Section 11: Methods for calculating band structure

The computational solid state physics is a very fast growing area of research. Modern methods for calculating the electronic band structure of solids allow predicting many important properties of solids. All these methods involve the development of quite complicated computer codes. Nowadays some of such programs are available in the market and can be purchased and used by researchers.

Tight-binding approximation

Tight-binding method uses atomic orbitals as basis wave functions. Let us see how the energy spectrum gradually changes as atoms are assembled to form the solid. Consider lithium as an example. In a free atom electrons moves in a potential well, as shown in Fig. 1a. The atomic spectrum consists of a series of discrete energy levels, which are denoted by 1s, 2s, 2p, etc. The lithium atom contains three electrons, two of which occupy the 1s shell which is completely full, and the third electron is in the 2s shell.

If the two Li atoms are assembled to form the molecule Li_2, the potential "seen" by electrons is now the double well shown in Fig.1b. Due to a coupling between atoms, each of the atomic levels - that is, the 1s, 2s, 2p, etc. - has split into two closely spaced levels. We may, therefore, speak of the 1s, 2s, 2p, etc., molecular energy levels, recognizing that each of these is, in fact, composed of two sublevels. The amount of splitting depends strongly on the internuclear distance of the two atoms in the molecule: the closer the two nuclei, the stronger the perturbation and the larger the splitting. The splitting also depends on the atomic orbital: The splitting of the 2p level is larger than that of the 2s level, which is larger still than that of the 1s level. The reason is that the radius of the 1s orbital, for instance, is very small, and the orbital is therefore tightly bound to its own nucleus. It is not greatly affected by the perturbation. The same is not true for the 2s and 2p orbitals, which have larger radii and are only loosely bound to their own nuclei. It follows that, generally speaking, the higher the energy, the greater the splitting incurred.

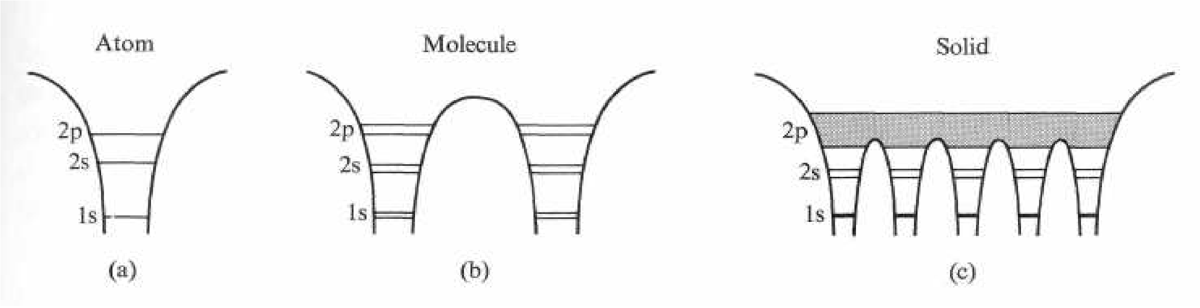

Fig.1 The evolution of the energy spectrum of Li from an atom (a), to a molecule (b), to a solid (c).

The above considerations may be generalized to a polyatomic Li molecule of an arbitrary number of atoms. Thus in a 3-atom molecule, each atomic level is split into a triplet, in a 4-atom molecule into a quadruplet, and so forth. The Li solid may then be viewed as the limiting case in which the number of atoms has become very large, resulting in a gigantic Li molecule. Each of the atomic levels is split into N closely spaced sublevels, where N is the number of atoms in the solid. Since N is so very large, about 10^{23}, the sublevels are so extremely close to each other that they coalesce, and form an *energy band*. Thus the 1s, 2s, 2p levels give rise, respectively, to the 1s, 2s, and 2p bands, as shown in Fig. 1c. The regions separating these bands are energy *gaps* - i.e., regions of forbidden energy -which cannot be occupied by electrons.

The width of the band varies, but in general the higher the band the greater its width, because a high energy state corresponds to a large atomic radius, and hence a strong perturbation, which is the cause of the level broadening. By contrast, low energy states correspond to tightly bound orbitals, which are affected but slightly by the perturbation.

The approximation which starts from the wave functions of the free atoms is known as the tight binding approximation or the LCAO (the linear combination of atomic orbitals) approximation. In general this approximation is quite good to describe the inner electronic shells of atoms and relatively localized bands. However, it is not very good for the description of the conduction electrons themselves. This is because the origin of the free electrons is completely different to the nature of the localized atomic states. Nevertheless, using tight-binding approximation it is possible to reproduce accurately the band structure of many solids including metals. Moreover, it is possible to describe reasonably well transport properties of metals as well.

Suppose that we have a solid which is composed of identical atoms which are characterized by an atomic orbitals $\varphi_\alpha(\mathbf{r})$. Here index α denote the orbital state of the atom. We assume for simplicity that we have one atom per unit cell. We are looking for the solution of the Schrödinger equation

$$H\psi(\mathbf{r}) = E\psi(\mathbf{r}), \tag{1}$$

it terms of linear combination of the atomic orbitals, so that the Bloch wave function has a form of

$$\psi_{\mathbf{k}}(\mathbf{r}) = \sum_m \sum_\alpha e^{i\mathbf{k}\mathbf{T}_m} c_\alpha(\mathbf{k}) \varphi_\alpha(\mathbf{r}-\mathbf{T}_m), \tag{2}$$

where we introduced index *m* describing the lattice vector $\mathbf{T}_m$ and the orbital located at site *m*, $\varphi_\alpha(\mathbf{r}-\mathbf{T}_m)$. The coefficients $c_\alpha(\mathbf{k})$ should be found from the Schrödinger equation. It is easy to show that (2) is the Bloch function which satisfies the requirement $\psi_{\mathbf{k}}(\mathbf{r}+\mathbf{T}) = e^{i\mathbf{k}\mathbf{T}}\psi_{\mathbf{k}}(\mathbf{r})$. Indeed,

$$\psi_{\mathbf{k}}(\mathbf{r}+\mathbf{T}) = \sum_{\alpha m} e^{i\mathbf{k}\mathbf{T}_m} c_\alpha(\mathbf{k}) \varphi_\alpha(\mathbf{r}-\mathbf{T}_m+\mathbf{T}) = e^{i\mathbf{k}\mathbf{T}} \sum_{\alpha m} e^{i\mathbf{k}(\mathbf{T}_m-\mathbf{T})} c_\alpha(\mathbf{k}) \varphi_\alpha[\mathbf{r}-(\mathbf{T}_m-\mathbf{T})] = e^{i\mathbf{k}\mathbf{T}}\psi_{\mathbf{k}}(\mathbf{r}), \tag{3}$$

Now we denote $\varphi_\alpha(\mathbf{r}-\mathbf{T}_m)$ by $\varphi_{\alpha m}$, substitute (2) in (1) to have

$$\sum_{\alpha m} e^{i\mathbf{k}\mathbf{T}_m} c_\alpha(\mathbf{k}) H\varphi_{\alpha m} = E\sum_{\alpha m} e^{i\mathbf{k}\mathbf{T}_m} c_\alpha(\mathbf{k}) \varphi_{\alpha m} . \tag{4}$$

Now we multiply this equation by $\varphi^*_{\alpha' m'}$ and take an integral over the entire volume of a solid which we denote in terms of the Dirac ket and bra vectors, so that

$$\sum_{\alpha m} e^{i\mathbf{k}\mathbf{T}_m} c_\alpha(\mathbf{k}) \langle \varphi_{\alpha' m'} | H | \varphi_{\alpha m} \rangle = E\sum_{\alpha m} e^{i\mathbf{k}\mathbf{T}_m} c_\alpha(\mathbf{k}) \langle \varphi_{\alpha' m'} | \varphi_{\alpha m} \rangle . \tag{5}$$

We assume that the basis set is orthogonal, i.e.

$$\langle \varphi_{\alpha' m'} | \varphi_{\alpha m} \rangle = \delta_{m' m}\delta_{\alpha' \alpha} . \tag{6}$$

This leads to the following equation for the coefficients

$$\sum_{\alpha} H_{a'a}(\mathbf{k}) c_{\alpha}(\mathbf{k}) = E c_{\alpha'}(\mathbf{k}) , \qquad (7)$$

where

$$H_{a'a}(\mathbf{k}) = \sum_{m} e^{i\mathbf{k}(\mathbf{T}_m - \mathbf{T}_{m'})} \langle \varphi_{\alpha'm'} | H | \varphi_{\alpha m} \rangle = \sum_{m} e^{i\mathbf{k}\mathbf{T}_m} \langle \varphi_{\alpha'0} | H | \varphi_{\alpha m} \rangle . \qquad (8)$$

The last equation came from the fact that the sum does not depend on index m'.

Eq.(7) represents a set of linear equations with respect to coefficients $c_{\alpha}(\mathbf{k})$. The number of equations is equal to the number of orbitals in an atom, N_{orb}. Therefore we obtain N_{orb} solutions. The eigenvalues and the eigenfunctions will be the functions of the wavevector **k**. Therefore, the solution will represent N_{orb} bands. In general these bands will represent mixed atomic states and can not be characterized by a particular angular momentum.

In order to solve eqs. (8) we need to know the matrix elements of the Hamiltonian. In some cases it is possible to calculate these matrix elements from atomic wave functions. In many cases however these matrix elements are considered as fitting parameters. For example, these parameters can be fitted to obtain accurate electronic band structure which is knows from more accurate calculations.

Now we consider two simple examples.

Example 1: linear chain of single-orbital atoms of lattice constant a with nearest-neighbor interactions. In this case we have just one equation (8)

$$\left[\langle \varphi_0 | H | \varphi_0 \rangle + \langle \varphi_0 | H | \varphi_{-1} \rangle e^{-ika} + \langle \varphi_0 | H | \varphi_1 \rangle e^{ika} \right] c(\mathbf{k}) = E c(\mathbf{k}) , \qquad (9)$$

and obtain for the energies:

$$E(\mathbf{k}) = E_0 - 2\gamma \cos ka , \qquad (10)$$

where $E_0 = \langle \varphi_0 | H | \varphi_0 \rangle$ is the on-site atomic energy and $\gamma = -\langle \varphi_0 | H | \varphi_{-1} \rangle = -\langle \varphi_0 | H | \varphi_1 \rangle$ is the hopping or bond integral which depends on the overlap of the wave functions of the two nearest neighbor atoms.

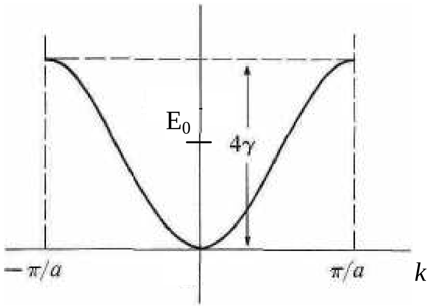

Fig. 2 The dispersion curve in the tight-binding model.

The tight-binding band is shown in Fig.2. The bandwidth, 4γ, is proportional to the overlap integral. This is reasonable, because the greater the overlap the stronger the interaction, and consequently the wider the band.

When the electron is near the bottom of the band, where k is small, we find that

$$E(\mathbf{k})-\left(E_0-2\gamma\right)=\gamma k^2a^2\,, \tag{11}$$

which has the same form as the dispersion relation of a free electron. An electron in that region of k-space behaves like a free electron with an effective mass

$$m^*=\frac{\hbar^2}{2\gamma a^2}\,. \tag{12}$$

It is seen that the effective mass is inversely proportional to the bond integral. This is intuitively reasonable, since the greater the overlap the easier it is for the electron to hop from one atomic site to another, and hence the smaller is the inertia (or mass) of the electron. Conversely, a small overlap leads to a large mass, and vice versa.

Example 2: The above treatment can be extended to three dimensions in a straightforward manner. Thus for a sc lattice, the band energy is given by

$$E(\mathbf{k})=E_0-2\gamma\left[\cos k_xa+\cos k_ya+\cos k_za\right]. \tag{13}$$

The energy contours for this band, in the k_x - k_y plane, are shown in Fig.3a, and the dispersion curves along the [100] and [111] directions are shown in Fig.3b. The bottom of the band is at the origin k = 0, and the electron there behaves as a free particle with an effective mass given by Eq.(12). The top of the band is located at the corner of the zone along the [111] direction, that is, at [π/a, π/a, π/a]. The width of the band is equal to 12γ.

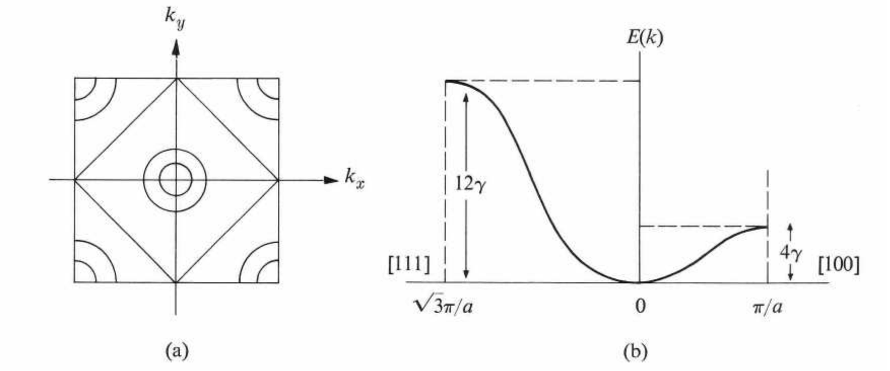

Fig. 3 (a) Energy contours for an sc lattice in the tight-binding model, (b) Dispersion curves along the [100] and [111] directions for an sc lattice in the TB model.

The cellular (Wigner-Seitz) method

The TB model is too crude to be useful in calculations of actual bands, which are to be compared with experimental results. Now we shall consider some of the common methods employed in calculations of actual bands.

The cellular method was the earliest method employed in band calculations by Wigner and Seitz. It was applied with success to the alkali metals, particularly to Na and K.

The Schrödinger equation whose solution we seek is

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r})\right]\psi_{\mathbf{k}}(\mathbf{r}) = E(\mathbf{k})\psi_{\mathbf{k}}(\mathbf{r})\,. \tag{14}$$

where $V(\mathbf{r})$ is the crystal potential and $\psi_{\mathbf{k}}(\mathbf{r})$ the Bloch function.

Within the cellular method, we divide the crystal into unit cells; each atom is centered at the middle of its cell, as shown in Fig.4. Such a cell, known as the *Wigner-Seitz* (WS) *cell,* is constructed by drawing bisecting planes normal to the lines connecting an atom *A*, say, to its neighbors, and "picking out" the volume enclosed by these planes. (The procedure for constructing the WS cell is analogous to that used in constructing the Brillouin zone in k-space.) For Na, which has a bcc structure, the WS cell has the shape of a regular dodecahedron.

In order to find the solution of Eq. (14), we assume that the electron, when in a particular cell, say *A,* is influenced by the potential of the ion in that cell only. The ions in other cells have a negligible effect on the electron in cell *A* because each of these cells is occupied, on the average, by another conduction electron which tends to screen the ion, thereby reducing its potential drastically. To ensure that the function $\psi_{\mathbf{k}}$ satisfies the Bloch form $\psi_{\mathbf{k}} = e^{i\mathbf{kr}}u_{\mathbf{k}}$, it is necessary that $u_{\mathbf{k}}$ be periodic, i.e. $u_{\mathbf{k}}$ be the same on opposite faces of the cell, e.g., points P_1 and P_2 in Fig. 4.

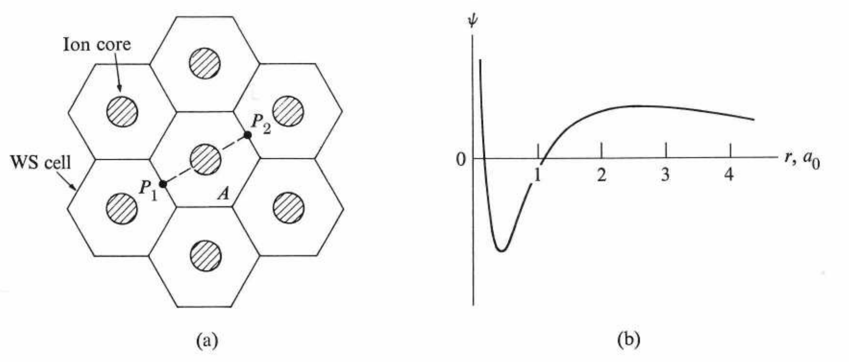

Fig. 4 (a) The WS cell, (b) The wave function ψ_0 at the bottom of the 3s band in Na.

The procedure is now clear in principle: We attempt to solve the Schrödinger equation in a single cell, using for $V(\mathbf{r})$ the potential of *a free* ion, which can be found from atomic physics. In Na, for instance, $V(\mathbf{r})$ is the potential of the ion core Na^+. It is still very difficult, however, to impose the requirements of periodicity on the function for the actual shape of the cell, and to overcome this difficulty Wigner and Seitz replaced the cell by a WS *sphere* of the same volume as the actual cell. Using these simplifying assumptions concerning the potential and the periodic conditions, one then solves the Schrödinger equation numerically, since an analytical solution cannot usually be found. The resulting wave function ψ_0 at the bottom of the band, $k = 0$, is shown in Fig. 4b. The wave functions at other values of **k** near the bottom of the band may then be approximated by

$$\psi_{\mathbf{k}} \simeq e^{i\mathbf{kr}}\psi_0 \tag{15}$$

which has the Bloch form.

One important feature of these results is the shape of the wave function in Fig.4b. The wave function oscillates at the ion core, but once outside the core the function is essentially a constant. This constancy of the wave function holds true for almost 90% of the cell volume. Thus the wave function behaves like a plane wave, as seen from Eq.(15), over most of the cell, and hence over most of the crystal. Looking at this in terms of the potential, we see that where the function is a plane wave, the potential must be a constant. Thus the *effective* potential acting on the electron is essentially a constant, except in the region at the ion core itself. Viewing the motion of the electron in the crystal as a whole, we conclude that the electron moves in a region of constant potential throughout most of the crystal; only at the cores themselves does the electron experience any appreciable potential. This result explains why the conduction electrons in Na may be regarded as essentially free electrons.

Despite its usefulness, the cellular method is greatly oversimplified, and is not currently much in use.

The augmented-plane wave (APW) method

The APW method was developed by Slater in 1937. Since the effective crystal potential was found to be constant in most of the open spaces between the cores, the APW method begins by assuming such a potential (Fig.5), which is referred to as the *muffin-tin potential.* The potential is that of a free ion at the core, and is strictly constant outside the core. The wave function for the wave vector k is now taken to be

$$\phi_{\mathbf{k}} = \begin{cases} e^{i\mathbf{kr}}, & r > r_0 \\ atomic\ \ function, & r < r_0 \end{cases} \tag{16}$$

where r_0 is the core radius. Outside the core the function is a plane wave because the potential is constant there. Inside the core the function is atom-like, and is found by solving the appropriate free-atom Schrödinger equation. Also, the atomic function in (16) is chosen such that it joins continuously to the plane wave at the surface of the sphere forming the core; this is the boundary condition here.

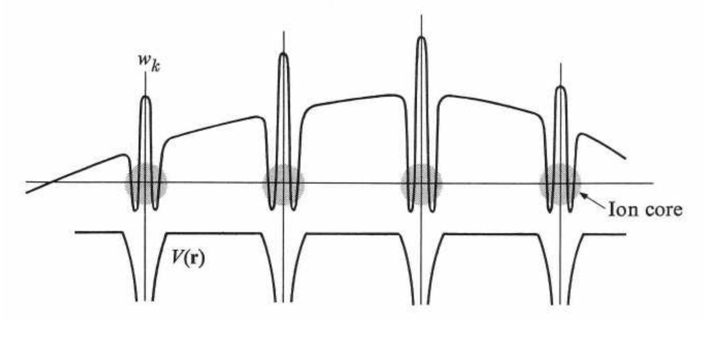

Fig. 5 The potential and wave function in the APW method.

The function $\phi_{\mathbf{k}}$ does not have the Bloch form, but this can be remedied by forming the linear combination

$$\psi_{\mathbf{k}} = \sum_{\mathbf{G}} a_{\mathbf{k+G}} \phi_{\mathbf{k+G}} \tag{17}$$

where the sum is over the reciprocal lattice vectors, which has the proper form. The coefficients $a_{\mathbf{k+G}}$ are determined by requiring that $\psi_{\mathbf{k}}$ minimizes the energy. In practice the series in (17) converges quite rapidly, and only a few terms are sufficient to give the desired accuracy.

The APW method is a sound one for calculating the band structure in metals, and has been used a great deal in the past few years.

The pseudopotential method

Yet another method popular among solid-state physicists for calculating band structure in solids is the pseudopotential method, which is distinguished by the manner in which the wave function is chosen. We seek a function which oscillates rapidly inside the core, but runs smoothly as a plane wave in the remainder of the open space of the WS cell. Such a function was chosen in the APW method according to (16), but this is not the only choice possible. Suppose we take

$$\phi_{\mathbf{k}} = \varphi_{\mathbf{k}} - \sum_{\alpha} c_{\alpha} \varphi_{\alpha} \tag{18}$$

where $\varphi_{\mathbf{k}}$ is a plane-wave-like wave function and φ_{α} is an atomic function. The sum over α extends over all the atomic shells which are occupied. For example, in Na, the sum extends over the 1s, 2s, and 2p shells. The coefficients c_{α} are chosen such that the function $\phi_{\mathbf{k}}$, representing a 3s electron, is orthogonal to the core function φ_{α}. By requiring this orthogonality, we ensure that the 3s electron, when at the core, does not occupy the other atomic orbitals already occupied. Thus we avoid violating the Pauli exclusion principle.

The function $\phi_{\mathbf{k}}$ has the features we are seeking: Away from the core, the atomic functions φ_{α} are negligible, and thus $\phi_{\mathbf{k}} \sim \varphi_{\mathbf{k}}$, a plane wave. At the core, the atomic functions are appreciable, resulting in rapid oscillations, as shown in Fig. 6.

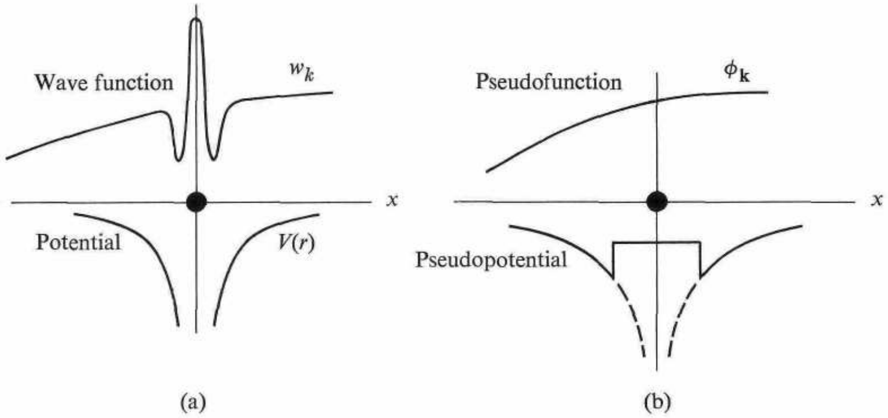

Fig. 6 The pseudopotential concept, (a) The actual potential and the corresponding wave function, as seen by the electron, (b) The corresponding pseudopotential and pseudofunction.

If one now substitutes $\phi_{\mathbf{k}}$ into the Schrodinger equation one can arrive at

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V^{pseudo}\right]\varphi_{\mathbf{k}}(\mathbf{r}) = E\varphi_{\mathbf{k}}(\mathbf{r}), \tag{19}$$

Because we took into account core levels in the representation of the wave function (18) the effective potential V^{pseudo} is weaker than the real potential V. This cancellation of the crystal potential by the atomic functions is usually appreciable, often leading to a very weak potential V. This is known as *the pseudopotential.* Since V^{pseudo} is so weak, the wave function is almost a plane wave and is called the *pseudofunction.*

The pseudopotential and pseudofunction are illustrated graphically in Fig. 6b. Note that the potential is quite weak, and, in particular, the singularity at the ion core is entirely removed. Correspondingly, the rapid "wiggles" in the wave function have been erased, so that there is a smooth plane-wave-like function.

The APW and pseudopotential methods, as well as other related systems, require much numerical work which can feasibly be carried out only by modern electronic computers.

Example: Electronic structure of copper

Copper has electronic configuration of ([Ar]3d^{10}4s^{1}). In the metallic state the closed-shell atomic levels of the argon configuration (1s^{2}2s^{2}2p^{6}3s^{2}3p^6) give rise to very tightly bound bands, lying well below the energies of any of the remaining electronic levels in the metal. The electrons in these low-lying levels can be considered as part of the (for most purposes) inert ion cores, and the remaining bands can be constructed by considering an fcc lattice of Cu^{11+} ions to which are added eleven electrons (3d^{10}4s^{1}) per primitive cell.

In the case of copper (and the other noble metals) at least six bands are required (and six turn out to be enough) to accommodate the eleven additional electrons. Their structure is shown in Figure 7. For almost all wave vectors **k** the six bands can be seen to separate into five lying in a relatively narrow range of energies from about 2 to 5 eV below E_F, and a sixth, with an energy anywhere from about 7 eV above to 9 eV below E_F.

It is conventional to refer to the set of five narrow bands as the d-bands, and the remaining set of levels as the s-band. However, these designations must be used cautiously, since at some values of **k** all six levels are close together, and the distinction between d-band and s-band levels is not meaningful. Also the “s-band” has a significant admixture from unoccupied 4p states. The nomenclature reflects the fact that at wave vectors where the levels do clearly group into sets of five and one. The five are derived from the five orbital atomic d-levels, in the sense of tight binding, and the remaining level accommodates what would be the 4s electron in the atom.

Note that the **k** dependence of the s-band levels, except where they approach the d-bands, bears a remarkable resemblance to the lowest free electron band for an fcc crystal (plotted in Figure 7b for comparison), especially if one allows for the expected modifications near the zone faces characteristic of a weak crystal potential. Note also that the Fermi level lies far enough above the d-band for the s-band to intersect E_F at points where the resemblance to the free electron band is still quite recognizable. Thus the calculated band structure indicates that for purposes of Fermi surface determination one might still hope for some success with a nearly free electron calculation.

However, one must always keep in mind that not too far below the Fermi energy lies a very complex set of d-bands, which can be expected to influence the metallic properties far more strongly than do any of the filled bands in the alkali metals.

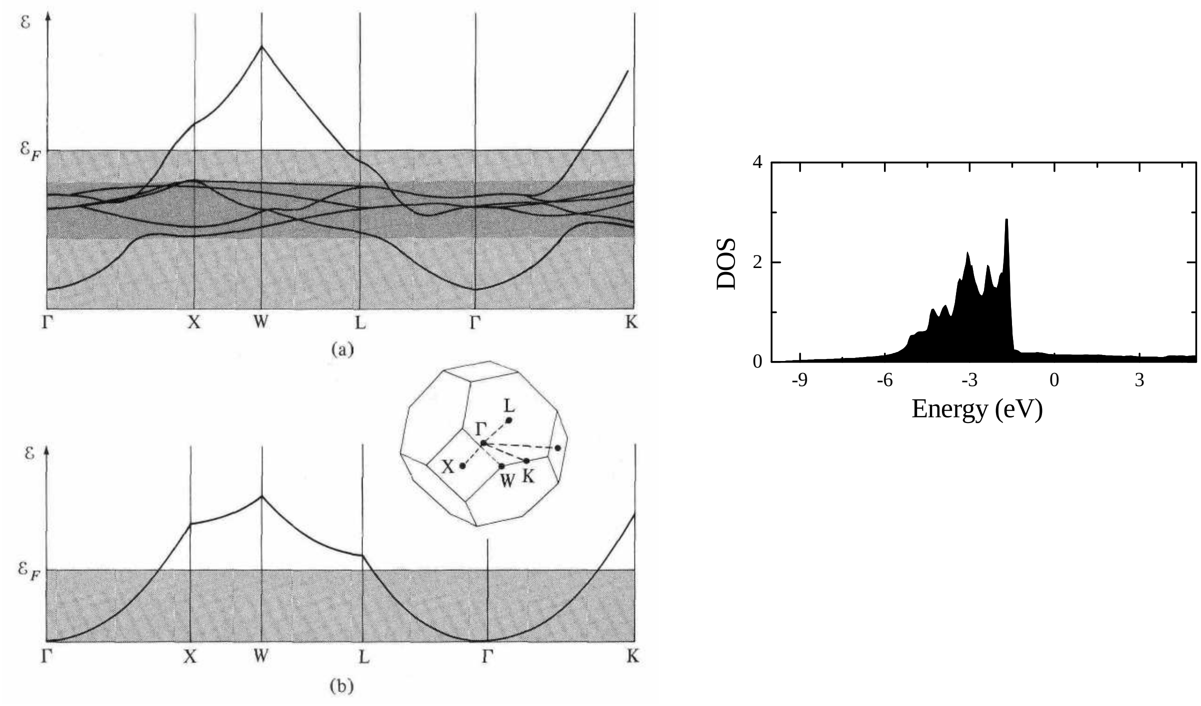

Fig.7 (a) Calculated energy bands in copper. The E vs. **k** curves are shown along several lines in the interior and on the surface of the first zone. (The point Γ is at the center of the zone.) The d-bands occupy the darkest region of the figure, whose width is about 3.5 eV. (b) The lowest-lying free electron energies along the same lines as in (a). (The energy scales in (a) and (b) are not the same.) (c) Density of states.

Fig. 1c shows the density of electronic states. The DOS is defined as the number of electronic states per unit energy per unit volume. The region with the high density of states corresponds to the d bands which are non dispersive, so that their energy does not change much with the k vector. On the other hand the energy of electrons within the s band changes significantly with the k vectors. The respective DOS of s electrons is low. Again it is important to note that strictly speaking there is a contribution of s character in the d band region, as well as the contribution of the d character within the s band region.