b^d contributes a small second-order quantity to δU_5 , which we neglect, and we can abbreviate the second and third terms by noting that $\rho_a + \rho_b$ gives rise to the Hartree potential V_H . We obtain

$$\delta U_5 = \delta U_e(\rho_a; \rho_b) + \int V_H \delta\rho_a d\mathbf{r}_a + \int V_H \delta\rho_b d\mathbf{r}_b + \int V_H \rho_c d\mathbf{r}_c,\quad (16.19)$$

where to sufficient accuracy V_H in region c may be taken as a constant equal to its value at the cut in the undisplaced system. Finally, in the sixth term we write the change in $u_{xc}\rho$ with the help of (15.11) as

$$\delta(u_{xc}\rho) = \frac{\partial(u_{xc}\rho)}{\partial\rho} \delta\rho = \mu_{xc} \delta\rho,\quad (16.20)$$

so that δU_6 becomes

$$\delta U_6 = \int \mu_{xc} \delta \rho_a d\mathbf{r}_a + \int \mu_{xc} \delta \rho_b d\mathbf{r}_b + \int u_{xc} \rho_c d\mathbf{r}_c. \quad (16.21)$$

Adding up the contributions δU_1 to δU_6 we find most terms canceling, leaving

$$\begin{aligned} \delta U = & \delta \left(\sum_n E_n \right) + \delta U_e(N_a + \rho_a; N_b + \rho_b) \\ & + \int (V_N + V_H + u_{xc} - V) \rho_c d\mathbf{r}_c. \end{aligned} \quad (16.22)$$

We now choose the previously undefined V in region c as

$$V(\mathbf{r}_c) = (V_N + V_H + \mu_{xc}) + (u_{xc} - \mu_{xc}), \quad (16.23)$$

so that the integral over region c in (16.22) vanishes. The first bracket in (16.23) is the self-consistent field (16.13) at the cut, and the second bracket is a discontinuity in the devised potential between region c and regions a and b . All terms may to sufficient accuracy be taken as constant through region c as before. When the displacement causes an overlap between regions a and b instead of a gap, as on the right in Fig. 25, the overlap becomes region c counted negatively, i.e., the discontinuity is reversed in sign.

The final result constituting *Andersen's theorem* is therefore

$$\delta U = \delta U_e(N_a + \rho_a; N_b + \rho_b) + \delta \left(\sum_n E_n \right). \quad (16.24)$$

It is necessary to interpret this precisely. The first term is simply *the electrostatic force across the boundary* between the two (usually electrically neutral) pieces, regions a and b . Note that U_e here is the genuine electrostatic energy, different from the so-called electrostatic energy U_{es} in (13.3); the latter in our present notation would be $U_e(N_a; N_b) - U_e(\rho_a; \rho_b)$. In the atomic sphere, approximation δU_e is zero because each sphere is electrically neutral giving no field outside it, but in a shear it should be given by an Ewald-like energy defined for a charge density equal to that at the cell boundary for the same reasons as given in Section IV,13. In some cases where atomic cells are only slid over one another, it would be just an Ewald energy, but a modified quantity would be required where the displacement produces gaps and overlaps as in Fig. 25. The second term in (16.24) is the interesting one, *the change in the one-electron levels*. The E_n for the undisplaced system are straightforward, with the levels $E_n + \delta E_n$ calculated from an artificially constructed potential as defined above.

The theorem is interesting in that the *double-counting* term of (13.3) and (15.12) is absent, as is the relaxation in self-consistent field due to the displacement. Indeed these two effects cancel, and stated that way the theorem is not so remarkable. Let $\delta\rho$ be the change due to the self-consistent rearrangement of charge. Then the change in the double-counting term of (13.3) and (15.12) is

$$-\int \frac{\rho(\mathbf{r}) \delta\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3\mathbf{r} d^3\mathbf{r}', \quad (16.25)$$

which from (15.7) is the change in $\sum E_n$ due to δV_H by first-order perturbation theory. Thus we must exclude this effect in $\delta(\sum E_n)$ if we omit the change (16.25) in the double-counting electrostatic term, and the same applies to the double-counting of exchange and correlation energy. *This cancellation is perhaps the most important conclusion of the present analysis because it helps to explain qualitatively the success in using the one-electron energy $U_1 = \sum E_n$ to calculate total energy changes* (see Section IV,13).

So far the theorem has only been applied to uniform volume changes.^{126,190,191,193,203} In the atomic sphere approximation, the energy values are determined by the logarithmic derivatives $D_l(E)$ (16.4) at the sphere radius s as described in Section 15 of Bullett's chapter.¹⁵ The change in $D_l(E)$ is obtained by simply integrating the radial wave equation out from s to $s + \delta s$ in the potential $\mu_{xc}(s)$. Indeed, we have the identity^{190,204}

$$\begin{aligned} \partial E / \partial s &= u_l^2 \partial D_l / \partial \ln s \\ &= -[E - \mu_{xc}(s)]s^2 - (D_l - l)(D_l + l + 1), \end{aligned} \quad (16.26)$$

so that we recover the previous result (16.3). Note that there $\frac{1}{3}u_{xc}$ is $u_{xc} - \mu_{xc}$ in Slater's $\rho^{1/3}$ approximation. We conclude that *the force theorem* (16.24) is equivalent to Liberman's local formulation (16.2) of the electronic pressure in the medium, and indeed Pettifor¹⁹⁰ had already used the transformation (16.26) backward to derive the energy formulation (16.24) in the particular case of uniform compression.

Can the force theorem be used to calculate *elastic constants*? The point at issue is that the theorem is valid to first order in δR , whereas an elastic constant or the force constant for the atom in Fig. 25 is a second derivative of U . To calculate the bulk modulus it is necessary to calculate the self-consistent field at one volume V_1 and then from the force theorem the pressure P_1 by a small volume change. It is then necessary to do a second self-consistent

²⁰³ H. L. Skriver, O. K. Andersen, and B. Johansson, *Phys. Rev. Lett.* **41**, 42 (1978).

²⁰⁴ Andersen,¹²⁵ Eq. (2.31) and the equation immediately below it. Also Eq. (77) of Makintosh and Andersen.¹²⁶

calculation at a second volume V_2 and calculate P_2 there. The two pressures can be fitted to the relation

$$P = (B/V_0)(V - V_0) \quad (16.27)$$

to determine both the bulk modulus B and equilibrium volume V_0 .

The procedure is a little simpler for a shear constant or any situation such as in Fig. 25 where the equilibrium position is dictated by symmetry. Let us consider a displacement x , say 10% from equilibrium. We can certainly determine the force $F(x)$ by calculating the self-consistent potential $V_{\text{SCF}}(\mathbf{r}; x)$ at displacement x and then δU from our theorem using a further displacement $\delta x = 1\%$, say. Then F/x is the required force constant since $F(x)$ must be linear in x . Thus only one self-consistent field need be calculated, through that with the displacement x . In the approximation of rigidly displaced muffin tins or atomic spheres the latter potential is constructed simply from the V_{SCF} for the symmetrical ground state. In such a case it in fact is sufficient²⁰⁵ to calculate U at $x = 0$ and at one finite $x = x_0$, say, and then to fit a parabola

$$U(x) = U(0) + \frac{1}{2}Cx^2 \quad (16.28)$$

to obtain the force constant C . We can prove that in the approximation of rigidly displaced potentials, this gives the same result as the use of the theorem to calculate $F(x_0)$ and hence $C = F(x_0)/x_0$. Consider $U(x_0)$ and the calculation of $U(x_0 + \delta x)$ by means of the theorem. In the latter calculation, the potential in region c occurs once positively and once negatively and so may be ignored in calculating $F(x_0)$. The rest of the potential is defined by rigidly displacing the $V(\mathbf{r}; x_0)$. But since in our approximation the latter is itself defined by rigid displacement from $V(\mathbf{r}; x = 0)$, our $U(x_0 + \delta x)$ calculated in the theorem is also a point on the parabola $U(x)$ obtained by calculating U over a range of values of x with the rigid displacement approximation. Thus the theorem gives the slope of $U(x)$ at x_0 and this together with the fact that $U(x)$ has a minimum at $x = 0$ is sufficient to determine the parabola uniquely. The rigid displacement of muffin tin potentials is probably a very good approximation for the d bands of transition metals, but in other systems such as SnTe²⁰⁶ the valence charge flow induced by the displacement x_0 can have important effects.

c. Review of One-Electron Approximations

Finally we return to comment on the Wigner approximation of Section III,8 and the use of one-electron energy U_1 in Section IV,13, in the light of the more quantitative analysis of the present section.

²⁰⁵ This argument is due to O. K. Andersen, private communication.

²⁰⁶ P. B. Littlewood and V. Heine, *J. Phys. C* **12**, 4431 (1979).

First consider, for example, the calculation of the energy difference between the fcc, hcp, and bcc structures for the transition metals discussed in Section IV,13. The discussion below (16.25) suggests that to first order the difference in the self-consistent field cancels out if one calculates U_1 with the same potential for all three. This difference in self-consistent potential is the effect we worried about at the end of Section IV,13 and found to be small. It is now seen to cancel out to first order by the argument of (16.25). Of course, the δU_e term in (16.24) or δU_{es} in (13.3) remain. Both are short-range, repulsive interactions. It is not clear physically why one should occur in (16.24) and the other in (13.3), or indeed how much they differ.

Second, let us consider the Wigner trick (Section III,8) for the exchange-correlation hole, including its extension by Bullett⁸¹ to spatially separated sp^3 bonding orbitals in the diamond structure or more generally (Section III,8). The $V_{SCF}(\mathbf{r})$ inside one atomic cell is simply that due to a singly charged ion, which considered as a function of \mathbf{r} is independent of atomic sphere radius s or interatomic separations. Thus its use at two values s and $s + \delta s$ automatically creates the potential specified for calculating the levels $E_n + \delta E_n$ in (16.24), aside from the jump in potential in region c . For shears the gaps and overlaps are of equal size, i.e., the positively and negatively counted regions c [see below Eq. (16.23)], and the jump in potential $\mu_{xc} - u_{xc}$ there cancels to lowest order but this is not true of bulk moduli. Thus shear constants, and in general conformational rearrangements, should be given well as noted in Section IV,13.

Third, there is the exchange and correlation term $u_{xc} - \mu_{xc}$ in (16.23), which defines the discontinuity of potential in region c . It cancels out for the energy differences among, say, the fcc, bcc, and hcp structures if these are compared at the same net volume, as already remarked in connection with shears. For bulk moduli the term remains, but it is small in the case of transition metals because it involves the charge density found at the cell boundary s both directly in $\rho(s)$ and in $u_{xc}(s) - \mu_{xc}(s)$: the $\rho(s)$ is much less than the average electron density, so that the contribution

$$\int [u_{xc}(s) - \mu_{xc}(s)]\rho(\mathbf{r}_c) d^3r_c \quad (16.29)$$

may not be very significant.

In conclusion, the success of some of the approximations of Sections III,8 and IV,13 can be traced in certain situations to exact or approximate cancellations revealed explicitly by the theorem (16.24).

17. SELF-CONSISTENCY IN THE POTENTIAL

With modern computers it is possible to carry out electronic structure calculations to self-consistency in the potential for simple perfect solids and

for very small atomic clusters. Even with these a large amount of computing time is required. However, it is easy and quick to achieve self-consistency within the atomic sphere approximation,^{120,125} and this substantial degree of self-consistency is probably quite adequate in some cases, e.g., close-packed transition metals, while being perhaps insufficient for describing the bonds in the diamond-structured semiconductors. Aside from charge transfer in compounds, self-consistency is achieved in each atom separately, which is the reason for including the topic in the present review.

There are two essential ingredients to the atomic sphere approximation. As the name of the method implies, the Wigner–Seitz atomic cell of each atom is approximated by a sphere, with the charge density ρ and potential being spherically symmetric inside. Let us suppose we are calculating the d band of a transition element or a compound of it. To a zeroth approximation we can take some typical energy in the band, e.g., E_0 in the sense of Section III,10, and take the wave function run out from the radial wave equation for that energy as typical for all the d electrons on the atom. The latter number can be obtained from the d-electron local density of states determined by a conventional band structure calculation or by the methods of Section IV,11. Either way, self-consistency can be achieved quite simply because all d electrons have been lumped together into an average wave function and some number of them taken.

The above procedure is too crude, but it can be developed. The charge density $|\psi|^2$ of an electron state depends on its energy E in the band. For example, Fig. 13 shows how the radial wave function changes with energy from the most bonding state at the bottom of the band to the antibonding state at the top. Thus we can write for the radial wave function

$$u_i(r, E) = u_i(r, E_0) + (E - E_0)\dot{u}_i(r, E_0), \quad (17.1)$$

where

$$\dot{u}_i(r, E_0) \equiv [\partial u_i(r, E)/\partial E]_{E=E_0}, \quad (17.2)$$

and, indeed, Andersen includes the next term in $\ddot{u}_i(r, E_0)$ in (17.1). The corresponding charge density has the form

$$\rho(r) = Au_i^2 + Bu_i\dot{u}_i + C\ddot{u}_i^2, \quad (17.3)$$

where A is equal to the total number of d electrons as before, B depends on the local density of states on the atom through

$$B = \int_{-\infty}^{E_F} (E - E_0)n_i(E) dE, \quad (17.4)$$

and C is usually neglected. Thus the charge density ρ can be represented very simply in terms of two (or three) functions $u_i(r, E_0)$ and $\dot{u}_i(r, E_0)$ [and perhaps

$\ddot{u}_i(r, E_0)$] and hence the potential calculated. For a very wide band structure one may divide it into two or three energy ranges in each of which such an expansion (17.1) can be applied.

The second important aspect of the atomic sphere approximation is the separation between structure and potential. In Section III,10 we saw that all three $dd\sigma$, $dd\pi$, $dd\delta$ hopping integrals are proportional to the parameter M^2 . Thus the band structure has a *canonical form*, which can be calculated once and for all for the desired configuration of atomic positions and then scaled by M^2 , and the same applies to integrals such as (17.4). We see that the coefficients in (17.3) are fixed by the canonical band structure, aside from a scaling factor. When going around the potential wave function loop to achieve self-consistency, it is not necessary to integrate explicitly over the occupied states in the band each time, only to scale the canonical result (calculated once) by a scaling factor determined by the logarithmic derivative and hence by the potential. In this way achieving self-consistency for the solid is akin in computation to self-consistency in a single isolated atom. That represents an enormous simplification, applicable to any solid with or without lattice periodicity.

We have sketched only the main idea of how self-consistency is achieved. Bullett in Section 15 of his chapter¹⁵ has given a somewhat more rounded view of the whole atomic-spheres method, but we refer the reader to the original sources^{120,125,126} for details. Application has been made in Andersen *et al.*^{118,120,125,126,132,193,203} among those cited in the present review.

VI. Historical Note

18. THE DEVELOPMENT OF THE CONCEPTS OF LOCAL ELECTRONIC STRUCTURE

There is a fashion in science to hide the process through which scientific discoveries are made and to expose only the results. This does a disservice both to the public understanding of scientific development and to new entrants to the profession. It may therefore be of interest to record how the main ideas contained in the present review emerged in the author's mind during the late 1960s and early 1970s. This is not a scholarly history of the development worldwide to adjudicate the credits for originality, but rather a picture of the organic growth in one consciousness. It will already be clear from the text of this chapter that I was not an originator of ideas that first opened chinks in the wall holding up progress. My role was to some extent that of cheerleader and campaign manager on the battering ram to finish the job off!

Two circumstances helped prepare me. When first I entered university I had intended to study chemistry, my first love from secondary school, so that the local point of view of atoms and bonds and valence orbitals, as throughout the present chapter, was already well established. In any case I have a very pictorial mind (in real space, not \mathbf{k} space!) and think about the forces the electron feels as it moves around and between the atoms of a solid, guided by the Schrödinger equation.

More importantly, Jacques Friedel¹¹ had originated the two main ideas of the present review in connection with his theory of random dilute alloys, namely the use of the *local density of states* $n(E, \mathbf{r})$ instead of wave functions and the *invariance property* that it has, as described in Sections I.1 and I.2. To him belongs the greatest credit. I was schooled in his work in the mid 1950s because he maintained personal contact with his previous research supervisor, Professor (now Sir Nevill) Mott who was then also my supervisor. My desk was in the Royal Society Mond Laboratory (a part of the Cavendish Laboratory) led by Professor David Shoenberg, and as a sideline to my main work interpreting the de Haas–van Alphen effect in aluminum I looked at his measurements with Uddin²⁰⁷ of the de Haas–van Alphen effect in bismuth alloys. It gave me first-hand experience²⁰⁸ of Friedel's concepts.

The next advance was (now Sir Sam) Edwards' application of *Green function* techniques to one-electron problems as a spin-off from the development of many-body theory. Friedel's local density of states became the imaginary part of the Green function, which could be solved for even in disordered systems.²⁰⁹ In Cambridge we were aware of Edwards' work through the considerable local interest in molten metals and other forms of disorder.²¹⁰

For over ten years there was otherwise little new development of Friedel's ideas. However, I had found myself appealing to the invariance theorem (Section I.2) more and more in personal discussion of a wide diversity of topics and explicitly in two or three pieces of research, particularly the similarity between the tails of the metal wave functions on the semiconductor side of a metal semiconductor junction and the surface states on a free semiconductor surface (Section II.6 and Heine²⁹). Another case was Mott's return from a conference on soft x-ray spectra, perturbed that the bandwidths measured on the two constituent elements were apparently not equal as expected, but more characteristic of the pure elements.^{25,26} My

²⁰⁷ D. Shoenberg and M. Z. Uddin, *Proc. R. Soc. London, Ser. A* **156**, 687 and 701 (1936).

²⁰⁸ V. Heine, *Proc. Phys. Soc., London, Sect. A* **69**, 505 and 513 (1956).

²⁰⁹ S. F. Edwards, *Philos. Mag. [8]* **6**, 617 (1961); *Proc. R. Soc. London, Ser. A* **267**, 518 (1962).

²¹⁰ See, for example, W. D. Knight, A. G. Berger, and V. Heine, *Ann. Phys. (N.Y.)* **8**, 173 (1959); J. M. Ziman, *Philos. Mag. [8]* **6**, 1013 (1961); N. F. Mott and W. D. Twose, *Adv. Phys.* **10**, 107 (1961).

qualitative explanation in terms of the invariance theorem has already been repeated in Section I,2, backed up by Jacobs²⁷ and Inglesfield's²³ subsequent calculations. With the advent of pseudopotential theory one had tools for describing electrons in sp-bonded nearly free-electron alloys, but with the serious defect discussed at the end of Section I,2: the theory for a MgAl alloy, for instance, had to start from a free-electron gas of 2.5 electrons per atom, instead of from something looking like magnesium and aluminum atoms, resulting in a perturbation series not as rapidly convergent as one would like. In such a theory there could be no way of obtaining quantities such as the heat of formation directly in terms of the change from element to alloy, and this state of stalemate is evident in the review³ by Weaire and myself on the cohesion and structure of sp-bonded metals.

This frustration led me to sharpen up my ideas of what one was looking for and I had got as far as Eq. (2.2):

$$\begin{aligned} G \text{ (in cell A in an AB alloy)} &= G \text{ (in pure metal A)} \\ &\quad + \text{(correction terms from nearby B atoms),} \end{aligned} \quad (18.1)$$

so that the invariance theorem is contained as the zeroth approximation. I always want a mathematical formalism that parallels the physical concepts as closely as possible. Being by temperament mainly interested in the application of theory and (foolishly) slightly disdainful of formalism per se, I did not have the mathematical tools or intuition to obtain the solution. However my mental preparation ensured that I instantly recognized the next clue toward the desired goal when I came across it: in fact I pounced like a vulture. On a brief visit to Professor Garcia-Moliner's group in Madrid (he had overlapped with me in Cambridge) he showed me a formalism for calculating surface states by himself and Dr. Juan Rubio.³⁷ I found it a jungle of *ad hoc* algebra that I never did fight my way through (the reader must interpret this in the light of my nonmathematical mind) all related at that stage to even and odd solutions with respect to a plane interface. But in conversation I learned that the idea was to match at the boundary the Green functions from the vacuum and the solid, as I had previously matched wave functions.^{29,50} That immediately connected it with the idea of Eqs. (2.2) and (18.1), and I satisfied myself in a couple of days that the method worked in practice for the surface-state energy in a one-dimensional model.³⁴ I was convinced it must be possible to generalize the theory to an arbitrarily shaped boundary and to solve for the whole Green function in the form (18.1) besides determining discrete energy levels. I wrote up this conviction with Eq. (18.1) and Fig. 6 (Section II,4) in a few pages of notes, emphasizing its relevance to the whole range of problems listed at the beginning of Section II,6: alloys, surface states, junctions, adsorbed atoms, molecules. I also noted the importance of the invariance theorem, and that a formalism like

(18.1) should provide for the first time a satisfactory proof and a statement of the correction terms (Sections I,2, II,4, and II,5). I circulated and discussed these notes among the Cambridge and Madrid groups, and Garcia-Moliner and Rubio³⁸ from Madrid and Inglesfield from Cambridge³⁵ published equivalent solutions. The subsequent applications by both groups have already been reviewed in Section II,6. This then formed the development of Friedel's and Edwards' ideas, particularly for nearly free-electron materials, though now seen as not restricted to them.

At about the same time Friedel took the step that led to the developments of Section IV,11 for electronic structure of tight-binding form.¹⁵⁸ He had shown how one could approximate to the local density of states on an atom in a strictly local way, which Dr. Françoise Cyrot-Lackmann applied first to molten transition metals²¹¹ and then to transition metal surfaces.²¹² She must have given a paper at a surface science conference at Cornell University, but I only appreciated the idea and importance of the method through sitting next to her on the airplane returning from the meeting. I had learned about the data on the atomic magnetic moments in the Fe₃Al alloy (Sections I,1 and I,2) through the experimental group of Dr. W. H. Taylor and Dr. Jane Brown in Cambridge working on magnetic alloys. I had been for some time very impressed by these data in relation to the invariance theorem as related in Section I,2. It seemed an excellent application of the new method, not competing with Cyrot-Lackmann's study of surfaces, and I suggested it to Haydock starting as a research student. At that stage the method consisted of calculating the first few moments by hand from Eq. (11.15) and then fitting a somewhat arbitrarily chosen form for the local density of states to them. Instead Haydock set about improving the method, the problem of Fe₃Al being passed in turn some years later to his student Mike You.⁷ The crucial factor was that Haydock transferred to the group from theoretical high-energy physics with a much stronger mathematical background, interest, and imagination than myself. The role of teacher and pupil were largely reversed: he was producing mathematical ideas so fast that at one time I found myself working to evaluate and develop them and discriminate between them! It was clear that the recursion method,^{138,139} when Haydock produced it in his second year as a research student, would have an impact on several lines of interest in the growing solid state theory group in Cambridge.

M. J. Kelly, also as a research student, took over as impresario for applications of the new method. While earlier interests of the group had been mainly in surfaces and metallic cohesion, he put together the collaboration between

²¹¹ F. Cyrot-Lackmann, *Adv. Phys.* **16**, 393 (1967).

²¹² F. Cyrot-Lackmann, *J. Phys. Chem. Solids* **29**, 1235 (1968).

Haydock's recursion method, Bullett's work on tight-binding parameters for silicon bonds, and the atomic coordinates (as yet unpublished) of the Polk and Connell-Temkin models of amorphous silicon²¹³ (Section IV,11). That was concerned with differentiating between rival models, respectively, containing or excluding odd-membered rings of atoms. Then through a summer school he became familiar with Grimley's theory of chemisorption and applied the method to that.²¹⁴ Thus through these two early pieces of work, Kelly broadened the range of applications of the recursion method. Subsequent development followed Kuhn's²¹⁵ "normal science" as is evidenced in the remainder of this volume.

I shall not stray to discourse on topics slightly peripheral to the main concept of local electronic structure. Brief historical references have already been included in the appropriate places on the treatment of exchange-correlation (Section III,8), Anderson and Bullett's development of the tight-binding method (Sections III,7 and III,9), the resonance theory of transition metals (Section III,10), Andersen's atomic sphere approximation (Sections III,10 and V,17), and the local formulation of pressure and force (Section V,16). Similarly, I have not presumed to describe how similar ideas undoubtedly developed in other minds in other places: such speculations would be certain to contain many errors. Finally I leave to others to draw from this narrative any lessons about temperament, education, serendipity in research, the old-boy network, international travel, Kuhnian²¹⁵ paradigms, and scientific minirevolutions: but I hope the reader has perceived it has been fun!

ACKNOWLEDGMENTS

I am grateful to the three other authors in this volume for their interactive collaboration and to C. M. M. Nex, D. G. Pettifor, and R. G. Woolley for help in various sections of this chapter. I am also indebted to many friends, speakers, and authors around the world from whom I have obtained ideas and information: I realize they often go unacknowledged because I forgot the fountain from which I drank, and I hope they will accept my sincere apology and gratitude. Likewise my main purpose in this review has been to expound the ideas, and I have not been able to keep up with all developments in all the fields touched on: I therefore extend my apologies also to those whose work might have been cited but has not been, and hope my co-authors in the other chapters will have done better than I.

²¹³ D. W. Bullett and M. J. Kelly, *Solid State Commun.* **16**, 1379 (1975).

²¹⁴ M. J. Kelly, *J. Phys. C* **7**, L157 (1974); *Surf. Sci.* **43**, 587 (1974).

²¹⁵ T. S. Kuhn, "The Structure of Scientific Revolutions," 2nd ed. Univ. Chicago Press, Chicago, Illinois, 1970.