

The Band Structure Problem*

J. M. Ziman

H. H. Wills Physics Laboratory, Bristol, BS 8, 1TL, England

(October 10, 1969)

The numerical solution of the Schrödinger equation for an electron in a dense assembly of atoms (*i. e.* a solid or liquid metal or semiconductor) has made great progress in the past ten years. This is not merely a consequence of greater computing power; we now have a much better grasp of the mathematical theory of such solutions.

By 1960 a number of practical methods had been devised for the computation of the electronic structure of ordered crystals, but these lacked intuitive interpretation. The first advance was to rewrite the OPW method in terms of pseudopotentials, thus making sense of the free-electron theory of metals. This development has proved particularly valuable in semiquantitative and empirical investigations of Fermi surfaces, transport properties, lattice dynamics, cohesion, *etc.*, but we have had to wait until recently for a rigorous analysis of the criteria for convergence of the various types of model potential or pseudopotential that have been postulated.

The next step was to show that the KKR (Green function) method could also be expressed as a pseudopotential, and then to demonstrate that this was also a form of APW expansion. The relative computational power of these two methods can thus be analyzed, and questions answered concerning the fulfillment of the empty lattice test, the apparent lack of uniqueness of the expansions, the advantages of "folding" matrix elements from distant points of the reciprocal lattice, and the introduction of contributions from the interstitial potential.

At this stage, the connections between the band structure problem and the t-matrix theory of scattering were uncovered, and d-bands were seen to arise as resonances of the muffin-tin wells. The KKR matrix could now be rewritten as a mixture of pseudopotential and tight-binding elements, in harmony with the empirical model Hamiltonian representations of hybridised s-p and d-bands. This method not only permits more rapid computations, but shows clearly how the width and position of such bands should depend on the atomic potential.

Some problems still remain. For example, present techniques do not seem adequate for first-principles calculations on molecular crystals, where the anisotropy of the interstitial potential (*i. e.* easy channels along bonds, but high hills between layers or chains) is probably the dominant feature.

As for disordered systems—we know little for certain and nothing quantitatively. The linear chain model has been fully studied but is quite irrelevant to the three-dimensional case. The present theoretical confusion is exemplified by the equiconcentration substitutional alloy in the tight-binding limit; some formulae give only one band, others allow two. Again, the very possibility of producing band gaps by diffraction of free electrons in a topologically disordered system (*e. g.* amorphous Ge) has not been demonstrated mathematically with any rigor.

Key words: APW; band structure; density of states; disordered systems; KKR; pseudopotential; t-matrix; molecular crystals.

1. Algebra vs. Arithmetic

Any scientific problem or puzzle can seem interesting and significant if one gets sufficiently involved in it: the difficulty sometimes is to persuade other people of this importance. How should we defend our interest in

the Electronic Density of States to an unprejudiced tribunal?

Not, surely, in terms of immediate use, but of long term understanding. Electrons being the glue of all "materials", their states within condensed matter are of fundamental importance. No quantitative estimate of any property of a metal, semiconductor, insulator, glass, liquid, mineral, *etc.*, can begin without information about these states. In fact, we want all the wave

*An invited paper presented at the 3d Materials Research Symposium, *Electronic Density of States*, November 3-6, 1969, Gaithersburg, Md.

functions of all the electrons outside of the closed shells—a tall order, which cannot be fulfilled by direct experiment. The next best thing is the energy spectrum, or “density of states,” although as we shall learn in the course of this symposium, that cannot always be deduced unambiguously from observed phenomena.

Progress in this field therefore depends on sound theoretical analysis of the hypothetical possibilities, as well as careful experimental investigation of the facts. The calculation of electronic band structure is thus the central mathematical problem of solid state physics. Every “exciting” topic or mysterious phenomenon—superconductivity, the Kondo effect, ferromagnetism, Fermi surfaces, the Gunn effect, Josephson tunnelling, *etc.*,—eventually depends for its computable parameters on this mundane task.

It is sometimes argued, by the deeply unimaginative follower of scientific fashion, that this problem has been solved long ago, and can safely be left to the brute strength of more and more powerful computers. This is quite wrong; these elephants must be goaded and guided by experienced *mahouts*, whose skill is to see in advance the type of answer that is to be obtained, and then to deploy the minimum of force to lever away the obstacles. A ream of computer print-out is useless unless it agrees so perfectly with experiment that we need never look back and see why and how it went wrong. Our task is to devise techniques for the theoretical mastery of ever more complex systems, which requires at every stage that we know exactly what we are doing, analytically as well as numerically.

This is well illustrated by the recent history of our subject. Let me express this in personal terms. A little more than ten years ago, in gathering material for a monograph [1] of which one chapter covered this topic, I found that many techniques of band structure computation had been proposed and tried out, but that there were very few cases where the results had been confirmed experimentally, or where they gave any insight into the actual electronic structure of the materials. It was simply not obvious, for example, that almost all the calculated band structures for metals could have been derived from the free electron system by perturbation effects at the zone boundaries, because nobody had programmed his computer to print out the data in that form. We knew from the success of the free electron *model* that this could not be very far from the truth, but we had not the imagination to rewrite the algebra so as to see how this must arise within whatever method of calculation we might happen to use.

In the past decade, of course, computational techniques have improved enormously in accuracy and power, so that a whole body of expertise is now availa-

ble for application in any particular case [2]. Given the exact one-electron potential of a crystalline solid, we can compute the band structure to almost any desired degree of accuracy. But the trouble is that we do not always have this potential, complete with all the electron-electron terms, spin-orbit interaction, core polarization, exchange and correlation effects and so on, so that our first-principles computations just miss the answers we are seeking. Without an appeal to the basic algebraic principles and governing features of the model, we then flounder around, trying to adjust the parameters by trial and error. Somebody else, using a different “method” may get a different answer: is this due to deep discrepancies in the fundamental assumptions, or to errors of approximation, or just numerical mistakes?

2. Pseudism

Now recall how our minds have been liberated by the pseudopotential concept. There is no need to explain this to the present audience. Let us suppose that we had tried to express the Bloch function of wave vector \mathbf{k} by a sum of simple plane waves

$$\psi_{\mathbf{k}} = \sum_{\mathbf{g}} \alpha_{\mathbf{g}} e^{i(\mathbf{k}+\mathbf{g}) \cdot \mathbf{r}}, \quad (1)$$

where \mathbf{g} runs through the reciprocal lattice: we should have to solve the infinite set of linear equations

$$\{|\mathbf{k} + \mathbf{g}|^2 - \mathcal{E}\} \alpha_{\mathbf{g}} + \sum_{\mathbf{g}'} \mathcal{V}(\mathbf{g} - \mathbf{g}') \alpha_{\mathbf{g}'} = 0 \quad (2)$$

where $\mathcal{V}(\mathbf{g} - \mathbf{g}')$ is a Fourier component of the periodic potential in the lattice and \mathcal{E} is the energy of the state we are after. By rewriting the equations in terms of orthogonalized plane waves we can show that the whole problem is equivalent to solving a very similar set of equations

$$\{|\mathbf{k} + \mathbf{g}|^2 - \mathcal{E}\} \beta_{\mathbf{g}} + \sum_{\mathbf{g}'} \Gamma_{\mathbf{g}\mathbf{g}'} \beta_{\mathbf{g}'} = 0, \quad (3)$$

in which the pseudopotential components $\Gamma_{\mathbf{g}\mathbf{g}'}$ are much smaller than the original set $\mathcal{V}(\mathbf{g} - \mathbf{g}')$. Thus, the whole problem is equivalent to the perturbation of free electron waves by a weak pseudopotential and can be solved by elementary computation. For a perfect Bravais lattice the value of $\mathcal{V}(\mathbf{g} - \mathbf{g}')$ or of $\Gamma_{\mathbf{g}\mathbf{g}'}$ is a function only of the potential associated with a single atom or ion—in the language of x-ray diffraction, it is just the “atomic form factor” in the formula for diffraction by an assembly of such objects at the appropriate Bragg angle. The band structures of most ordinary metals, and many semiconductors, can be read at a glance. Not only does this provide us with an admirable

parametrization of Fermi surfaces, optical spectra, *etc.*, in perfect crystals, but can be extended to include almost all the properties of thermally excited, impure or disordered materials—electron-phonon interactions, electrical conductivity of solid and liquid metals, lattice dynamics, phase stability of alloys, *etc.* In moments of enthusiasm [3,4,5] we may perhaps be forgiven for pretending that all the problems of the theory of metals are cured by a strong dose of “pseudism”. It is a wonderful model for zeroth order calculations, and the ideal do-it-yourself kit for the enthusiastic amateur. It had the effect of turning band structure theory from a rule of thumb *technology* into an elegant *science*.

Nevertheless, the pseudopotential method is not the ultimate solution to the band structure problem. In the first place, the program of replacing the true atomic potential by a localized pseudopotential, independent of energy and momentum, cannot be fulfilled exactly. If, like Herman and his colleagues [6] one is trying to make very accurate first principles calculations, nothing is gained by rewriting the OPW equations in this form. Indeed, there is a danger that the apparent simplicity and rapid convergence of the pseudopotential equations may seduce us into further approximations which hide important effects; once having lost touch with the exact equations, we slide easily into a sloppy mess where qualitative and quantitative, first principles and parametrized, features are inextricably confused.

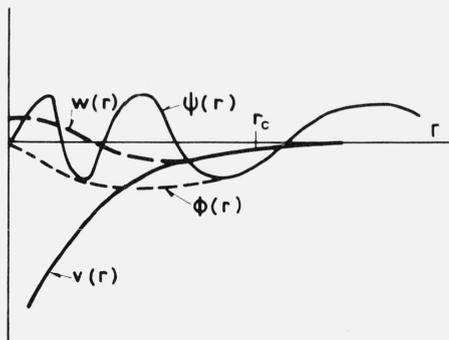


FIGURE 1. The true wave function $\psi(\mathbf{r})$ in the true potential $\theta(\mathbf{r})$ is replaced by the pseudo wave function $\phi(\mathbf{r})$ in the pseudopotential $w(\mathbf{r})$.

This type of confusion is compounded by the non-uniqueness of pseudopotentials. The original algebraic proof of this arbitrariness came as something of a surprise, but it is really quite obvious. We are asked, in effect to construct a weak potential that will reproduce the effect of a strong potential on an electron wave of given energy impinging on the atom. The boundary condition on the pseudo wave function—that it should match the true wave function on the outside—is very

weak, and amounts to little more than fixing the value of a few integrals over the pseudopotential. We know, for example, that the s-wave scattering phase shift of the true potential will be reproduced at low energies if we choose the spatial average of the pseudopotential correctly—and so on. Almost any function containing a few adjustable parameters can be made to fit these conditions. Of course the problem of finding a fixed local pseudopotential that will imitate the effects of the true potential over a wide range of energy is much more difficult, and has not been solved, but that is not what we are asked to do.

This arbitrariness was exploited to the full by Heine and Abarenkov [7] who chose the most elementary pseudopotential functions so as to simplify the rest of the algebra. It was natural to reproduce the core potential of a metallic ion with a square well of depth $A_l(\mathcal{E})$, which could be continued outwards as a simple Coulomb potential; or as a screened Coulomb potential, according as one is thinking of an isolated free atom or of a “pseudo atom” in a condensed phase (fig. 2). In fact, the value of $A_l(\mathcal{E})$ for a given angular momentum can then be estimated from the optical term values, in the tradition of the quantum defect method of Kuhn and Van Vleck.

Such a “model potential” is obviously good physics, and can be more or less justified mathematically. It copes very elegantly with one of the most difficult aspects of the whole theory—the self-consistency problem for the valence electrons—about which, for reasons of brevity, I shall say very little here. According to Shaw [8], the screening corrections can be calculated accurately, although it pays to eliminate the discontinuity at the surface of the square well by treat-

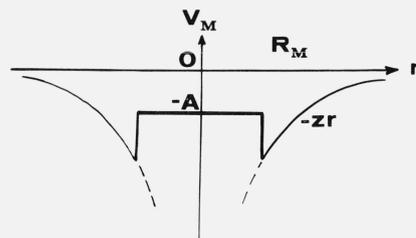


FIGURE 2a. Heine-Abarenkov pseudopotential; before screening.

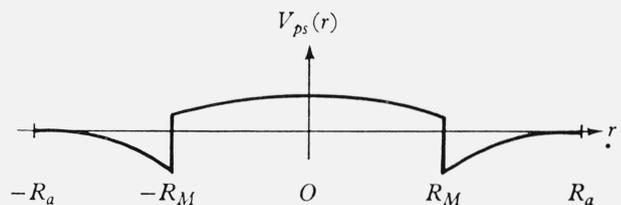


FIGURE 2b. After screening (from [5]).

ing the radius of this internal flat region as another adjustable parameter, depending also on energy and momentum (fig. 3).

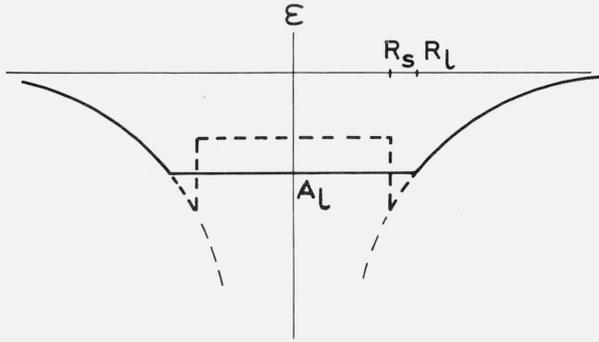


FIGURE 3. *Shaw pseudopotential.*

Notice, however, the dangers of overelaboration. An arbitrarily defined model potential in real space is valuable only in proportion to its algebraic or geometrical simplicity, and will not bear much “improvement” in the name of numerical precision or in order to get better agreement with experiment. In the event the electronic structure depends on the “form factor”—the Fourier transform of the pseudopotential—which might then just as well be derived directly from the true potential by some more powerful method, or which we could also represent by some simple empirical function [9].

From a formal point of view, the arbitrariness of the pseudopotential is certainly quite worrying. How can the electronic band structure depend uniquely on the periodic lattice potential if this arbitrary function can be interposed in the calculation? Well now, suppose we had tried to solve the equations (2) for the Bloch functions expanded in simple plane waves. Since these are an infinite set we should have had to proceed by successive approximations, just as if we are trying to sum a series term by term. But these equations really have many solutions of much lower energy than the one we are looking for, corresponding to all the narrow tight-bound bands and an expansion in powers of $\mathcal{V}(\mathbf{g}-\mathbf{g}')$ simply does not converge for energies in the valence band. We are trying to sum the Born series for scattering by one of the atomic potentials, ignoring the fact that it has numerous deep bound states. The pseudopotential trick removes all the effects of these bound states, and gives us a convergent series. It is rather like wanting to evaluate $1/(1+\chi)$ when χ is about 10: a power series in χ will not converge, but we can easily construct a new series in some new variable $y = (\chi-a)$, say, which can be made to converge in the region of interest. The actual terms in the series will depend on the value of a , which may be any arbitrary number larger

than about 5—but the final answer will be independent of this choice. Thus the final value of the energy as a function of wave vector comes out the same, whatever form of pseudopotential we introduce into the equations.

This suggests a possible criterion for a “best” pseudopotential: choose the form of $\Gamma_{gg'}$ that causes the series expansion for the Bloch functions to converge most rapidly. There is a rather elaborate mathematical theory of the Born series, due to Weinberg, which can be applied to this problem [10] and which does discriminate in principle between various formulae. These investigations are not, perhaps, of very great practical value to the horny-handed programmer of computers, but they are healthy in establishing the basic mathematical foundations of the whole technique.

3. The Problem of Bound Bands

The most serious limitation of the pseudopotential concept is that it applies only to the so-called “simple” metals—those without d-states in the valence band. There is, of course, a long tradition of representing such states by the tight binding method, as a linear combination of atomic orbitals. The coefficients a_L in such combinations then have to satisfy a set of linear equations of the form

$$(\mathcal{E}_L - \mathcal{E})a_L + \sum_{L'} V_{LL'}(\mathbf{k})a_{L'} = 0, \quad (4)$$

where the index L stands for different angular momentum and magnetic quantum numbers; for example, the five values of the component of angular momentum in a band of d-states. The original bound state at \mathcal{E}_L is broadened into a band by the various overlap integrals $\mathcal{V}_{LL'}(\mathbf{k})$, which can in principle be evaluated, although in practice this is so complicated and inaccurate that one treats them as adjustable parameters.

It used to be thought that *all* the states in metals could be described in this way, by bringing in enough different atomic orbitals. The picture of states overlapping and broadening to make nice valence and conduction bands illustrates one of the nursery rhymes of our subject (fig. 4). Unfortunately, this is quite misleading. What happens is that as the atomic *potentials* overlap, and the barriers fall between atomic cells, most of these atomic bound-state orbitals disappear. The ordinary s and p valence levels of the atoms vanish into a nearly free-electron band which can only be described if one includes propagating wave functions from above the spectrum of bound states of the separate ions or atoms.

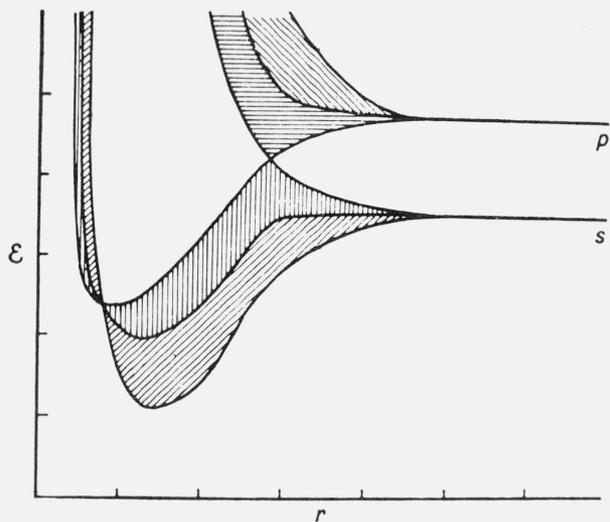


FIGURE 4. Conventional picture of energy bands from overlap of atomic orbitals.

We thus arrive at an impasse: we can describe ordinary $s-p$ bands in pseudopotential language, and d -bands in tight binding language, but there seems no common tongue, even when these bands overlap and hybridize as in the transition metals.

This difficulty never seems to have worried the active calculators of band structures: they used two techniques that gave good numerical results in all cases—the augmented plane wave method and the Green function method. One of the main developments in band structure theory in the past 5 years has been to show the mathematical connections between these useful techniques and the concepts of pseudopotential and tight-binding.

The idea of an augmented plane wave is quite simple. At some given energy \mathcal{E} , one solves the Schrödinger equation inside a spherical potential well, of radius R_s , say. The solution is a linear combination of products of radial functions and spherical harmonics of different values of angular momentum. Now determine these coefficients so that this solution matches on to a plane wave of wave vector \mathbf{k} outside the sphere. This function is still not an exact solution of the Schrödinger equation, and has a discontinuity of slope at R_s ; but we can build up our Bloch function by combining a set of these with wave vectors $\mathbf{k}, \mathbf{k} + \mathbf{g}$, etc. just as in (1) and then using the variational principle for the energy. The coefficients satisfy a set of equations exactly like the pseudopotential equations (3) so that we can find

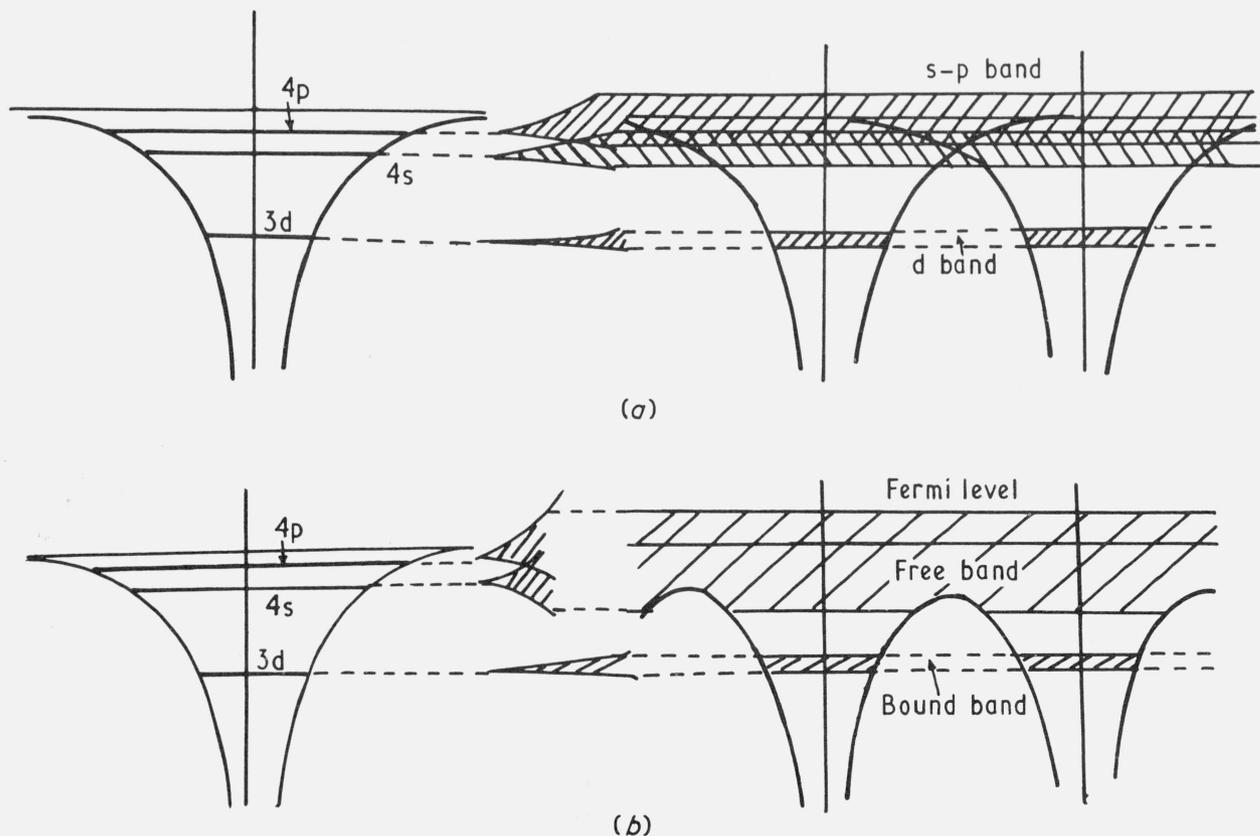


FIGURE 5. (a) Conventional LCAO description of formation of metallic conduction band; (b) Description in terms of muffin-tin potentials.

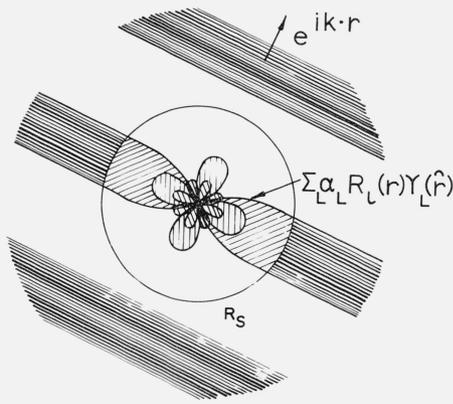


FIGURE 6. An augmented plane wave.

the energy \mathcal{E} as a function of \mathbf{k} by finding the roots of the determinant in the usual way.

The actual formula for $\Gamma_{gg'}^{APW}$ is rather elaborate, so I will not write it down; it depends upon \mathbf{k} , and also upon \mathcal{E} through the first derivatives of the radial solutions of the Schrödinger equation at R_s . At first sight one might have thought that this could be interpreted as an elaborate energy- and momentum-dependent form factor, derivable from a pseudopotential; but this is not the case. The difficulty is that Γ^{APW} does not vanish in the elementary case of an empty lattice—whereas we should certainly expect a pseudopotential to be zero when we remove the true potential to which it is supposed to be equivalent. The connection with the tight binding formalism appears even more obscure, even though one can compute perfectly good d-bands by this method.

In desperation, we turn to the KKR method of Korringa and of Kohn and Rostoker. This is called the Green function method because it was originally derived in that somewhat abstract language, but it really depends upon a self-consistency argument; as the Bloch wave proceeds through the crystal lattice, and encounters the various atomic spheres, it suffers scattering or diffraction—but this diffraction must be exactly what is needed to reproduce the wave and keep it on the move without loss. Again, I will spare you from the algebra, and merely report that, as in the APW method, one uses the radial solutions of the Schrödinger equation in

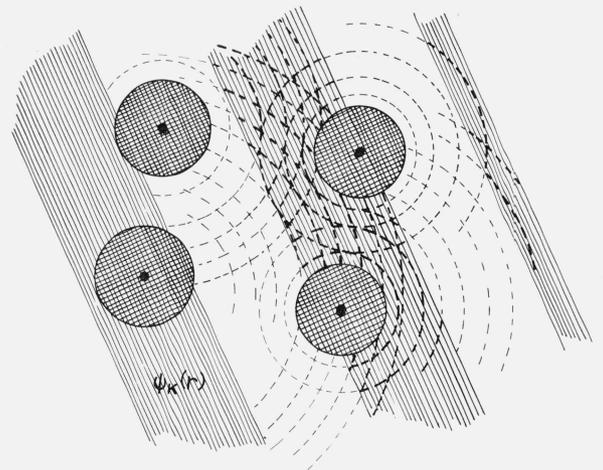


FIGURE 7. Scattered waves recombining as plane waves in KKR method.

each atomic sphere and plane waves outside. The result is yet another set of linear equations—this time for the coefficients of the mixture of solutions of various angular momentum in the sphere:

$$\kappa \{ \cot \eta_l(\kappa) - i \} b_L + \sum_{L'} B_{LL'}(\mathbf{k}, \kappa) b_{L'} = 0. \quad (5)$$

In this formula, the energy \mathcal{E} is κ^2 , and $\eta_l(\kappa)$ is the phase shift that would have been produced by the atomic sphere in scattering a plane wave of this energy. The “structure constants” $B_{LL'}(\mathbf{k}, \kappa)$ depend on the energy and momentum of the state being studied, but otherwise can be laboriously computed from the geometrical structure of the lattice.

This does not look very much like either of our previous formulae. Indeed, from the pseudopotential point of view it looks quite wrong, for when we apply the empty lattice test we make η_l tend to zero, which causes $\cot \eta_l$ to blow up. In fact these equations need to be turned upside down if we are to understand them physically [11]. The algebra is again a bit heavy, and depends essentially on some of the analytic properties of the structure constants, each of which is in fact a sum over reciprocal lattice vectors of products of spherical harmonics and Bessel functions *etc.* The result is a set of algebraic equations of the form of (3), with the following expression for the “matrix elements of the pseudopotential”:

$$\Gamma_{\mathbf{g}\mathbf{g}'}^{\text{KKR}} = -\frac{4\pi N}{\kappa} \sum_l (2l+1) \tan \eta_l' \frac{j_l(|\mathbf{k}-\mathbf{g}|R_s) j_l(|\mathbf{k}-\mathbf{g}'|R_s)}{|j_l(\kappa R_s)|^2} P_l(\cos \theta_{\mathbf{g}\mathbf{g}'}) \quad (6)$$

where

$$\cot \eta_l' = \cot \eta_l - \frac{n_l(\kappa R_s)}{j_l(\kappa R_s)}. \quad (7)$$

In this formula, j_l and η_l are spherical Bessel functions, and $P_l(\cos\theta_{gg'})$ is the ordinary Legendre polynomial for the angle between vectors $\mathbf{k}-\mathbf{g}$ and $\mathbf{k}-\mathbf{g}'$.

This formula is highly instructive, for a number of reasons.

- (i) Consider an empty lattice, for which $\eta_l=0$. Then η'_l will also vanish, and with it $\tan \eta'_l$. Thus $\Gamma_{gg'}$ is a genuine pseudopotential, which goes to zero with the true potential.
- (ii) When η_l is small, the difference between, say $\tan \eta'_l$ and $\sin \eta_l \exp(i\eta_l)$ is negligible. Ignoring the ratios of spherical Bessel functions, $\Gamma_{gg'}$ looks just like a scattering amplitude for the effect of our given potential on a single plane wave. This is good physics: the crystal is made up of an assembly of objects, each of which scatters the Bloch wave into itself.
- (iii) A strong potential, with many deep bound states may, nevertheless, have quite small phase shifts, so may behave like a weak pseudopotential. Thus, the principle of subtracting away the divergences due to the bound states amounts to simply representing each phase shift as the smallest possible angle, modulo (π) . This is a well-known property of phase shifts.
- (iv) As shown by Lloyd [12], this form of matrix element can be derived from a simple model potential. We merely put a delta function singularity of potential over the surface of the sphere of radius R_s , of strength to match the phase shift η_l outside, for each value of l .
- (v) The connection with the APW formula was discovered by Morgan [13]. Suppose we write $\Gamma^{APW}(0)$ for the values of the APW matrix elements in an empty lattice. Then

$$\Gamma^{APW} = \Gamma^{KKR} + \Gamma^{APW}(0). \quad (8)$$

The APW matrix elements have these extra parts to them, which do not really contribute to the band structure, and which do not vanish for any value of l , even for empty space. One can even derive Γ^{APW} from a model potential [12], but this is much more complicated in form than the one for Γ^{KKR} and does not vanish in empty space.

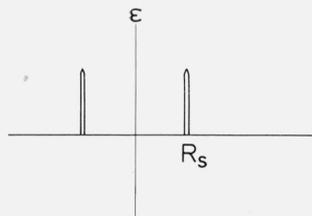


FIGURE 8. Pseudopotential for Γ^{KKR} .

These properties of this new form of pseudopotential suggest that it should be much easier than the APW method to use in practice for simple metals, where we need only introduce small phase shifts for a few values of angular momenta. We may also use the computational device of "folding" the determinant for large values of $\mathbf{g}-\mathbf{g}'$, as if we were treating the diffraction from distant zone boundaries as a small perturbation [14]. This form is also said to be the best for convergence of the Born series in the Weinberg sense [10], whatever that may imply. But the whole question of the relative computational efficiency of these methods and their minor variants is quite complicated; all I would say here is that the effort of comparing them is made much more fruitful when we understand the basic algebraic connections.

One further mystery needs clarification. Let us recall that the basic algebraic equations (3) are for the purpose of discovering the coefficients β_g in some expansion of the wave function in the appropriate plane waves. Thus, if we had been using Γ^{APW} in these equations, we should have been writing

$$\psi_{\mathbf{k}} = \sum_{\mathbf{g}} \beta_{\mathbf{g}} \phi^{APW}(\mathbf{k} + \mathbf{g})$$

where $\phi^{APW}(\mathbf{k} + \mathbf{g})$ is augmented plane wave having the form $\exp\{i(\mathbf{k} + \mathbf{g}) \cdot \mathbf{r}\}$ outside of the atomic sphere. Now it turns out [13] that the KKR equations also suppose that the wave function has been expanded in augmented plane waves — but since the matrix elements (8) are different in these equations the coefficients β_g will be different. In other words, the Bloch function $\psi_{\mathbf{k}}$, which is supposed to be a unique solution to our band structure problem, has two entirely different representations in terms of the same set of basic functions.

This is permissible, because in fact we are only combining APW's to satisfy the Schrödinger equation *outside* the spheres; the part within each sphere is automatically determined by its adjustment to the boundary condition [15]. It is well known that a periodic function defined over only part of the unit cell can be

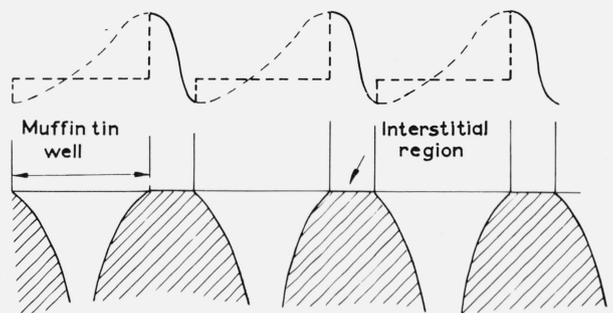


FIGURE 9. Function defined as Bloch wave in interstitial region may have arbitrary form in muffin-tin well.

represented by many Fourier expansions, depending on what properties it is allowed to have in the excluded region. The APW and KKR expansions both represent ψ_k correctly – yet they are not made up of exactly the same combinations of simple plane waves in the interstitial regions. This point is perhaps worth emphasizing because in either case we have a very explicit representation of the wave function of the Bloch state, in a form that is quite convenient for calculations of electron-electron interactions, self-consistency of potentials, and optical, x-ray, photoemission, and positron-annihilation matrix elements, *etc.*

It has sometimes been held against the APW & KKR methods that they can only be used for a “muffin-tin potential” – *i.e.* for a periodic lattice of spherically symmetric wells with “empty space” in between. But this is not an absolute restriction. Suppose there really is a significant nonconstant potential \mathcal{V}_i in the interstitial region. Then we can take this into account by adding to $\Gamma_{gg'}$ the corresponding Fourier component $\mathcal{V}_i(\mathbf{g}-\mathbf{g}')$ of this potential – made explicit by being given a constant value across the mouths of the muffin-tin wells [16]. Thus, the level which I call the “muffin-tin zero” [17] cuts across the equipotential surfaces, producing muffin-tin wells with bound states, which are eliminated by a pseudopotential device, and ranges of weak potential hills through which the valence electrons easily tunnel, and which can be represented adequately by their Fourier transforms. If we go further, and suppose that this interstitial potential had been produced by the superposition of screened Coulomb potentials, or charge clouds, carried by the individual atoms, then we can imagine \mathcal{V}_i analysed into these spherically symmetrical constituents arranged in a lattice, and reassign these to the corresponding muffin-tin wells, whose deep potentials have by now been replaced by a model potential or pseudopotential. In other words, we arrive back precisely at the sort of analysis implied by figure 2 or figure 3: the effect of the atoms on the electrons is equivalent to diffraction by an

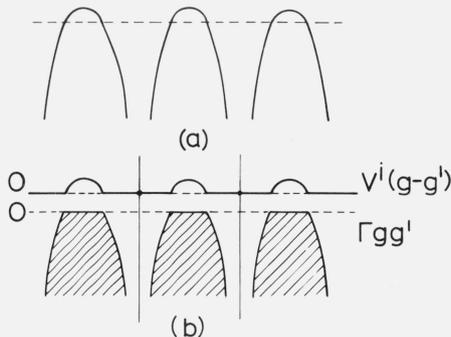


FIGURE 10. Lattice potential (a) dissected into an interstitial potential and muffin-tin wells.

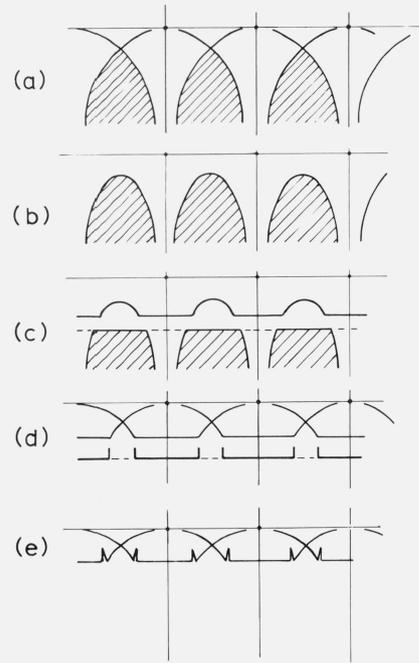


FIGURE 11. Overlapping potentials (a), summed to make lattice potential (b), dissected into an interstitial potential and muffin-tin wells (c), redefined as pseudopotentials and overlapping external parts (d), and recombined as pseudo-atom potentials (e).

assembly of screened model potentials, whose outer fields may, within reason, be superposed without hindrance. Thus we could use $\Gamma^{KKR} + \mathcal{V}_i$ as the form factor in any calculation where model potentials are employed, *e.g.* resistivity of liquid metals, lattice dynamics, *etc.*

This final demonstration of the equivalence of all three methods of band structure – OPW, APW and KKR – in the case of simple metals and semiconductors is very satisfactory, but I am now worried about one general point. Suppose we have a very anisotropic lattice – for example, the chain structure of Te, or the layer structure of graphite. The separation of the potential into muffin-tin wells and an interstitial potential must be done at a level below the lowest barriers between the atoms – for example, at the level of the potential half way between neighbors along a chain. But this may leave very high hills in the interstitial potential between the chains or layers – and the unwillingness of the electron to tunnel through such hills may not be well expressed by an expansion in plane waves in this region. Perhaps this is not a serious point after all; but I mention it to show that we are now gaining confidence to attack the electronic structure of more complex molecular crystals, a field which has up to now been dominated by an army of theoretical chemists wielding innumerable linear combinations of atomic orbitals – a

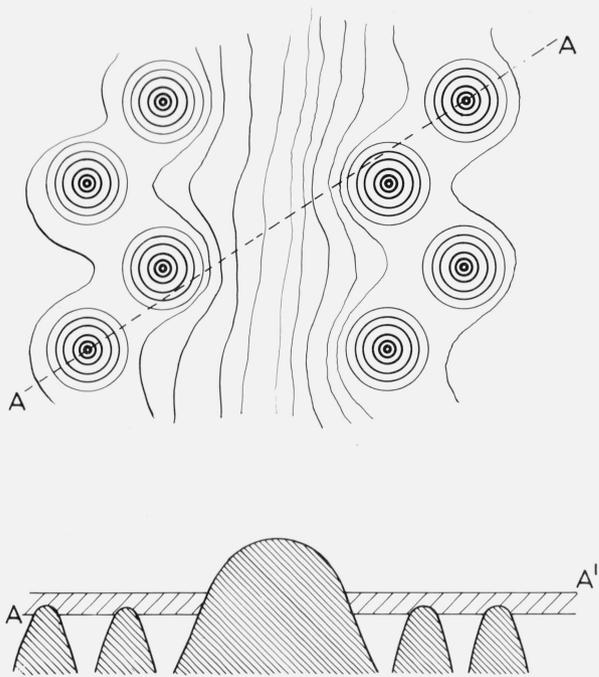


FIGURE 12. Potentials in a crystal of long chain molecules: electrons occupy the valleys containing muffin-tin wells, separated by high potential hills.

weapon whose fundamental efficacy I now take leave to doubt.

4. Resonance Bands

What about d-bands, which can be computed numerically by the APW and KKR method, but whose empirical description has usually been handled by the tight binding formula? The answer to this question is perhaps one of the most elegant results of the recent theory. Let us proceed from, say (5), the original KKR equations, which are not unlike the tight-binding equations (4), in that the index L , labelling the unknown coefficients, refers to various spherical harmonics, or

components of angular momentum. We might ask, for example, what would happen to the phase shift $\eta_l(\kappa)$ if the energy happened to coincide exactly with a bound state \mathcal{E}_l of the atomic potential. To answer this question in general, we should need to study the theory of scattering in the unphysical regions where \mathcal{E} lies below the muffin-tin zero, making κ pure imaginary; but it turns out that a factor like $\mathcal{E}_l - \mathcal{E}$ then appears in $\cot \eta_l(\kappa)$ just as we might expect. Now look at our formula (6) for the KKR pseudopotential in the reciprocal lattice representation: if $\cot \eta'_l$ were to vanish, at any energy, then $\tan \eta'_l$ would become infinite, and everything would go wrong. Thus, if η'_l should ever go through $\pi/2$ the band structure would be seriously affected.

Now this is a familiar situation in the general theory of scattering by atoms, molecules or nuclei: the phase shift η_l goes through $\pi/2$ in the positive energy region whenever there is a "resonance" of angular momentum. Thus, if the atomic or ionic potential has such a resonance, this will give rise to significant band effects in this neighborhood. There is a standard theory of such phenomena, which tells us that we may write

$$\tan \eta_l \sim \frac{W}{\mathcal{E} - \mathcal{E}_l} \quad (9)$$

for the phase shift of a resonance of width W centered on the energy \mathcal{E}_l . It is easy to show, using (6), that this has the effect of introducing a band of states of about this width, at about this energy, in the nearly-free-electron spectrum [11].

This argument can be carried further. Starting from the KKR formulae and making systematic transformations and approximations, Heine [18] showed how one could separate out a particular resonance term, and keep this in the angular momentum representation, with indices m, m' for the different components of l , while reproducing a typical pseudopotential expression in the reciprocal lattice representation \mathbf{g}, \mathbf{g}' . The matrix of these equations can thus be written in the form

$$\left(\begin{array}{ccc|ccc} \mathcal{E} - \mathbf{k}^2 & \Gamma_{0\mathbf{g}} & \dots & & & \\ \Gamma_{\mathbf{g}0} & \mathcal{E} - (\mathbf{k} + \mathbf{g})^2 & & & & \\ \vdots & & \ddots & & & \\ \hline & & & \gamma_{\mathbf{g}m} & & \\ \hline & & & & \mathcal{E} - \mathcal{E}_l & V_{mm'} & \dots \\ & & & & V_{m'm} & \mathcal{E} - \mathcal{E}_l & \dots \\ & & & & \vdots & & \ddots \\ & & & & \vdots & & \ddots \end{array} \right) \quad (10)$$

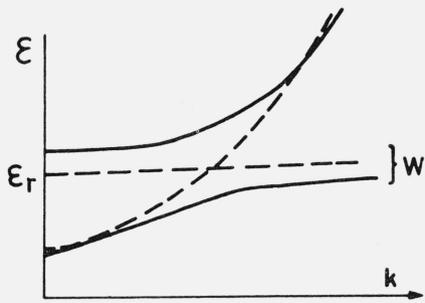


FIGURE 13. Resonance band crossing nearly free band.

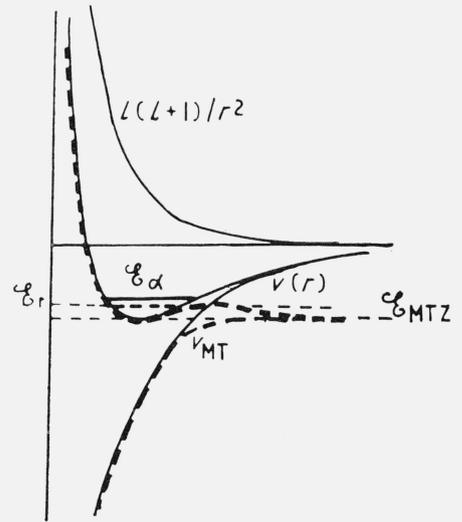


FIGURE 14. How a bound state of the atom becomes a resonance level of the muffin-tin well (See [17]).

Without the submatrices γ_{gm} *etc.*, this would factorize into a nearly-free-electron, pseudopotential matrix, such as we might expect to find in a simple metal with an s-p band, together with an ordinary tight binding matrix, corresponding to the overlapping and mixing of the 5 degenerate d-levels of the free atom. The coefficients γ_{gm} *etc.* then describe the hybridization of these two systems of states, which must necessarily occur when these bands cross one another.

As it happens (but not accidentally!) an empirical “model Hamiltonian” of just this form had already been proposed for transition and noble metals [19] before it was deduced directly from the KKR equations. We can now, therefore, justify this type of expression in principle, and even calculate the various coefficients directly from the atomic potential. In fact there are now several different versions of these equations, of varying computability, convergence and analytical simplicity [20] but all essentially equivalent of Heine’s formula [5,18].

This reinterpretation of the tight-binding formalism, and its unification with the other band structure methods is very pleasing, but to my mind there is a greater gain. Let us ask how resonances actually arise? For an ordinary one-electron potential, we need to think of the effects of the centrifugal barrier term $l(l+1)/r^2$ in the radial Schrödinger equation, which becomes important for $l=2$. A bound d-state is really constrained to avoid the nucleus by this “potential”. Now lower the ordinary potential at the outer edges of the atom: the effect may be to leave a potential dip within the core, where a “virtual”, long-lived level could still exist, even though, eventually, it would have to decay as the electron tunnelled out into free space. Thus, the original bound d-state has become a d-resonance; if the potential barrier is sufficiently thick, the resonance will be sharp; it is not surprising that the language of overlapping bound states applies to the bands produced in such cases.

From this picture we can learn a lot about the gross features of the density of states of the metal. We see,

for example, that although the little peaks and dips of the d-band can be derived from general tight-binding theory, especially when aided by group theory, the width of this complex of bands will depend chiefly on the width of the resonance, which is governed in turn by the potential barrier produced by the centrifugal force in the outer part of each muffin-tin well. Again, the actual position of this band will be determined mainly by the energy of the original d-state from which it derives—and this is fixed on a scale relative to, say, some deep state of the core. On this scale, however, the position of the ordinary conduction band does not depend on any atomic orbitals, but is determined mainly by the muffin-tin zero, which can only be calculated correctly by taking very careful account of screening, correlation energy, overlaps of potential, *etc.* We thus discover the reason for a well-known difficulty in band structure calculations—that the width of the d-band, and its position relative to the Fermi level is very sensitive to the model, and cannot apparently be calculated with the precision we would like.

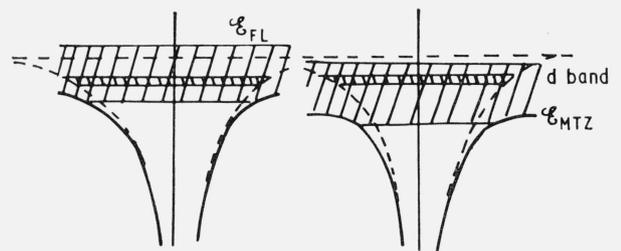


FIGURE 15. How the position of the d-band within the conduction band depends on the muffin-tin zero (See [17]).

5. Some Thoughts in Disorder

Now that we understand the electronic structure of crystalline solids so very well, we are tempted to attack disordered materials—liquids, alloys, amorphous and glassy substances. This campaign has been actively waged now for about a decade, but I am not sure that it has yielded many great prizes. The major difficulty, of course, is that we must abandon Bloch's theorem, which reduces the complexity of the problem in the perfect lattice by a divisor of the order of 10^{23} . Without crystal momentum as a good quantum number, we flounder about in a mixture of approximate algebra and incomplete intuition, hoping to find some clearcut concepts that will guide the interpretation of complicated experiments on messy materials.

It is true that the spectrum of the disordered linear array is now well understood [21]—and turns out to be much more spiky than one would have guessed from simple statistical considerations. Some of these features may persist in three-dimensional systems, but unfortunately the mathematical methods used in the one-dimensional case seem ill-adapted to generalization. In particular, real solid systems have two properties that cannot be simulated at all by a linear chain. In three dimensions, a localized defect or impurity can be avoided by a detour, so that it does not present an absolute barrier to an incident particle or excitation. In three dimensions, also we may have “structural disorder”, which is no longer topologically equivalent to any regular lattice, whereas in a linear chain the mere succession of atoms prescribes an ordering, however wildly we vary the properties of the individual potential wells.

Let me give two examples of simple cases where our present theory is inadequate. It is obvious enough that a disordered transition metal—*e.g.* liquid iron—should have a d-band arising from the d-resonance, just as in any crystalline phase of about the same atomic volume [22]. The mathematical theory of such a band is still rather uncertain [23], but there is no doubt about the physics. Suppose, however, that we make an alloy—*e.g.* of Ag and Au—whose constituent atoms have their resonance at different energies; how far apart would these energies need to be to give us two distinct d-bands, and how would this depend on the relative concentrations and relative ordering of the constituents? The model can be made extremely elementary—equal numbers of A and B type atoms, with a single bound s-state on each, substituted at random on a regular lattice with a constant overlap integral V between nearest neighbors. Some highly respected statistical theories which rely upon defining an average propagator in

such a medium, seem to insist that the bands will be drawn out into a continuous broad spectrum as the two levels move apart; others would allow a split to occur when the spacing is rather larger than the width of either band [24]. I feel sure, myself, that the latter prediction is correct, but we have still a great deal to do before we can calculate the width of each band the shape of the tails into the gap, and the nature of any levels in these regions. How far, for example, do these bands depend upon the possibilities of “percolation”, from one atom to another of the same type, through large distances—a property that depends peculiarly on the dimensionality of the lattice and the relative concentrations of the components?

Another contradiction between mathematical theories and physical intuition occurs in the case of



FIGURE 16. Does a mixed crystal have one bound band or two?

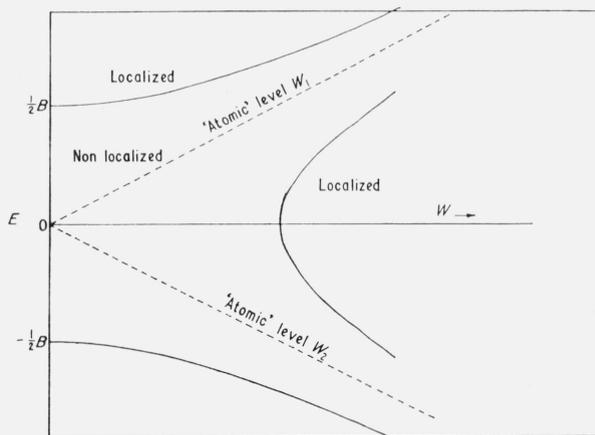


FIGURE 17. Regions of localized and non-localized states for an “equiconcentration alloy”.

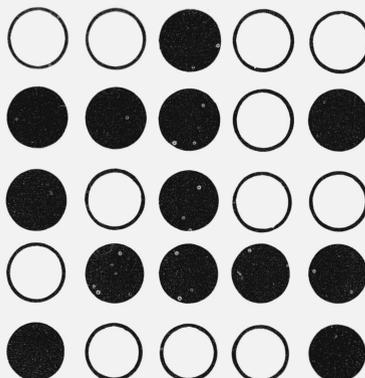


FIGURE 18. A percolation chain in an equiconcentration alloy.

amorphous semiconductors. Let it be granted, for the sake of argument, that amorphous Ge and Si are "tetrahedral glasses"; each atom has four neighbors, arranged more or less in the regular tetrahedral orientation, just as in the regular diamond lattice, but the connectivity of the structure has been altered in a random way, so that there is no long-range order. From the point of view of a chemist, this system is a single covalently bonded molecule: the saturation of all the bonds implies that some energy of excitation is required to create a carrier, so we should expect the material to be a semiconductor. The substantial gap in the optical spectrum of amorphous Ge supports this reasonable interpretation. But suppose we were to treat this by the conventional pseudopotential procedure, assigning a model potential to each atom and then calculating the diffraction effect on a free electron gas. In the absence of long-range order, there would be no strong Bragg reflections from well-defined lattice planes, and thus no proper band gaps at the zone boundaries, *etc.*; from the point of view of solid-state theory, this material ought to be a metal. This antinomy needs to be resolved if we are to understand the theory of disordered systems—or even the theory of the chemical bond. There is some evidence—as yet merely qualitative [27]—that the diffraction approach can be made to give a band gap if one takes into account the higher-order particle correlations. Thus, a glass differs from a liquid in that three neighboring atoms may have a strong tendency to be oriented so as to make a good bond angle; this is a form of short-range order, implying a strong constraint on the three-and four-body statistical distributions of atoms. At the same time, the relationship between the localized molecular orbitals of the chemical bonds and the delocalized "Bloch states" of the crystal or amorphous solid needs to be clarified [28]. But these are only two of the numerous unsolved problems in this field.

The above account of the band structure problem is obviously very sketchy and incomplete—especially in the total neglect of all electron-electron effects. We shall obviously learn much more about it as this conference proceeds. But I think it is good to look back and see what progress has been achieved—and even better

to look forward to whole Alps of ignorance still to be surmounted.

6. References

- [1] **Electrons and Phonons**: (Oxford, Clarendon Press 1960).
- [2] **Methods in Computational Physics: Vol. 8: Energy Bands of Solids**: Edited by B. Alder, S. Fernbach, M. Rotenberg (New York: Academic Press 1968).
- [3] Ziman, J. M., *Advances in Physics*, **13**, 89 (1964).
- [4] Harrison, W. A., **Pseudopotentials in Metals** (New York: Benjamin 1966).
- [5] Heine, V., **The Physics of Metals** (ed. J. M. Ziman: Cambridge, University Press, 1969) Chap. 1.
- [6] Herman, F., R. L. Kortum, C. D. Kuglin, J. P. Van Dyke and S. Skillman in Ref. 1, p. 193.
- [7] Heine, V. and I. V. Abarenkov, *Phil. Mag.*, **9**, 451, (1964).
- [8] Shaw, R. W. and W. A. Harrison, *Phys. Rev.* **163**, 604 (1967). R. W. Shaw, *Phys. Rev.* **174**, 769 (1968).
- [9] Ashcroft, N. W., *J. Phys. C*, **1**, 232 (1968).
- [10] Rubio, J. and F. Garcia-Moliner, *Proc. Phys. Soc.* **91**, 739; **92**, 206 (1967). J. B. Pendry, *J. Phys. C*, **1**, 1065 (1968).
- [11] Ziman, J. M., *Proc. Phys. Soc.* **86**, 337 (1965).
- [12] Lloyd, P., *Proc. Phys. Soc.* **86**, 825 (1965).
- [13] Morgan, G. J., *Proc. Phys. Soc.* **89**, 365 (1966). K. H. Johnson, *Phys. Rev.* **150**, 429 (1966).
- [14] Lawrence, M. J., Thesis, Bristol University (1969).
- [15] Slater, J. C., *Phys. Rev.* **145**, 599 (1966).
- [16] Scholsser, H. C. and P. M. Marcus, *Phys. Rev.* **131**, 2529 (1963). F. Belezny and M. J. Lawrence, *J. Phys. C*, **1**, 1288 (1968).
- [17] Ziman, J. M., *Proc. Phys. Soc.* **91**, 701 (1967).
- [18] Heine, V., *Phys. Rev.* **153**, 673 (1967).
- [19] Hodges, L. and H. Ehrenreich, *Phys. Letters* **16**, 203 (1965). F. M. Mueller, *Phys. Rev.* **153**, 659 (1967).
- [20] Hubbard, J., *Proc. Phys. Soc.* **92**, 921 (1967). R. L. Jacobs, *J. Phys. C*, **1**, 492 (1968).
- [21] See, *e.g.* J. Hori, **Spectral Properties of Disordered Chains and Lattices** (Oxford: Pergamon Press 1968).
- [22] Anderson, P. W. and W. L. McMillan, **Teoria del magnetismo nei metalli di transizione**: Verenna Summer School XXXVII (1967).
- [23] Lloyd, P., *Proc. Phys. Soc.* **90**, 207, 217 (1967). G. J. Morgan, *J. Phys. C*, **2**, 1446, 1454 (1969).
- [24] Ziman, J. M., *J. Phys. C*, **2**, 1230 (1969) (II).
- [25] Ziman, J. M., *J. Phys. C*, **2**, 1532 (1968) (I). D. F. Holcomb and J. J. Rehr (to be published).
- [26] Tauc, J., R. Grigorivici and A. Vancu, *Phys. Stat. Solid* **15**, 627 (1966).
- [27] Fletcher, N. H., *Proc. Phys. Soc.* **91**, 724; **92**, 265 (1967). J. M. Ziman, *J. Phys. C*, (1969) (III).
- [28] Anderson, P. W., *Phys. Rev. Letters* **21**, 13, (1968).

(Paper 74A2-596)