*Ab initio* tight binding

**REVIEW ARTICLE**

# *Ab initio* tight binding

A P Horsfield† and A M Bratkovsky‡

† Fujitsu European Centre for Information Technology, 2 Longwalk Road, Stockley Park,
Uxbridge, Middlesex UB11 1AB, UK
‡ Hewlett-Packard Laboratories, 3500 Deer Creek Road, Palo Alto, CA 94304-1392, USA

**Abstract.** Empirical tight binding has proven to be very popular in recent years on account of its computational efficiency and accuracy. However, it has limitations, notably the difficulties associated with fitting parameters and improving models when the desired accuracy cannot be achieved. In the light of this, a number of efforts have been made to derive tight-binding models from first principles. Here are described a number of formalisms based on density functional theory which span the range of approaches currently being used.

## 1. Introduction

Mathematical models are central to the interpretation of physical phenomena. The great advantage of computer models is that they can be made very sophisticated, and so describe rather accurately the phenomena we wish to understand. Indeed, the best quantum chemical calculations rival experiment in the accuracy they can achieve. Computer simulations can be carried out over the complete span of length scales from the cosmological to the sub-atomic. Here we focus on the atomic length scale.

For the overwhelming majority of problems of interest which are best described at the atomic level, achieving the most accurate account of some phenomenon requires that a balance must be struck between two competing requirements: a large enough number of non-equivalent atoms must be considered to remove effects due either to periodic boundaries or cluster surfaces; the model used to describe the interactions between atoms must be precise enough to include all the relevant features. The nature of the final compromise depends very sensitively on the problem being studied, and so there can be no universal method. In this review we will concentrate on those problems where a few hundred atoms are sufficient to describe the process being simulated, but where an explicit account of the electrons is needed to describe the interatomic interactions. It should be pointed out though that the rapid improvement in both algorithms and performance of hardware are shifting ever higher the maximum number of atoms that can be treated with the methods described below.

## 2. The role of quantum mechanical simulations

In the following, a basic understanding of total-energy quantum mechanical methods is assumed. An introduction to methods appropriate to solids can be found in Ashcroft and Mermin [1].

All interatomic interactions involve the motion of electrons. For some systems this can be accounted for in a very simple way leading to simple models. Notable examples based on

perturbation theory include: noble gases where the electrons are perturbed only slightly about a very stable atomic ground state leading to dispersion forces accurately described by a pair potential; nearly free-electron metals where most of the total energy can be described by a uniform electron gas in the potential of pseudo-ions with the small residual interactions being well described by a pair potential derived from second-order perturbation theory.

There are many other systems where simple perturbation theory is inadequate. A general class of such systems is where strong covalent bonds are made and broken leading to *large* redistributions of electron density. This redistribution leads to complicated non-local changes in interatomic interactions which are most easily described by treating the electrons explicitly. One example system is the carbon vacancy in titanium carbide [2]. The removal of the carbon atom results in charge being distributed preferentially into some bonds over others. The resultant atomic relaxation can only be understood using a many-centre analysis.

To describe electronic motion we must use quantum mechanics. Since exact solutions can only be found for a very limited range of problems approximations must be made. Almost all the effort in practice goes into constructing suitable numerical approximations. Which approximation is chosen depends strongly on the problem being solved. There are two basic decisions that always need to be made: the choice of theory (either the many-body Schrödinger equation or density functional theory); the choice of basis set in terms of which to expand the wavefunctions. *Ab initio* tight binding makes use of the Kohn and Sham [3] formulation of density functional theory. This is chosen on account of the accuracy that has been achieved consistently with rather simple approximations (notably the local density approximation). Possible choices of basis set are described below.

## 3. Empirical tight binding

In order to understand the interest in *ab initio* tight binding it is necessary to go back to its precursor, empirical tight binding [4–7]. This is the simplest *quantitative* quantum mechanical model. Its simplicity allows analytic results to be produced for a number of systems; thus it has been used extensively in the past to provide qualitative understanding of a wide range of electronic phenomena. Recently it has been rediscovered as a quantitative total-energy method, often being combined with molecular dynamics. The main reasons for this are: it is a quantum mechanical model and thus allows for a proper description of electronic motion; it is very simple and thus can be implemented very efficiently; it is a real-space method and thus can be used with linear scaling algorithms; it is a parametrized model and thus can give remarkably high accuracy for some systems.

But it also has major limitations: fitting the parameters is often a very lengthy business; constructing models for systems with more than one type of atom is usually much more difficult than creating models for a single atom type; when the model breaks down it is very hard to decide how to improve the model. A number of attempts have been made to produce more accurate models [8–17], and the success of Hartree–Fock-based empirical methods such as CNDO [18–22] suggests that for some systems improved empirical methods have value. However, the fundamental weaknesses remain.

## 4. *Ab initio* tight binding

The strengths of empirical tight binding are so clear that they provide a strong motivation to overcome its limitations. Since the limitations are related to fitting parameters and extending the model, the natural way to proceed is to derive a tight-binding model from first principles.

The decisions made during the derivation will define the limitations of the model. If greater accuracy or speed are required, different decisions can be made. In a first-principles description every term has a clear definition; thus evaluating terms for mixed systems should be no harder than for single-element systems.

Empirical tight binding is efficient for several reasons: the basis set is minimal, thus minimizing the time spent on diagonalizing the Hamiltonian matrix; the integrals are all given by formulae that are rapid to evaluate; the range of interaction between atoms is short (the orbitals are localized in space), thus allowing the construction of the Hamiltonian to be carried out in a time that scales linearly with the number of atoms. We would like to retain these properties in any first-principles formalism. There is no problem in principle with constructing a localized minimal basis set. However, in general the integrals will not be able to be represented by simple functions, but they can be evaluated once, and stored in tables that can be interpolated later on.

We now take a brief look at the fundamental theory underlying *ab initio* tight binding, and then look at a number of practical implementations. A comparison of the methods, using the isolated vacancy in silicon as an example, is given at the end of the review. However, it is important to note that the method which is most appropriate will depend strongly on the application. Thus cited applications of the methods should be referred to in order to determine which is the best for a given problem.

## 4.1. The Harris–Foulkes functional

A fundamental decision underpinning every *ab initio* tight-binding model is to make it non-self-consistent (though this constraint can be relaxed, as discussed below). The Harris–Foulkes functional [23–25] ($U_{HF}$) is very similar to the Kohn–Sham functional, except that it is defined entirely in terms of an *input* charge density ($n_{in}$) (whereas the Kohn–Sham functional is defined in terms of both an input and an output charge density):

$$U_{HF}[n_{in}] = \sum_i f_i \varepsilon_i - \frac{1}{2} \int d\vec{r} \, d\vec{r}' \frac{n_{in}(\vec{r}) n_{in}(\vec{r}')}{|\vec{r} - \vec{r}'|} + \frac{1}{2} \sum_{I \neq J} \frac{Z_I Z_J}{|\vec{R}_I - \vec{R}_J|}$$
$$+ E_{xc}[n_{in}] - \int d\vec{r} \, v_{xc}[n_{in}; \vec{r}] n_{in}(\vec{r}) \tag{1}$$

where $\varepsilon_i$ is an eigenvalue of the effective Hamiltonian $\hat{h} = \hat{T} + \sum_I v_I + v_{xc} + v_{Ha}$, $f_i$ is the corresponding single-particle state occupancy, $Z_I$ is the charge on ion $I$, $\vec{R}_I$ is the position of ion $I$, $E_{xc}$ is the exchange and correlation functional, $v_{xc}[n_{in}; \vec{r}] = \delta E_{xc}[n_{in}]/\delta n_{in}(\vec{r})$, $\hat{T}$ is the electron kinetic energy operator, $v_I$ is the interaction potential for an electron and ion $I$, and $v_{Ha}(\vec{r}) = \int d\vec{r}' \, n_{in}(\vec{r}')/|\vec{r} - \vec{r}'|$. The set of terms following the sum of eigenvalues is called the double-counting term. It corrects for the fact that part of the potential the electrons move in is generated by the electrons themselves. The eigenvalues are found by solving the equation

$$\hat{h}\psi_i = \varepsilon_i \psi_i \tag{2}$$

where $\psi_i$ is a single-particle wavefunction.

Given the original motivation for considering the electrons explicitly, namely that there are often cases where large charge transfers occur when bonds are made or broken, we need to justify the use of a non-self-consistent scheme in which we work with a fixed input charge density. The justification is that the error in the total energy is second order in the difference between the input charge density and the self-consistent charge density [26]. Provided the first-order terms dominate over all others, this is a good approximation. However, the electrostatic

terms are second order in the density, so if there is significant charge transfer leading to long-ranged internal fields, errors may occur.

This functional has been tested on a wide range of systems, and has been found to be surprisingly accurate. In particular, Polatoglou and Methfessel [27] looked at the bulk properties of Be, Al, V, Fe, Si, and NaCl. They found that the bulk modulus and lattice constant were well described in each case (even in ionic NaCl), though the energy was less well converged. Finnis [26] found that it is important that the input charge density be contracted relative to the free-atomic charge density. Using the contracted density he was able to obtain well converged results even for the surface and vacancy in aluminium.

One notable set of systems where it fails is transition metals [28]. The problem here is that the electronic configuration in the atom is quite different from that in the solid, even in the neighbourhood of the core. One cycle of self-consistency greatly improves the results.

Having discussed the general underlying theory, we now look at some specific implementations.

### 4.2. Atomic-like orbital-based tight binding

A small number of atomic orbitals are capable of giving good convergence. There are two factors responsible for this. The first is that, by construction, the orbitals have the correct form near the ion cores. The second is more general. The correct wavefunctions will minimize the energy of the system. Thus we can make a linear combination of atomic orbitals, and adjust the coefficients so as to minimize the energy. This gives us a best-fit approximation with the property that any errors in the final energy will vary as the square of the error in the wavefunctions. Thus the energy is rather insensitive to errors in the wavefunction.

There are three formalisms that will be discussed here that begin by expanding the wavefunctions in terms of atomic orbitals $\phi_{I\alpha}$, where $\alpha$ is an orbital index covering principal and angular quantum numbers. We begin by describing the formalism of Sankey and Niklewski [29], followed by the variant due to Horsfield [30], and finally the two-centre formalism of Frauenheim [31].

### 4.2.1. Formalism of Sankey and Niklewski.

Conceptually this method is straightforward. It is a non-self-consistent linear combination of atomic orbitals (LCAO) method in which the integrals are evaluated prior to a simulation, and then obtained from tables by interpolation during a simulation. There are, however, a number of important technical points that make the method viable, and hence interesting.

If we expand the wavefunctions as $\psi_i = \sum_{I\alpha} C_{I\alpha}^{(i)} \phi_{I\alpha}$, and define

$$h_{I\alpha,J\beta} = \int d\vec{r} \; \phi_{I\alpha} \hat{h} \phi_{J\beta}$$
$$O_{I\alpha,J\beta} = \int d\vec{r} \; \phi_{I\alpha} \phi_{J\beta}$$

$$(3)$$

then equation (2) can be transformed into the following matrix equation:

$$\sum_{J\beta} h_{I\alpha,J\beta} C_{J\beta}^{(i)} = \varepsilon_i \sum_{J\beta} O_{I\alpha,J\beta} C_{J\beta}^{(i)}. \tag{4}$$

The method then consists of a choice of input charge density, basis set and ways to evaluate and tabulate the integrals given in equation (3) (the Hamiltonian and overlap matrices) and equation (1) (the double-counting terms).

The orbitals are taken to be atomic-like, and the input charge density is taken as a sum of atomic-like charge densities ($N_{int}(\vec{r}) = \sum_I n_I(\vec{r})$), where $n_I$ is the atomic-like charge centred

on site $I$). We have already seen from the work of Finnis [26] that it is important to compress the charge density relative to the free atom. There are also two problems associated with taking the orbitals from the free atom. In the first place it makes calculations slow since the orbitals are long ranged and many neighbours must be considered when constructing the Hamiltonian and overlap matrices. The second problem is that, according to the virial theorem for particles that interact with potentials ($V$) that vary with distance as $1/r$, the electronic kinetic energy ($\hat{T}$) should increase as atoms bond to form a condensed system ($\langle\hat{T}\rangle = -\frac{1}{2}\langle V\rangle$). Both these problems can be overcome very elegantly by taking the orbitals (and also the charge density) from a confined atom. This puts the atom into a slightly excited state. In the method of Sankey and Niklewski the atom is confined by forcing the orbitals (but not their derivatives) to go to zero at some radius (5 Bohr radii for silicon). This is equivalent to confining the atoms in an infinitely deep spherical square well potential (see figure 1).
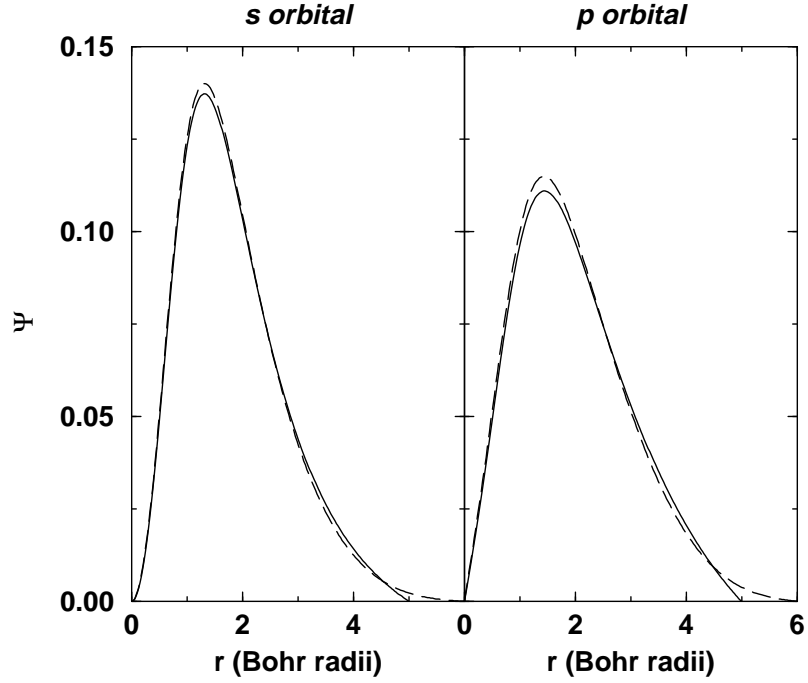


**Figure 1.** A comparison of two sets of orbitals for a minimal basis set for silicon. The full lines correspond to the orbitals used by Sankey and Niklewski, and the broken line to those used by Horsfield. The qualities of the basis sets are essentially the same. However, the basis set of Sankey and Niklewski has a discontinuity in its first derivative at the cut-off radius.

The overlap integrals ($O_{I\alpha,J\beta} = \int d\vec{r}\ \phi_{I\alpha}\phi_{J\beta}$) and kinetic energy integrals ($T_{I\alpha,J\beta} = \int d\vec{r}\ \phi_{I\alpha}\hat{T}\phi_{J\beta}$) clearly consist of one-centre ($I = J$) and two-centre ($I \neq J$) terms. The electrostatic part of the double counting ($-\frac{1}{2}\sum_{IJ}\int d\vec{r}\ d\vec{r}'\ n_I(\vec{r})n_J(\vec{r}')/|\vec{r} - \vec{r}'| + \frac{1}{2}\sum_{I\neq J} Z_I Z_J/|\vec{R}_I - \vec{R}_J|$) also consists of one- and two-centre terms.

The Hartree potential and the local part of the atomic pseudopotential can be combined to form a short-ranged neutral-atom potential

$$v^{(NA)}(\vec{r}) = \sum_I \left\{ v_I(\vec{r}) + \int d\vec{r}'\ \frac{n_I(\vec{r'})}{|\vec{r} - \vec{r}'|} \right\} = \sum_I v_I^{(NA)}(\vec{r}). \tag{5}$$

A typical matrix element involving this potential is

$$\int d\vec{r}\ \phi_{I\alpha}(\vec{r})v^{(NA)}(\vec{r})\phi_{J\beta}(\vec{r}) = \sum_K \int d\vec{r}\ \phi_{I\alpha}(\vec{r})v_K^{(NA)}(\vec{r})\phi_{J\beta}(\vec{r}). \tag{6}$$

If $I = J = K$ we have a one-centre integral. If $I \neq J$, but $K = I$ or $K = J$, then we have a two-centre integral. Otherwise we have a three-centre integral.

The one-centre integrals can be evaluated once and stored. The two-centre integrals can be tabulated as a function of separation on a one-dimensional grid, with rotations being taken into account by means of Slater–Koster [32] tables. The three-centre integrals are tabulated as a function of three variables (see figure 2): the bond length ($r$), the distance between the bond centre and the site on which the potential appears ($s$), and an angle ($\theta$). The tables are created for a specific geometry, and the integrals for other geometries are obtained by means of rotations. The integrals are three dimensional, and are performed in reciprocal space as this allows two out of the three integrals to be performed analytically, leaving a one-dimensional integral to be performed numerically.
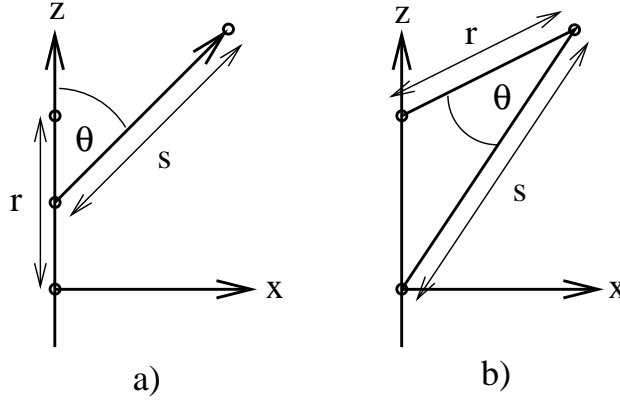


**Figure 2.** The geometry used to define variables in terms of which the three-centre tables are constructed. Panel (a) is for the method of Sankey and Niklewski, and panel (b) is for the method of Horsfield.

The electrostatic integrals are easy to tabulate because the total potential can be expressed as a linear combination of spherical single-site quantities. However, the integrals involving the exchange and correlation potential and energy are more difficult to handle because the functions involved have a strongly sub-linear dependence on density (roughly $n^{1/3}$). In the Sankey–Niklewski method this functional dependence is exploited by approximating the density in some region by an optimized constant value, for which the integrals are easy to evaluate. Consider the integral of the exchange and correlation potential in the local density approximation

$$\int d\vec{r}\ \phi_{I\alpha}(\vec{r})v_{xc}(n_{in}(\vec{r}))\phi_{J\beta}(\vec{r})$$

$$\approx \int d\vec{r}\ \phi_{I\alpha}(\vec{r})v_{xc}(\bar{n})\phi_{J\beta}(\vec{r}) + \int d\vec{r}\ \phi_{I\alpha}(\vec{r})v'_{xc}(\bar{n})(n_{in}(\vec{r}) - \bar{n})\phi_{J\beta}(\vec{r}) + \cdots$$

$$= v_{xc}(\bar{n})O_{I\alpha,J\beta} + v'_{xc}(\bar{n})(n_{I\alpha,J\beta} - \bar{n}O_{I\alpha,J\beta}) + \cdots \tag{7}$$

where $n_{I\alpha,J\beta} = \int d\vec{r}\ \phi_{I\alpha}(\vec{r})n_{in}(\vec{r})\phi_{J\beta}(\vec{r})$. The optimum constant value for the density ($\bar{n}$) is chosen so as to make the second term zero ($\bar{n} = n_{I\alpha,J\beta}/O_{I\alpha,J\beta}$). Care has to be taken when the overlap goes to zero. Dipole and quadrupole fluctuation corrections are included.

The method has been applied to such systems as amorphous silicon [33–39], silicon clusters [40], silicon surfaces [41, 42], various carbon structures [43–51], a number of multicomponent problems such as GeSe [52] and Si–C alloys [53], as well as many others [54–70].

*4.2.2. Formalism of Horsfield.* The formalism of Horsfield [30] is rather similar to that of Sankey and Niklewski. The important differences are the choice of basis set, and the way in which the exchange and correlation integrals are handled.

A minimal basis set was found to be inadequate for an accurate description of fluoro-carbons, but a double-numeric basis set gave good agreement with accurate density functional calculations. The orbitals were taken from the neutral atom and a positively charged ion [71]. As with the formalism of Sankey and Niklewski the atomic calculations were performed in a confining potential. The potential had the form $r^6$. The square-well potential was not used because it produces a discontinuity in the first derivative of the wavefunction at the cut-off radius (see figure 1). The integrals were all performed in real space using partition functions [71].

The perspective taken when evaluating the exchange and correlation integrals is rather different from that of Sankey and Niklewski. The key point is no longer the sub-linear dependence of the functionals on density, but rather the localized character of the confined atomic charge densities. For the exchange and correlation potential integrals, this localization allows us to write a many-centre expansion of the form

$$\int d\vec{r} \; \phi_{I\alpha}(\vec{r}) v_{xc}[n_{in}; \vec{r}] \phi_{J\beta}(\vec{r})$$

$$\approx \int d\vec{r} \; \phi_{I\alpha}(\vec{r}) v_{xc}[n_I + n_J; \vec{r}] \phi_{J\beta}(\vec{r})$$

$$+ \sum_{K(\neq I,J)} \int d\vec{r} \; \phi_{I\alpha}(\vec{r}) \{ v_{xc}[n_I + n_J + n_K; \vec{r}] - v_{xc}[n_I + n_J; \vec{r}] \} \phi_{J\beta}(\vec{r}) + \cdots.$$

$$(8)$$

As for the electrostatic terms we have one-, two-, and three-centre (and higher) terms. These are added to the electrostatic and kinetic energy terms to create a single set of tables (see figure 2 for the geometry used for the three-centre integrals). It was found to be necessary to carry out an additional numerical integral for the on-site exchange and correlation term since the few-centre approximation is not accurate enough in this case. However, the integral requires only a few points, and so is fast to evaluate. This method gives very good agreement with accurate self-consistent calculations for molecules [30].

*4.2.3. Formalism of Frauenheim.* It could be argued that this is not strictly a first-principles method in that it requires the fitting of a pair potential, but it does make use of tabulated integrals for the hopping integrals and overlap matrix, and has been used very successfully; thus it is considered here. The form used for the total energy is

$$U_{Frauenheim} = \sum_i f_i \varepsilon_i + \frac{1}{2} \sum_{I \neq J} \phi_{IJ}(|\vec{R}_I - \vec{R}_J|) \qquad (9)$$

where $\phi_{IJ}$ is a short-ranged repulsive pair potential. This form can be justified in terms of screening arguments [72].

As for the previous two formalisms the orbitals are obtained from atomic calculations with the atoms being confined by a localizing potential, the form used being $r^2$ [73]. Once the orbitals are chosen, the overlap and Hamiltonian matrices can be generated. This method

eliminates two terms from the Hamiltonian matrix that are present in the previous formalisms, namely the crystal-field terms and the three-centre integrals [31]. This can again be justified in terms of pseudopotential and screening arguments [74]. The potential used to evaluate the two-centre hopping integrals is the sum of the two spherical atomic potentials. The on-site terms in the Hamiltonian are taken from the free atom (not confined). Using this Hamiltonian and overlap matrix the band-structure energy can be evaluated. The pair potential is generated by subtracting the band energy from first-principles total-energy calculations for certain selected structures. The absence of three-centre terms makes this method faster and more economical with its use of memory as compared with the previous two methods.

This formalism has been applied widely. Materials that have been studied using this formalism include carbon [31], silicon [75], boron nitride [76], germanium [77], gallium arsenide [78], $Si_nN_m$ clusters [79], and gallium nitride [80].

### 4.3. Muffin-tin orbital-based tight binding

As stated above, we wish to make our basis set localized in real space and as small as possible. Making use of orbitals obtained from calculations of an atom in a confining potential is one way of approaching this problem. A less *ad hoc* approach is to construct the basis set from solutions to an exactly solvable problem that is as close as possible to the one we are interested in. The muffin-tin potential [81] provides a popular and successful solvable problem for close-packed systems. The potential for a solid is approximated by a series of non-overlapping atomic-like spherical potentials, and a constant potential between the spheres (a two-dimensional representation of the regions looks rather like a muffin tin). Schrödinger's equation can be solved exactly in both regions. These solutions are matched at the boundaries of the spheres to produce muffin-tin orbitals (MTOs). One can further reduce the effect of the interstitial regions by working with space-filling, overlapping atomic spheres instead of MT spheres (so-called atomic spheres approximation, ASA). Unfortunately these MTOs are very long ranged. However, a unitary transformation can be applied to these long-ranged orbitals to render them short ranged, and hence suitable for tight binding [82]. The formalism is somewhat intricate, so the relevant theory has been reproduced in the appendix.

### 4.3.1. Nearly orthogonal tight-binding LMTOs.

In the linear MTO method, MTOs are used to construct a basis set which is (i) energy independent, (ii) exact to linear order in energy, and (iii) rapidly convergent. The latter means that for the valence electrons it is generally sufficient to retain one orbital per site per orbital quantum number $L$ ($=lm$), where $l$ and $m$ are the orbital and azimuthal quantum numbers respectively. Typically nine standard LMTOs ($\chi^0$) per site (corresponding to an spd basis set) produce sufficient accuracy for most transition metals [83]. This is a minimal basis set, which is what we would like for tight binding.

The LMTOs represent a variational basis constructed from the solutions of the Schrödinger equation within the muffin-tin spheres $\phi(\epsilon_\nu, r)$, where $\epsilon_\nu$ is selected within the region of energies occupied by the valence electrons. If we consider an arbitrary energy $\epsilon$, then the LMTOs have an error of order $(\epsilon - \epsilon_\nu)^2$ within the muffin-tin (MT) spheres and of order $(\epsilon - \epsilon_\nu)^1$ in the interstitial region (not belonging to any of the muffin-tin spheres). To improve the accuracy of the method one usually selects *overlapping* MT spheres whose volume is equal to the volume of the respective Wigner–Seitz spheres and neglects completely the contribution of the interstitial regions (so-called atomic sphere approximation, ASA, or LMTO-ASA).

The MTOs are long ranged with tails that decay with distance from the origin $r$ as $1/r^{l+1}$. (Thus an s-MTO decays as $1/r$, and p- and d-MTOs decay as $1/r^2$ and $1/r^3$ respectively.) These are not suitable for real-space calculations. However, these tails are formally analogous

to a Coulomb field produced by a superposition of electric multipoles placed at atomic sites, and thus they may be *screened* (made short ranged) by appropriate unitary transformations of the initial basis [83] to form a tight-binding representation. We shall denoted the screened LMTO by $\chi^\alpha$.

By screening the linearized MTOs, one can produce the *most localized* basis set $\chi^\alpha$, most advantageous for real-space calculations. These orbitals are non-orthogonal, but can be transformed into *nearly orthogonal* (though somewhat longer-ranged) orbitals $\chi^\gamma$. A corresponding nearly orthogonal tight-binding Hamiltonian $H^\gamma$ can then be generated and used in much the same way as an orthogonal empirical tight-binding Hamiltonian. A summary of the underlying theory is given in the appendix. Here the central results are presented.

A brute-force method to transform the most localized orbitals $\chi^\alpha$ into orthogonal ones is to apply the Löwdin transformation

$$\mathcal{H} = \frac{1}{\sqrt{O^\alpha}} H^\alpha \frac{1}{\sqrt{O^\alpha}} \equiv \langle (O^\alpha)^{-1/2} \chi^\alpha | -\nabla^2 + v | \chi^\alpha (O^\alpha)^{-1/2} \rangle \tag{10}$$

where $\mathcal{H}^\alpha \equiv \langle \chi^\alpha | -\nabla^2 + v | \chi^\alpha \rangle$ is the standard LMTO Hamiltonian, $v$ is the electron potential, and $O^\alpha \equiv \langle \chi^\alpha | \chi^\alpha \rangle$ is the matrix of overlap integrals between the screened LMTOs $\chi^\alpha$. We would like to avoid doing this transformation by careful choice of our LMTO representation.

Consider the representation where $\alpha = \gamma_{II}$, i.e. where we use *site-dependent* screening. In this representation the parameter $\ddot{P}^\alpha(\varepsilon_\nu) = 0$, as follows from equations (A.47), (A.52), and (A.53). Thus we have

$$\begin{aligned} O^\gamma &= 1 + h^\gamma p h^\gamma + {}^i \langle \kappa^\gamma | \kappa^\gamma \rangle^i \approx 1 \\ H^\gamma &\approx \varepsilon_\gamma + h^\gamma + h^\gamma \varepsilon_\nu p h^\gamma + {}^i \langle \kappa^\gamma | -\nabla^2 + v | \kappa^\gamma \rangle^i \approx \varepsilon_\gamma + h^\gamma. \end{aligned} \tag{11}$$

Therefore, we can easily construct a *nearly* orthogonal representation which is sufficiently accurate for applications using the recursion method. To get the expression for $h^\gamma$ we have to use the fact that the first term in the overlap (A.53) is the major one, so that $(O^\alpha)^{-1/2} \approx (1 + \sigma^\alpha h^\alpha)^{-1}$ and we can introduce (see equation (10)) the nearly orthogonalized LMTOs

$$|\chi^\gamma\rangle = |\chi^\alpha\rangle (1 + \sigma^\alpha h^\alpha)^{-1}$$

and using this to calculate the Hamiltonian matrix, we get

$$H^\gamma \equiv \langle \chi^\gamma | -\nabla^2 + v | \chi^\gamma \rangle = (1 + h^\alpha \sigma^\alpha)^{-1} H^\alpha (1 + \sigma^\alpha h^\alpha)^{-1}.$$

Substituting here the expression for $H^\alpha$ from equation (A.52), we get equation (11) with

$$h^\gamma \equiv (1 + h^\alpha \sigma^\alpha)^{-1} h^\alpha = h^\alpha (1 + \sigma^\alpha h^\alpha)^{-1} = h^\alpha - h^\alpha \sigma^\alpha h^\alpha + \cdots \tag{12}$$

so we can use the most-localized $h^\alpha$ for constructing the nearly orthogonal $H^\gamma$. We should keep in mind that, in principle, the $\gamma$-representation is longer ranged than the *most-localized* $\alpha$-representation.

The actual computational procedure starts from the transformation $S^0 \rightarrow S^\alpha$; then one calculates the potential parameters for a trial potential (or density) and evaluates $h^\gamma$ from equation (12). One then proceeds to calculate $H^\gamma$ and spectral functions (the electron density of states, etc) needed to reach a self-consistent solution to the electronic problem. This is especially useful in combination with the recursion method [84, 85] for large disordered solids [86], where the $k$-space method breaks down.

We now briefly consider a few examples of the applications of this method. It has proven very useful in studies of amorphous magnetism, such as that in $Fe_{1-x}B_x$ metallic glasses [87] and Al–Mn [88, 89]. In the latter case it was possible to make a prediction of an unusual transition when the paramagnetic crystalline $Al_6Mn$ becomes magnetic on being *melted*. The

real-space recursion TB-LMTO method was extended to treat arbitrary *non-collinear* magnetic structures (i.e. with arbitrary directions of the exchange potential $\frac{1}{2}\vec{\Delta}_I\widehat{\vec{\sigma}}$, where $\widehat{\vec{\sigma}}$ are the Pauli matrices) which made it possible to establish the existence of a random magnetic order in Al–Mn liquids in [89], and investigate complex magnetism and magnetization direction switching in ultrathin Co/Cu films [90, 91]. In the latter case the convergence was achieved to better than 0.0004 electrons au$^{-3}$ for electron densities and about 0.003° for spin directions. The number of iterations necessary for such a self-consistency was very large, about 1000. Additional examples are mentioned in e.g. [106, 108].

*4.3.2. Effective-medium-theory-based tight binding.*  In this model the total energy is given by [92]

$$U_{EMTB} = \sum_I e_c(s_I) + \left[ E_{as} - \sum_I e_{as}(s_I) \right] + \left[ E_{1el} - \sum_I e_{1el}(s_I) \right]. \qquad (13)$$

A central quantity in this expression is $s_I$, which is the neutral-atom radius. This is the radius of the smallest sphere centred on site $I$ which is charge neutral. The function $e_c(s_I)$ gives the energy per atom in a reference system (the diamond structure for the case of silicon) with the same neutral-atom radius. This is the large term in the expression. The following two pairs of terms are corrections to this term which take into account the details of the environment. The pair of terms (the atomic sphere terms) $\left[ E_{as} - \sum_I e_{as}(s_I) \right]$ is the difference in a combined electrostatic and exchange and correlation term between the real system ($E_{as}$) and the reference system. The term for the real system is expressed as a density-dependent pair potential. The final pair of terms is the difference in the one-electron energies between the real system and the reference system. A two-centre TB-LMTO Hamiltonian is used to evaluate the one-electron energies for the real system. This method has been applied to a number of problems involving silicon [93–95].

## 5. Introducing self-consistency

One of the main limitations of the formalisms described above is the absence of charge self-consistency. This is important for many systems, and so a number of attempts have been made to include self-consistency into the tight-binding models. The schemes range from simple monopolar corrections to full self-consistency with no shape approximations for the charge density.

Elstner *et al* [96] introduced self-consistency into the formalism of Frauenheim. We begin with this approach because it can be applied to any tight-binding formalism (empirical as well as *ab initio*). Self-consistency is introduced by means of the addition of the following term to the tight-binding expression for the total energy:

$$\Delta E = \frac{1}{2} \sum_{I,J} q_I q_J \gamma_{I,J} \qquad (14)$$

where $\gamma_{I,J} = \int d\vec{r}\, d\vec{r}'\, F_I(|\vec{r} - \vec{R}_I|) F_J(|\vec{r}' - \vec{R}_J|)/|\vec{r} - \vec{r}'|$. The function $F_I(r)$ is a spherical charge density, normalized to 1, and is taken to have an exponential form. Gaussians could also be used, as they make the algebra simpler [97]. The charges $q_I$ are given by Mulliken population analysis ($q_I = \sum_{\alpha,J\beta,i} C_{I\alpha}^{(i)} f_i C_{J\beta}^{(i)} O_{J\beta,I\alpha} - Z_I$). Minimizing the total energy with respect to the expansion coefficients $C_{I\alpha}^{(i)}$ gives the self-consistent Hamiltonian

$$h_{I\alpha,J\beta} = h_{I\alpha,J\beta}^{(0)} + \frac{1}{2}(w_I + w_J) O_{I\alpha,J\beta} \qquad (15)$$

where $h^{(0)}_{I\alpha,J\beta}$ is the non-self-consistent Hamiltonian, and where the energy shift $w_I = \sum_J q_J \gamma_{I,J}$. The force on an atom is given by

$$\vec{F}_K = \vec{F}_K^{(0)} - \sum_{I\alpha,J\beta,i} C_{I\alpha}^{(i)} f_i C_{J\beta}^{(i)} \frac{(w_I + w_J)}{2} \frac{\partial O_{I\alpha,J\beta}}{\partial \vec{R}_K} - \frac{1}{2} \sum_{I,J} q_I q_J \frac{\partial \gamma_{I,J}}{\partial \vec{R}_K} \qquad (16)$$

where $\vec{F}_K^{(0)}$ is the force evaluated using the expression for non-self-consistent tight binding.

Formally this expression ignores any contribution from exchange and correlation. However, the main contribution to self-consistency is from the on-site term, and the decay rates of the functions $F_I$ are adjusted to give an on-site value ($\gamma_{I,I}$) that agrees with either experiment or first-principles calculations, and thus includes correlation.

There have been two attempts to include self-consistency into the formalism of Sankey and Niklewski within the spherical charge approximation [55,61]. Here we will only consider the later method [61]. Consider silicon. It has s and p valence electrons. Self-consistency is achieved by varying the number of s and p electrons contributing to the charge density on each site. This allows both for promotion of electrons from s to p, and for the net accumulation of charge. The number of s and p electrons is determined from the output wavefunctions using the projection onto the Löwdin orbitals [98]. All the integrations are carried out exactly as before, with one exception. When there is a net charge on a site there will be a long-ranged electrostatic field. Integrals involving this field are treated within a dipole approximation.

The formalism of Horsfield also includes approximate self-consistency. It assumes spherical input charges, but allows them to vary as $n_I(\vec{r}) = n_I^{(0)}(\vec{r}) + q_I \Delta_I(\vec{r})$, where $n_I^{(0)}$ is the charge density from the neutral atom. The perturbing charge density $\Delta_I$ is spherical, normalized to one, and constructed from partially occupied orbitals. The Hamiltonian matrix elements are assumed to vary linearly with $q_I$, and the double-counting term to vary quadratically. For the electrostatic terms this is the correct behaviour, but for the exchange and correlation terms this is approximate. The long-ranged electrostatic fields are handled using a low-order Gaussian expansion for the orbitals which allows analytic expressions to be used. The values of $q_I$ are found by maximizing the total energy. This follows from the fact that the Harris–Foulkes functional is maximized by the correct input charge density [99] (unlike the Kohn–Sham functional which is minimized).

The formalism of Lin and Harris [99] is self-consistent from the beginning. The main difference as compared with the tight-binding formalisms described above is that analytic functions are used for the orbitals and input charge density, allowing all the integrals except those involving exchange and correlation to be expressed in closed form. A quadratic approximation is made for the exchange and correlation energy. The charge on each site is found by maximizing the total energy.

All the above self-consistent schemes make the approximation that the input charge density is a sum of atom-centred spherical charge densities. Ordejon *et al* [100] relaxed this approximation and represent the difference between the sum of atomic densities and the self-consistent density on a uniform mesh. They thus turned the formalism of Sankey and Niklewski into a fully self-consistent LCAO method.

## 6. Comparison of methods

In order to give some idea of the relative efficiencies and accuracies of the methods described above, calculations of the formation energy of the relaxed isolated vacancy in silicon were performed. All the calculations use a 64-atom cell with one atom removed. Only the $\Gamma$ point is used for sampling the Brillouin zone. The basis set is always minimal (s and p).

The calculations were performed using the program Plato (package for linear combination of atomic-type orbitals).

The times we consider are for the evaluation of one total energy and set of atomic forces, and we will measure them relative to the time for non-self-consistent orthogonal empirical tight binding. The methods considered are: empirical orthogonal tight binding using the parameters of Bowler *et al* [101]; the non-orthogonal tight binding of Frauenheim; the *ab initio* tight binding of Horsfield; and full linear combination of atomic orbitals (LCAO) with many of the integrals evaluated during the calculation. With the exception of the full LCAO method, self-consistency is imposed using the monopole method of Elstner *et al* [96]. From table 1 we can draw several general conclusions.

**Table 1.** Results of calculations of the formation energy of the relaxed vacancy of silicon. The times are measured relative to orthogonal tight binding. Note the following abbreviations: TB = tight binding, SC = self-consistent, AI = *ab initio*, LCAO = linear combination of atomic orbitals, DNP = double numeric with polarization. The times in curly brackets for the *ab initio* tight binding correspond to the use of the Chebyshev expansion for the three-centre integrals, whereas the other numbers correspond to the use of linear interpolation.

| Method | Formation energy (eV) | Time |
|---|---|---|
| Orthogonal TB | 3.2 | 1 |
| SC orthogonal TB | 3.2 | 13 |
| Non-orthogonal TB | 3.1 | 2 |
| SC non-orthogonal TB | 3.4 | 19 |
| AI TB | 4.3 | 72 {180} |
| SC AI TB | 4.4 | 87 {210} |
| AI LCAO | 3.8 | 44 |
| SC AI LCAO | 4.1 | 110 |
| DNP AI LCAO | 3.9 | 880 |
| DNP SC AI LCAO | 3.2 | 2100 |

First, we consider the cost of simulations in terms of computer time: introducing three-centre terms makes the simulations much slower; compared with the non-self-consistent calculations, self-consistency is cheap for *ab initio* tight binding, but expensive for everything else; full LCAO can be cheaper than *ab initio* tight binding (this is because the three-centre integrals do not have to be evaluated one at a time in the full LCAO calculations).

Second, we consider the accuracies. For the four results that use the same basis set (*ab initio* tight binding and the full LCAO method), we see that no one of the approximations is obviously better than the others. The four results based on two-centre tight binding show remarkable agreement with each other, but are all about 1 eV smaller than the corresponding *ab initio* results. In table 1 are also given results obtained with a well converged basis set (two sets of s and p orbitals, and one set of d orbitals) using the full LCAO method. The self-consistent result (3.2 eV) is in remarkably good agreement with the two-centre tight-binding results. This follows from the fact that the two-centre models are fitted to well converged self-consistent results.

The accuracies cited above have been relative to other atomic-type orbital methods. The result from well converged plane-wave calculations is about 3.6 eV [102], which is rather larger than the most accurate result obtained from the self-consistent LCAO calculation with the large basis set. The origin of the discrepancy is the *k*-point sampling. Using only the Γ point is insufficient. If the non-self-consistent orthogonal tight-binding simulation is repeated

with a $3 \times 3 \times 3$ mesh of $k$-points, the vacancy formation energy is found to increase from 3.2 eV to 3.8 eV.

For the case of the vacancy it is clear that the two-centre tight-binding models are faster and more accurate than the *ab initio* minimal-basis-set methods. This is because the fitting procedures used to generate the models were appropriate for this problem. However, the two-centre models are not as transferable as the *ab initio* models. This can be clearly seen from the phase diagram of silicon. The non-orthogonal two-centre model of Frauenheim gives a description of the close-packed phases of silicon that is rather less good than that of the more open structures [75], whereas the *ab initio* model of Sankey and Niklewski [29] gives a rather better description. It should also be remembered that the motivation for developing the *ab initio* methods was to provide a scheme that can be improved systematically and which can be applied easily to multicomponent systems. To see both factors coming into play in the case of hydrocarbons and fluorocarbons, see Horsfield [30].

## Appendix. Derivation of the tight-binding LMTO method

In the muffin-tin approximation we have spherically symmetric potentials $v_I(r_I)$ within atomic spheres (AS) of radius $s_I$ centred on sites $I$, and a constant potential ($v_0 = 0$) in the interstitial region. Thus the total potential ($v(\vec{r})$) is given by

$$v(\vec{r}) = \sum_I v_I(r_I) \tag{A.1}$$

where $r_I = |\vec{r} - \vec{R}_I|$, and is the distance from the centre of the site at $\vec{R}_I$. Orbitals with eigenvalue $\varepsilon$ ($\varphi_l(\varepsilon, r)$) and energy derivatives $\dot{\varphi}_L(\varepsilon, r) = \mathrm{d}\varphi_l(\varepsilon, r)/\mathrm{d}\varepsilon$ are found from

$$[\nabla^2 + \varepsilon - v_I(r_I)]\varphi_l(\varepsilon, r) = 0 \tag{A.2}$$

with the energy measured in Ryd, and the distances in Bohr radii.

An orbital with arbitrary energy $\varepsilon$ can be approximated inside the muffin-tin sphere by

$$\varphi_l(\varepsilon, r) = \varphi_l(\varepsilon_v, r) + (\varepsilon - \varepsilon_v)\dot{\varphi}(\varepsilon_v, r) + \mathrm{O}\left[(\varepsilon - \varepsilon_v)^2\right]. \tag{A.3}$$

These are linearized muffin-tin orbitals. This approximation is accurate for energies spanning about 1 Ryd around $\varepsilon_v$.

We shall denote by $|\varphi(\varepsilon, r)\rangle$ the solutions terminated (that is, set to zero) outside their respective atomic spheres. The orbitals are normalized to unity within the AS:

$$\langle \varphi | \varphi \rangle \equiv \int_0^{s_I} \mathrm{d}r \, r^2 \varphi^2(\varepsilon, r) = 1$$

$$\langle \varphi | \dot{\varphi} \rangle = 0 \tag{A.4}$$

where the second orthogonality condition follows directly from normalization.

In the interstitial region (where $v = v_0$) the Schrödinger equation reduces to the Helmholtz *wave equation*

$$(\nabla^2 + \kappa^2)\varphi(\varepsilon, \vec{r}) = 0 \tag{A.5}$$

where $\kappa^2 = \varepsilon - v_0$.

A complete orbital is constructed from a part inside the AS (the head) and a part that lies outside (the tail). For the tail region one selects a solution of (A.5) in terms of the Bessel functions $H_l(\kappa^2)$ at given energies $\varepsilon_v$, which are generally different from the values used for the head. The simplest, but very effective, choice is $\kappa_v^2 = 0$. In this case equation (A.5) reduces to the Laplace equation, with the LMTO tail in the interstitial region being proportional to

$$K_{Ilm}^0(\vec{r}_I) \equiv \left(\frac{w}{r_I}\right)^{l+1} Y_{lm}(\vec{r}_I) \tag{A.6}$$

where $w$ is a scaling length, such as the average Wigner–Seitz radius. The envelope function $K^0$ has the form of an electrostatic potential produced by a $2^l$ pole at $I$. It is regular everywhere except at $R_I$, and can be expanded about any site $I' \neq I$ [103]:

$$K^0_{IL}(\vec{r}_I) = -\sum_{L'} |J^0_{I'L'}(\vec{r}_{I'})\rangle S^0_{I'L'IL} \tag{A.7}$$

in terms of the regular solutions of the Laplace equation

$$J^0_{IL}(\vec{r}_I) \equiv \frac{(r_I/w)^l}{2(2l+1)} Y_L(\vec{r}_I). \tag{A.8}$$

The ket-vector notation is used to denote the orbital terminated beyond the atomic sphere it is centred on. We shall make use of the convention that the orbitals, which depend explicitly on the direction $\vec{r}_I$, contain the spherical function $Y_L(\vec{r}_I)$. Otherwise, only the radial part depending on the absolute value of $r_I$ is used. The expansion coefficients $S^0$ are the so-called canonical structure constants which are independent of energy, lattice constant, and sphere radii. They have a very simple form in terms of the Slater–Koster notation [32]:

$$\begin{aligned}
S^0_{ss\sigma} &= -2w/d \qquad S^0_{sp\sigma} = 2\sqrt{3}(w/d)^2 \\
S^0_{pp}(\sigma, \pi) &= 6(w/d)^3(2, -1) \\
S^0_{sd\sigma} &= -2\sqrt{5}(w/d)^3 \\
S^0_{pd}(\sigma, \pi) &= 6\sqrt{5}(w/d)^4(-\sqrt{3}, 1) \\
S^0_{dd}(\sigma, \pi, \delta) &= 10(w/d)^5(-6, 4, -1).
\end{aligned} \tag{A.9}$$

The constants for arbitrary $\vec{R}_{I'} - \vec{R}_I$ required in the expansion (A.7) can be trivially found from Slater–Koster tables [32]. The same expansion (A.7) holds true if one uses the envelope solutions $K^0$ for $\kappa_\nu^2 \neq 0$, with a more complicated functional form for the $S(\kappa^2)$ matrices [103].

The construction of the LMTOs proceeds as follows. Consider one atomic sphere centred at site $I$. The solution of Schrödinger's equation inside the sphere is approximated by a linear combination of $\varphi_\nu$ and $\dot{\varphi}_\nu$ by matching the function and its first radial derivative to the envelope function $K^0$ at the atomic sphere surface:

$$K^0_{Il}(r_I) \equiv (w/r_I)^{l+1} \to A^K_{Il}(\varepsilon)\varphi_l(\varepsilon, r_I) + B^K_{Il}(\varepsilon)\dot{\varphi}_l(\varepsilon, r_I) \tag{A.10}$$

where $A^K$ and $B^K$ are the expansion coefficients found from the matching conditions. Note that since $\langle \varphi(\varepsilon)|\dot{\varphi}(\varepsilon)\rangle \equiv 0$ for *any* $\varepsilon$ we can perform this matching procedure for *arbitrary* energy and then generate a set of energy-*dependent* basis functions. As we will see shortly, this reduces the Schrödinger equation to a non-linear eigenvalue problem, which is difficult to solve. Instead, we shall use our freedom in constructing these orbitals to go over directly to an energy-*independent* basis set, which can be used effectively in a variational solution of the Schrödinger equation, since it results in the linear eigenvalue problem.

The tail of the envelope function (A.7) is also matched in the same way inside neighbouring spheres to the orbitals and their first energy derivatives:

$$J^0_l(r_{I'}) \equiv (r_I/w)^l/2(2l+1) \to \tilde{J}^0_{I'l}(\varepsilon, r_{I'}) \equiv A^J_{Il}(\varepsilon)\varphi_l(\varepsilon, r_I) + B^J_{Il}(\varepsilon)\dot{\varphi}_l(\varepsilon, r_I) \tag{A.11}$$

where the expansion coefficients $A^J_{Il}$ and $B^J_{Il}$ are found from this matching condition, as $A^K_{Il}$ and $B^K_{Il}$ have been before them. By using these substitutions we arrive at some atom-centred, energy-dependent orbitals which are called the MTOs $\chi$, and we write them in the following form for any point in the crystal:

$$\chi^N_{IL}(\varepsilon, \vec{r}) = |\varphi_L(\varepsilon, \vec{r}_I)\rangle A^K_{Il}(\varepsilon) + |\dot{\varphi}_L(\varepsilon, \vec{r}_I)\rangle B^K_{Il}(\varepsilon) - \sum_{I'L'} |\tilde{J}^0_{I'L'}(\vec{r}_{I'})\rangle S^0_{I'L'IL} + |K\rangle^i \tag{A.12}$$

where all $|\ \rangle$s are non-zero *only* in their respective atomic spheres, $|\ \rangle^i$ is non-zero only in the interstitial region, and $N_v$ is the normalization coefficient. It is more convenient to write MTOs in a slightly different, though equivalent, fashion. Namely, instead of $\varphi$ and $\dot{\varphi}$ one can use $\varphi$ and $\tilde{J}^0$ to arrive at a more frequently used representation of the energy-dependent MTOs:

$$\chi_{IL}^N(\varepsilon, \vec{r}) = |\varphi_L(\varepsilon, \vec{r}_I)\rangle N_{Il}^0(\varepsilon) + |\tilde{J}_L^0(\varepsilon, \vec{r}_I)\rangle P_{Il}^0(\varepsilon) - \sum_{I'L'} |\tilde{J}_{I'L'}^0(\varepsilon, \vec{r}_{I'})\rangle S_{I'L'IL}^0 + |K\rangle^i \quad (A.13)$$

where $P_{Il}^0(\varepsilon) = B_{Il}^K(\varepsilon)/B_{Il}^J(\varepsilon)$, $N_{Il}^0(\varepsilon) = A_{Il}^K(\varepsilon) - P_{Il}^0(\varepsilon)A_{Il}^K(\varepsilon)$. Omitting the site and orbital indices and summation over repeating indices, one can rewrite MTOs in the compact form

$$\chi^N(\varepsilon) = |\varphi(\varepsilon)\rangle N(\varepsilon) + |\tilde{J}^0(\varepsilon)\rangle[P^0(\varepsilon) - S^0] + |K\rangle^i \quad (A.14)$$

where $N(\varepsilon)$ and $P^0(\varepsilon)$ are the so-called *potential parameters* which are diagonal with respect to combined site and orbital index $IL$. Since $\chi_{IL}(\varepsilon, \vec{r}) = \chi_{IL}(\varepsilon, \vec{r} - \vec{R}_I)$, we can use the standard construction for the Bloch wave in a perfect crystal. Since the normalization factor is a constant we can work with

$$\chi(\varepsilon) \equiv \chi_{IL}^N(\varepsilon)/N_{IL}(\varepsilon_v). \quad (A.15)$$

We can write a general electron wavefunction as

$$\psi_i(\varepsilon) = \sum_{IL} C_{IL}^{(i)} \chi_{IL}(\varepsilon). \quad (A.16)$$

One is free to select the coefficients $C_{IL}^{(i)}$ such that the second term in equation (A.14) vanishes:

$$\sum_{I'L'} C_{I'L'}^{(i)}(\varepsilon)[P_{I'l'}^0(\varepsilon)\delta_{I'L'IL} - S_{I'L'IL}^0] = 0. \quad (A.17)$$

Then the remaining part, which is a linear combination of $|\varphi(\varepsilon)\rangle$ and $|K\rangle^i$, satisfies the Schrödinger equation for the muffin-tin potential *exactly*. Thus, we have recovered the famous KKR tail-cancellation condition. This equation, which defines the electron energy bands, is very inconvenient indeed, since it is a non-linear equation for the eigenvalues $\varepsilon$. Instead of solving this equation, we would like to construct from MTOs the most compact energy-*independent* basis to use for the *variational* solution of the Schrödinger equation. Moreover, we can use our flexibility in choosing energy-independent linear MTOs $\chi$ (LMTOs) to make them accurate to first order in $(\varepsilon - \varepsilon_v)$ inclusive, so that the errors will be proportional to $(\varepsilon - \varepsilon_v)^2$ *inside* the atomic spheres. As we mentioned earlier, our approximations mean that in the *interstitial* region the error of the LMTO is proportional to $(\varepsilon - \varepsilon_v)^1$, but this contribution is neglected in the atomic sphere approximation, where one selects *space-filling* overlapping MT spheres. We need to analyse MTOs in detail to see all this.

The matching conditions at the atomic sphere boundaries give us a two-by-two linear system of equations, and it is convenient to write their solutions in the following form:

$$P^0(\varepsilon) = \frac{W[\varphi(\varepsilon), K^0]}{W[\varphi(\varepsilon), J^0]} = 2(2l + 1)\left(\frac{w}{s}\right)^{2l+1} \frac{D(\varepsilon) + l + 1}{D(\varepsilon) - 1} \quad (A.18)$$

$$N^0(\varepsilon) = \frac{W[K^0, J^0]}{W[\varphi(\varepsilon), J^0]} = \frac{w/2}{W[\varphi(\varepsilon), J^0]} = \left[\frac{w}{2}\dot{P}^0(\varepsilon)\right]^{1/2} \quad (A.19)$$

where $D(\varepsilon) \equiv s\varphi'(\varepsilon, s)/\varphi(\varepsilon, s)$ is the logarithmic derivative of the wave function at the atomic sphere surface, which is simply related to the usual scattering phase shift $\delta_l$ [104], and $W[f_1, f_2] \equiv s^2[f_1'(s)f_2(s) - f_1(s)f_2'(s)]$ is the Wronskian. In both cases the prime stands for differentiation with respect to the radial variable. To derive (A.19) for $N$ via $\dot{P}$ one should differentiate the Wronskian relation (A.18) and use

$$W[\dot{\varphi}, \varphi] = s^2(\dot{\varphi}\varphi' - \dot{\varphi}'\varphi)_{r=s} = \int_0^s dr\ r^2\varphi^2 = 1. \quad (A.20)$$

The latter relation follows from the Schrödinger equation (A.2) and its first derivative with respect to energy

$$(-\nabla^2 + v - \varepsilon)\dot\varphi = \varphi.$$

It is important to note that when the tail-cancellation condition (A.17) is obeyed, the contributions of $\tilde{J}^0(\varepsilon)$ *cancel out* in the atomic sphere at the origin exactly, at least for low orbital angular momenta $l$. However, if one were to use the MTOs $\chi(\varepsilon)$ with a variational method, these unwanted contributions, generally varying linearly with the energy $\varepsilon$, would remain and reduce the precision of the band calculations. The energy dependence of $\tilde{J}^0(\varepsilon)$ is obviously weak, since it matches continuously to energy-independent envelope function $K^0$, and can be eliminated completely to first order when using fixed energy orbitals. At any given energy $\varepsilon = \varepsilon_{\nu ll}$ in a region of interest we select the head

$$\tilde\chi^{\text{head}} \equiv |\varphi(\varepsilon)\rangle \frac{N(\varepsilon)}{N(\varepsilon_\nu)} + |\tilde{J}^0(\varepsilon)\rangle \frac{P^0(\varepsilon)}{N(\varepsilon_\nu)}$$

(the part of the MTO in its own atomic sphere) and choose $\tilde{J}^0(\varepsilon_\nu)$ from the condition $\dot\chi^{\text{head}}(\varepsilon_\nu) = 0$ which gives us a condition for $\tilde{J}^0(\varepsilon_\nu)$:

$$|\dot\varphi(\varepsilon_\nu)\rangle + |\varphi(\varepsilon_\nu)\rangle \frac{\dot N(\varepsilon_\nu)}{N(\varepsilon_\nu)} + |\tilde{J}^0(\varepsilon_\nu)\rangle \frac{\dot P^0(\varepsilon_\nu)}{N(\varepsilon_\nu)} = 0. \tag{A.21}$$

This relation has a pure $L$-character since the matrices of potential parameters are diagonal matrices with respect to site and orbital indices. Thus, we arrive at the following substitution:

$$\tilde{J}^0(\varepsilon_\nu, r) \rightarrow -\frac{\dot\varphi(\varepsilon_\nu) + \varphi(\varepsilon_\nu)\dot N^0(\varepsilon_\nu)/N^0(\varepsilon_\nu)}{\dot P^0(\varepsilon_\nu)/N^0(\varepsilon_\nu)} \equiv -\dot\varphi^0(r)\left(\frac{2}{w}\dot P^0\right)^{-1/2} \tag{A.22}$$

where we have used (A.19) and introduced *the definitions*

$$\dot\varphi^0(r) \equiv \dot\varphi(r) + \varphi(r)\sigma^0 \tag{A.23}$$
$$\sigma^0 \equiv \dot N^0/N^0 = \ddot P^0/(2\dot P^0) \tag{A.24}$$

where we imply that all quantities without arguments are evaluated at $\varepsilon = \varepsilon_\nu$. After we perform the substitution (A.22) in the head *and the tails* of the MTO we obtain the energy-*independent* basis $\tilde\chi(\varepsilon_\nu)$, which is accurate to terms $O(\varepsilon - \varepsilon_\nu)^2$.

Substituting (A.22) in the equation (A.15), and using (A.19) we finally arrive at the working expression for the standard linear muffin-tin orbital (we mark this representation by *zero* as the superscript):

$$\begin{aligned}
\chi^0 &\equiv |\varphi\rangle + |\tilde{J}^0\rangle(P^0 - S^0)/N^0 + |K^0\rangle^i/N^0 \\
&= |\varphi\rangle - |\dot\varphi^0\rangle(\dot P^0)^{-1/2}(P^0 - S^0)(\dot P^0)^{-1/2} + |K^0\rangle^i/N^0 \\
&= |\varphi\rangle + |\dot\varphi^0\rangle h^0 + |K^0\rangle^i \left(\frac{2}{w}\dot P^0\right)^{-1/2}.
\end{aligned} \tag{A.25}$$

Here we have introduced the two-centre *first-order Hamiltonian*

$$h^0 = -(\dot P^0)^{-1/2}(P^0 - S^0)(\dot P^0)^{-1/2}$$

and have used (A.19). One can now use the constructed LMTOs to solve the Schrödinger equation by minimizing $\langle\psi_i|\hat H - E^{(i)}|\psi_i\rangle$ with $\psi_i$ taken from equation (A.16). This leads to the standard linear problem

$$\sum_{I'L'} C_{I'L'}^{(i)}(H_{I'L'IL} - E^{(i)}O_{I'L'IL}) = 0$$

where the matrix elements are

$$H \equiv \langle \chi^0 | -\nabla^2 + v | \chi^0 \rangle$$
$$= h^0 + h^0 \sigma^0 h^0 + (1 + h^0 \sigma^0) \varepsilon_v (\sigma^0 h^0 + 1) + h^0 \varepsilon_v p h^0 + {}^i \langle \kappa | -\nabla^2 + v | \kappa \rangle^i$$
(A.26)

$$O \equiv \langle \chi^0 | \chi^0 \rangle = (1 + h^0 \sigma^0)(\sigma^0 h^0 + 1) + h^0 p h^0 + {}^i \langle \kappa | \kappa \rangle^i$$
(A.27)

where we have introduced an extra potential parameter

$$p_{Il} = \langle (\dot{\varphi}_{Il})^2 \rangle = -\ddot{\varphi}_l(s_I)/3\varphi_l(s_I)$$

and

$$|\kappa\rangle^i \equiv |K^0\rangle^i \left( \frac{2}{w} \dot{P}^0 \right)^{-1/2}.$$

The last terms in (A.26) and (A.27) are the so-called combined corrections. Recalling that $|\kappa\rangle^i$ is a solution of the Schrödinger equation at $\varepsilon = \kappa^2$ we see that ${}^i \langle \kappa | -\nabla^2 + v | \kappa \rangle^i = \kappa^2 ({}^i \langle \kappa | \kappa \rangle^i)$. With the use of Green's theorem and the definition of $|\kappa\rangle^i$ we can express these integrals in terms of the canonical structure constants $S^0$ and their first energy derivatives $\dot{S} \equiv \partial S / \partial \kappa^2$ [105, 106]. Therefore, all one needs to know in order to solve the Schrödinger equation for a crystal is the values of the partial wavefunctions and their derivatives with respect to the coordinate and energy at the atomic sphere surface at the energy $\varepsilon_v$ in the window of interest. These values are easily found from radial solutions of the Schrödinger equation for the muffin-tin potentials.

Let us now discuss in more detail the first-order Hamiltonian $h^0$:

$$h^0 = -|\dot{\varphi}^0\rangle (\dot{P}^0)^{-1/2} (P^0 - S^0)(\dot{P}^0)^{-1/2} = c^0 - \varepsilon_v + \sqrt{d^0} S^0 \sqrt{d^0}$$
(A.28)

where

$$\sqrt{d^0} \equiv (\dot{P}^0)^{-1/2} = \left( \frac{2}{w} \right)^{1/2} W[\varphi, J^0] = \left( \frac{s}{2} \right)^{1/2} \left( \frac{s}{w} \right)^{l+1/2} \varphi(s) \frac{l - D}{2l + 1}$$

$$c^0 - \varepsilon_v \equiv -P^0 / \dot{P}^0 = -\frac{2}{w} W[\varphi, K^0] W[\varphi, J^0] = s\varphi^2(s) \frac{(D + l + 1)(l - D)}{2l + 1}$$

where $D \equiv D[\varphi_v(r)] \equiv \partial \ln \varphi_v(r) / \partial \ln r$ at $r = s$. Potential parameters $c^0 \equiv c^0_{Il} \delta_{I'L'IL}$ and $d^0 \equiv d^0_{Il} \delta_{I'L'IL}$ are the diagonal matrices easily found from the preceding equations.

It is instructive to relate the potential parameters $c^0$ and $d^0$ to the scattering characteristics of the muffin-tin potential for which they are calculated. First we recall the relation between the potential function and the scattering phase shift, namely $(P^0)^{-1} \propto -\tan \delta_l$. Since $\tan \delta_l$ has a resonant character [104], it is customary to parametrize $[P^0(\varepsilon)]^{-1}$ in the following functional form:

$$[P^0(\varepsilon)]^{-1} = \frac{\Delta}{\varepsilon - C} + \gamma$$
(A.29)

where the *canonical potential parameters* $C_l$ (centre), $\Delta_l$ (width), and $\gamma_l$ of the band $l$ are readily found from $\varphi'(\varepsilon, s)$, $\varphi(\varepsilon, s)$, and their derivatives with respect to energy. $C_l$ is the energy at which $D(C_l) = -l - 1$. Substituting this expression into equation (A.28) we arrive at the following relations for the parameters of the first-order Hamiltonian $h^0$:

$$\sqrt{d^0} = \sqrt{\Delta} \left( 1 + \gamma \frac{\varepsilon_v - C}{\Delta} \right)$$

$$c^0 - \varepsilon_v = (C - \varepsilon_v) \left( 1 + \gamma \frac{\varepsilon_v - C}{\Delta} \right).$$
(A.30)

If one selects $\varepsilon_v = C$, then $d^0 = \Delta$, $c^0 = C$, and the first-order Hamiltonian becomes

$$h^0 = \sqrt{\Delta} S^0 \sqrt{\Delta} \qquad H \approx H^{(1)} \equiv \varepsilon_v + \sqrt{\Delta} S^0 \sqrt{\Delta}.$$

Thus we have arrived at the simplest two-centre first-order Hamiltonian in real space, useful in qualitative discussions of the band problem. For an ideal crystal one transfers $H$ and $O$ into $k$-space by substituting for the only site-non-diagonal matrix $S$ with its Fourier transform, calculated by the standard Ewald procedure. Since the canonical potential parameters $C_l$, $\Delta_l$, and $\gamma_l$ are somewhat less sensitive to a choice of $\varepsilon_v$, one can substitute equation (A.3) into equation (A.18) and compare with the resonant expression for $P^0$, equation (A.29), with the following results suitable for practical calculations:

$$C - \varepsilon_v = -\frac{W[\varphi, K^0]}{W[\dot{\varphi}, K^0]} = -\frac{\varphi(s)}{\dot{\varphi}(s)} \frac{D[\varphi(s)] + l + 1}{D[\dot{\varphi}(s)] + l + 1}$$

$$\sqrt{\Delta} = \frac{\sqrt{w/2}}{|W[\dot{\varphi}, K^0]|} = \frac{1}{\sqrt{2}} \frac{(s/w)^{l+1/2}}{\sqrt{s}|\dot{\varphi}(s)|} \frac{1}{|D[\dot{\varphi}(s)] + l + 1|} \tag{A.31}$$

$$\gamma \equiv \frac{1}{P[\dot{\varphi}(s)]} = \frac{W[\dot{\varphi}, J^0]}{W[\dot{\varphi}, K^0]} = \frac{1}{2(2l+1)} \left(\frac{s}{w}\right)^{2l+1} \frac{D[\dot{\varphi}(s)] - l}{D[\dot{\varphi}(s)] + l + 1}.$$

In order to see that the width $\Delta$ has the correct dimensionality of Ryd we recall that we are working in atomic units, where $s$ has the dimensionality $(\text{Ryd})^{-1/2}$ and $\varphi$ has a dimensionality of $s^{-3/2} = (\text{Ryd})^{3/4}$. This completes our discussion of the *standard* LMTO method.

In order to treat large systems (meaning a few hundred atoms or more) one would like to have a localized, accurate, and minimal basis set which one can use in a *fully ab initio* method. The standard LMTO method operates with long-range orbitals and is not suitable for this purpose. However, the LMTOs can be made localized with the use of the information about the environment of each atom [83]. To screen the long-ranged tail of $K^0(\vec{r}_I)$ one introduces, instead of $J^0$, a function $J^\alpha$ in the expansion (A.7) which includes the screening multipoles with charges $\alpha_{Il}$ (=0 for $l > l_\alpha$, where usually $l_\alpha \leqslant 2$):

$$J_{Il}^\alpha(r_I) \equiv J_{Il}^0(r_I) - \alpha_{Il} K_{Il}^0(r_I). \tag{A.32}$$

Introducing such screening charges will inevitably change the multipole field everywhere, including the vicinity of the origin, where $K^0$ was centred. Thus, one has to introduce new envelope functions $K^\alpha$ which are defined in all space and are *linearly* related to $K^0$ everywhere. The tail expansion now reads

$$K_{IL}^\alpha(\vec{r}_I) = -\sum_{L'} J_{I'L}^\alpha(\vec{r}_{I'}) S_{I'L'IL}^\alpha \tag{A.33}$$

*where now $S^\alpha$ has non-zero on-site elements*, since $K^\alpha \neq K^0$ at the origin. Let us use compressed matrix indices $a \equiv IL$ which imply a summation over repeated indices, and mark by the superscript $\infty$ the envelope functions defined in the whole space, while the functions without this sign are assumed to be truncated beyond their atomic sphere of origin. Then we have

$$^\infty K_a^\alpha \equiv K_{a'}^0 \delta_{a'a} - J_{a'}^\alpha S_{a'a}^\alpha = K_{a'}^0(\delta_{a'a} + \alpha_{a'} S_{a'a}^\alpha) - J_{a'}^0 S_{a'a}^\alpha$$

$$^\infty K_a^0 \equiv K_{a'}^0 \delta_{a'a} - J_{a'}^0 S_{a'a}^0. \tag{A.34}$$

From our earlier discussion we expect these envelope functions to be related by a unitary transformation $U_{a'a}$ such that $^\infty K_a^\alpha = {}^\infty K_{a'}^0 U_{a'a}$. Comparing this with equation (A.34) we obtain a relation between $S^\alpha$ and $S^0$ by comparing the coefficients before $K^0$ and $J^0$ in two of these expansions:

$$U_{a'a} = \delta_{a'a} + \alpha_{a'} S_{a'a}^0 \equiv 1 + \alpha S^0$$

$$S_{a'a}^\alpha = S_{a'a''}^0 U_{a''a} = S^0(1 + \alpha S^\alpha) = S^0 + S^0 \alpha S^\alpha. \tag{A.35}$$

Thus we have obtained a Dyson-like equation for the screened structure constants. A solution of the Dyson equation (A.35) can be given in the following convenient form:

$$S^\alpha = S^0 + S^0 \alpha S^0 + \cdots \equiv \alpha^{-1}(1 + \alpha S^0 + \alpha S^0 \alpha S^0 + \cdots - 1)$$
$$= \alpha^{-1}\left[(1 - \alpha S^0)^{-1} - 1\right] = \alpha^{-1}(\alpha^{-1} - S^0)^{-1}\alpha^{-1} - \alpha^{-1}. \tag{A.36}$$

Since in a regular lattice with lattice vectors $\vec{R}$ and $\vec{T}$ and atomic position $\vec{\tau}$ within the unit cell,

$$\left[\alpha_{\vec{R}\vec{\tau}l}^{-1}\delta_{\vec{R}\vec{\tau}L,\vec{R}+\vec{T}\vec{\tau}'L'} - S^0_{\vec{R}\vec{\tau}L,\vec{R}+\vec{T}\vec{\tau}'L'}\right]^{-1} = (1/N)\sum_{\vec{k}}\left[\alpha_{\vec{R}\vec{\tau}l}^{-1} - S^0_{\vec{\tau}L,\vec{\tau}'L'}(\vec{k})\right]^{-1}\mathrm{e}^{-\mathrm{i}\vec{k}\cdot\vec{T}} \tag{A.37}$$

where by definition

$$S^0_{\tau L,\tau'L'}(\vec{k}) = \sum_{\vec{T}} S^0_{\vec{R}\vec{\tau}L,\vec{R}+\vec{T}\vec{\tau}'L'}\mathrm{e}^{-\mathrm{i}\vec{k}\cdot\vec{T}} \tag{A.38}$$

the long-range behaviour of $S^\alpha$ is defined by the analytical properties of the Fourier-transformed function $S^0(\vec{k})$, which is customarily evaluated by means of a Ewald summation. It follows from the form of the canonical structure constants (A.9) that $S^0(\vec{k})$ is bounded from above; therefore it is possible to find sufficiently small and positive $\alpha$s for which $\det[\alpha^{-1} - S^0(\vec{k})] \neq 0$. Since there are no poles for real values of $k$ in equation (A.37), its Fourier transform will be falling off exponentially with distance $|\vec{R} - \vec{R}'|/w$. This conclusion does not, of course, depend on whether the lattice is ordered or not. As mentioned above, it should just be sufficiently close packed or, for open structures, packed with empty spheres so that this recipe for screening might work.

We now give a simple example illustrating this procedure for screening LMTOs. Consider a system with only one s orbital per site [106]. The relevant structure constant is $S_{\mathrm{ss}\sigma}$, equation (A.9), whose lattice Fourier transform is

$$S(\vec{k}) = \sum_{\vec{T}} \frac{-2w}{T}\mathrm{e}^{-\mathrm{i}\vec{k}\cdot\vec{T}} \approx -\int \frac{\mathrm{d}^3T}{\Omega}\frac{2w}{T}\mathrm{e}^{-\mathrm{i}\vec{k}\cdot\vec{T}} + \mathrm{constant} = -\frac{6}{(kw)^2} + \mathrm{constant} \tag{A.39}$$

where the constant is determined from the condition that the on-site element of the $S$-matrix in real space vanishes, $S_{\vec{R}=0} = 0 = (1/N)\sum_{\vec{k}} S(\vec{k})$, $\Omega = (4\pi/3)w^3$ is the unit-cell volume, and the Brillouin zone was approximated by a sphere with radius $k_0$ such that $(4\pi/3)k_0^3 = (2\pi)^3/\Omega$. After performing the integration we obtain

$$S(\vec{k}) = -\frac{6}{(kw)^2} + \frac{1}{\alpha_c} \tag{A.40}$$

where $\alpha_c = 2^{-5/3}3^{-2/3}\pi^{2/3} = 0.32$. Substituting this into (A.36) we obtain for $\vec{R} \neq 0$

$$S^\alpha_{0\vec{R}} = (1/N)\sum \frac{1}{\alpha}\left(\frac{1}{\alpha} - \frac{1}{\alpha_c} + \frac{6}{(kw)^2}\right)^{-1}\frac{1}{\alpha}\mathrm{e}^{-\mathrm{i}\vec{k}\cdot\vec{R}}. \tag{A.41}$$

Note that this expression has no poles only if $\alpha < \alpha_c = 0.32$, i.e. the screening charges cannot exceed a certain limit. In this case, if we were to extend the integration over $k$ to infinity, we would get

$$S^\alpha_{0\vec{R}} = -\frac{\alpha_c^2}{(\alpha_c - \alpha)^2}\frac{2w}{R}\mathrm{e}^{-\sqrt{6\xi}(R/w)} \tag{A.42}$$

where we have introduced $\xi \equiv \alpha\alpha_c/(\alpha_c - \alpha)$. Indeed, this matrix decays exponentially in real space. This seemingly obvious statement [83] does not actually apply to real lattices. Obviously, the decay will be strictly exponential only when *all* multipole moments of the

given charge distribution are *exactly zero*. Note that an actual integration is cut off at the Brillouin-zone boundary, and the actual functional dependence on the distance $R/w$ is given in terms of a damped oscillatory asymptotic expansion. The complicated shape of the actual Brillouin zone, however, produces strong compensation of different oscillatory terms, so the actual behaviour is close to fast exponential decay. The optimal numerical choice for close-packed structures is $\alpha_s = 0.3485$, $\alpha_p = 0.053\,03$, and $\alpha_d = 0.0107$, and the orbitals are essentially limited to first and second neighbours [83].

A screened LMTO $\chi^\alpha$ can be constructed in *exactly* the same way as the standard LMTO (A.25), with the replacements of all superscripts $0 \to \alpha$, and by substituting for the potential parameter $\gamma$ with $\gamma - \alpha$ in (A.29), (A.30). Indeed, the matching conditions now read

$$J_{Il}^\alpha(r) \equiv \frac{(r_I/w)^l}{2(2l+1)} - \alpha_l \left(\frac{w}{r_I}\right)^{l+1} \to \tilde{J}_{Il}^\alpha(r_I) \tag{A.43}$$

$$K_{Il}^0(r) \equiv \left(\frac{w}{r_I}\right)^{l+1} \to \varphi_{Il}(\varepsilon, r_I) N_{Il}^\alpha(\varepsilon) + \tilde{J}_{Il}^\alpha(r_I) P_{Il}^\alpha(\varepsilon). \tag{A.44}$$

We immediately have from these matching conditions

$$P^\alpha(\varepsilon) = \frac{W[\varphi(\varepsilon), K^0]}{W[\varphi(\varepsilon), J^\alpha]} = \frac{P^0(\varepsilon)}{1 - \alpha P^0(\varepsilon)} \tag{A.45}$$

$$N^\alpha(\varepsilon) = \frac{W[K^0, J^\alpha]}{W[\varphi(\varepsilon), J^\alpha]} = \left[\frac{w}{2} \dot{P}^\alpha(\varepsilon)\right]^{1/2} \tag{A.46}$$

where $P^0$ is from (A.18). Rewriting (A.45) in the form

$$\left[P^\alpha(\varepsilon)\right]^{-1} = \left[P^0(\varepsilon)\right]^{-1} - \alpha$$

and substituting the resonant expansion (A.29), we arrive at

$$\left[P^\alpha(\varepsilon)\right]^{-1} = \frac{\Delta}{\varepsilon - C} + \gamma - \alpha. \tag{A.47}$$

In the expression for the Hamiltonian in the screened LMTO representation below (A.54) we shall see that all potential parameters are defined by $P^\alpha(\varepsilon)$ and its energy derivatives. Comparing (A.47) with (A.29) we see that indeed the only difference is the replacement $\gamma \to \gamma - \alpha$. Repeating our reasoning for the choice of $\tilde{J}^\alpha$, we arrive at the following substitution, which makes the screened LMTO independent of energy everywhere:

$$\tilde{J}^\alpha(\varepsilon_\nu, r) \to -\frac{\dot{\varphi}(\varepsilon_\nu) + \varphi(\varepsilon_\nu)\dot{N}^\alpha(\varepsilon_\nu)/N^\alpha(\varepsilon_\nu)}{\dot{P}^\alpha(\varepsilon_\nu)/N^\alpha(\varepsilon_\nu)} \equiv -\dot{\varphi}^\alpha(r)\left(\frac{2}{w}\dot{P}^\alpha\right)^{-1/2} \tag{A.48}$$

where

$$\dot{\varphi}^\alpha(r) \equiv \dot{\varphi}(\varepsilon_\nu) + \varphi(\varepsilon_\nu)\sigma^\alpha \tag{A.49}$$

$$\sigma^\alpha \equiv \frac{\dot{N}^\alpha(\varepsilon_\nu)}{N^\alpha(\varepsilon_\nu)} = \ddot{P}^\alpha/(2\dot{P}^\alpha) \tag{A.50}$$

and obtain a screened LMTO (again, we want it to start from just the atomic orbital $|\varphi\rangle$, so we divide the initial MTO (A.44) by $N^\alpha(\varepsilon_\nu)$):

$$\begin{aligned}
\chi^\alpha &\equiv |\varphi\rangle + |\tilde{J}^\alpha\rangle(P^\alpha - S^\alpha)/N^\alpha + |K^\alpha\rangle^i/N^\alpha \\
&= |\varphi\rangle - |\dot{\varphi}^\alpha\rangle(\dot{P}^\alpha)^{-1/2}(P^\alpha - S^\alpha)(\dot{P}^\alpha)^{-1/2} + |K^\alpha\rangle^i/N^\alpha \\
&= |\varphi\rangle + |\dot{\varphi}^\alpha\rangle h^\alpha + |K^\alpha\rangle^i \left(\frac{2}{w}\dot{P}^\alpha\right)^{-1/2}.
\end{aligned} \tag{A.51}$$

Here we have introduced the two-centre *first-order screened Hamiltonian*

$$h^\alpha = -(\dot{P}^\alpha)^{-1/2}(P^\alpha - S^\alpha)(\dot{P}^\alpha)^{-1/2}.$$

One can now use the screened LMTOs to solve the Schrödinger equation, which leads to the standard linear problem

$$\sum_{I'L'} C_{I'L'}^{\alpha(i)}(H_{I'L'IL}^\alpha - E^{(i)} O_{I'L'IL}^\alpha) = 0$$

where the matrix elements are

$$H^\alpha \equiv \langle \chi^\alpha | -\nabla^2 + v | \chi^\alpha \rangle$$
$$= h^\alpha(1 + \sigma^\alpha h^\alpha) + (1 + h^\alpha \sigma^\alpha)\varepsilon_\nu(\sigma^\alpha h^\alpha + 1) + h^\alpha \varepsilon_\nu p h^\alpha + {}^i\langle \kappa^\alpha | -\nabla^2 + v | \kappa^\alpha \rangle^i$$
$$(A.52)$$

$$O^\alpha \equiv \langle \chi^\alpha | \chi^\alpha \rangle = (1 + h^\alpha \sigma^\alpha)(\sigma^\alpha h^\alpha + 1) + h^\alpha p h^\alpha + {}^i\langle \kappa^\alpha | \kappa^\alpha \rangle^i \qquad (A.53)$$

where

$$|\kappa^\alpha\rangle^i \equiv |K^\alpha\rangle^i \left(\frac{2}{w}\dot{P}^\alpha\right)^{-1/2}.$$

One can gain more insight into the screened LMTOs by looking at the parametrization of the Hamiltonian

$$h^\alpha = -(\dot{P}^\alpha)^{-1/2}(P^\alpha - S^\alpha)(\dot{P}^\alpha)^{-1/2} = c^\alpha - \varepsilon_\nu + \sqrt{d^\alpha} S^\alpha \sqrt{d^\alpha} \qquad (A.54)$$

where

$$\sqrt{d^\alpha} \equiv (\dot{P}^\alpha)^{-1/2} = \left(\frac{2}{w}\right)^{1/2} W[\varphi, J^\alpha] = \sqrt{\Delta}\left[1 + (\gamma - \alpha)\frac{\varepsilon_\nu - C}{\Delta}\right]$$

$$c^\alpha - \varepsilon_\nu \equiv -P^\alpha/\dot{P}^\alpha = -\frac{2}{w}W[\varphi, K^\alpha]W[\varphi, J^\alpha] = (C - \varepsilon_\nu)\left[1 + (\gamma - \alpha)\frac{\varepsilon_\nu - C}{\Delta}\right]$$

where we again expressed everything in terms of our primary potential parameters $C$, $\Delta$, and $\gamma$ (equation (A.31)). Note that it differs from (A.30) merely by substitution $\gamma \to \gamma - \alpha$, as we discussed earlier; see equation (A.47).

This finalizes our construction of the tight-binding LMTO. We see that the first-order Hamiltonian (A.54) is short ranged because it contains the screened $S^\alpha$-matrices. The range of $H^\alpha$ (equation (A.52)) is larger, since it contains two-hop terms, like $h\sigma h$. These terms may be treated exactly or perturbatively.

Finally we offer some comments on the accuracy of the LMTO method and recent improvements. We have used the atomic spheres approximation (ASA) by taking the energy of our basis wave functions to be $\kappa_\nu^2 = 0$. Thus we make an error that is linear in $E - E_\nu$ in our basis in the interstitial region, whereas inside the MT spheres the corrections are proportional to $(\varepsilon - \varepsilon_\nu)^2$. The LMTO-ASA is a powerful tool for close-packed solids, but these inaccuracies make it unsuitable for calculation of forces and dynamics. Full potential LMTO has been designed to allow the inclusion of any non-ASA corrections and the accurate treatment of any interstitial region (see [107] and references therein). It is somewhat involved, making use of multiple-$\kappa_\nu^2$ basis sets to describe accurately the interstitial region.

It is worth mentioning that it is in principle possible to reformulate the standard LMTO method so that (i) the error of the basis in the interstitial region is not $\varepsilon - \varepsilon_\nu$ but $(\varepsilon - \varepsilon_\nu)^2$, (ii) the basis set is localized in space, and (iii) the full (non-spherical) charge density and potential are expanded via the same functions as were used to construct the basis [108]. The main idea of using solutions of the Schrödinger equation in the MT spheres as well as the interstitial region (A.5) with matching at the MT spheres does not change. However, in order

to localize the basis one has to solve the wave equation (A.5) such that the solution has zero $Y_{L'}(r_I)$-projections on other non-touching spheres with radii $a_{I'l'}$ (these spheres are analogous to hard impenetrable spheres). Such solutions always exist, they produce a complete set, and they may be localized. One needs these functions to obey the matching conditions of the wave functions at the MT spheres, which is done by introducing so-called *kinky* partial waves. The matching conditions at the MT spheres are then reduced to algebraic form involving the *kink* (KKR) matrix $K$. The LMTO overlap matrix and the Hamiltonian matrix can then be expressed solely in terms of the kink KKR matrix at the selected energy $E_\nu$ and its first *three* energy derivatives $\dot{K}$, $\ddot{K}$, and $\dddot{K}$. This, as we have seen, extends the spatial range of the matrix elements.

## References

[1]  Ashcroft N W and Mermin D 1976 *Solid State Physics* (Tokyo: Holt-Saunders)
[2]  Tan K E, Bratkovsky A M, Harris R M, Horsfield A P, Nguyen-Manh D, Pettifor D G and Sutton A P 1997 *Modell. Simul. Mater. Sci. Eng.* **5** 187
[3]  Kohn W and Sham L J 1965 *Phys. Rev.* **140** A1133
[4]  Sutton A P, Finnis M W, Pettifor D G and Ohta Y 1988 *J. Phys. C: Solid State Phys.* **21** 35
[5]  Goodwin L, Skinner A J and Pettifor D G 1989 *Europhys. Lett.* **9** 701
[6]  Xu C H, Wang C Z, Chan C T and Ho K M 1992 *J. Phys.: Condens. Matter* **4** 6047
[7]  Kwon I, Biswas R, Wang C Z, Ho K M and Soukoulis C M 1994 *Phys. Rev.* B **49** 7242
[8]  Tang M S, Wang C Z, Chan C T and Ho K M 1996 *Phys. Rev.* B **53** 979
[9]  Allen P B, Broughton J Q and McMahan A K 1986 *Phys. Rev.* B **34** 859
[10]  Khan F S and Broughton J Q 1989 *Phys. Rev.* B **39** 3688
[11]  Sawada S 1990 *Vacuum* **41** 612
[12]  Mercer J L and Chou M Y 1993 *Phys. Rev.* B **47** 9366
[13]  Mercer J L and Chou M Y 1994 *Phys. Rev.* B **49** 8506
[14]  Sigalas M M and Papaconstantopoulos D A 1994 *Phys. Rev.* B **49** 1574
[15]  Cohen R E, Mehl M J and Papaconstantopoulos D A 1994 *Phys. Rev.* B **50** 14 694
[16]  Menon M and Subbaswamy K R 1997 *Phys. Rev.* B **55** 9231
[17]  Liu F 1995 *Phys. Rev.* B **52** 10 677
[18]  Pople J A, Santry D P and Segal G A 1965 *J. Chem. Phys.* **43** S129
[19]  Pople J A and Segal G A 1965 *J. Chem. Phys.* **43** S136
[20]  Pople J A and Segal G A 1966 *J. Chem. Phys.* **44** 3289
[21]  Santry D P and Segal G A 1967 *J. Chem. Phys.* **47** 158
[22]  Pople J A, Beveridge D L and Dobosh P A 1967 *J. Chem. Phys.* **47** 2026
[23]  Harris J 1985 *Phys. Rev.* B **31** 1770
[24]  Foulkes W M C 1987 *PhD Thesis* University of Cambridge
[25]  Foulkes W M C and Haydock R 1989 *Phys. Rev.* B **39** 12 520
[26]  Finnis M W 1990 *J. Phys.: Condens. Matter* **2** 331
[27]  Polatoglou H M and Methfessel M 1988 *Phys. Rev.* B **37** 10 403
[28]  Chan C T, Vanderbilt D and Louie S G 1986 *Phys. Rev.* B **33** 2455
[29]  Sankey O F and Niklewski D J 1989 *Phys. Rev.* B **40** 3979
[30]  Horsfield A P 1997 *Phys. Rev.* B **56** 6594
[31]  Porezag D, Frauenheim Th, Köhler Th, Seifert G and Kaschner R 1995 *Phys. Rev.* B **51** 12 947
[32]  Slater J C and Koster G F 1954 *Phys. Rev.* **94** 1498
[33]  Drabold D A, Fedders P A, Sankey O F and Dow J D 1990 *Phys. Rev.* B **42** 5135
[34]  Drabold D A, Fedders P A, Klemm S and Sankey O F 1991 *Phys. Rev. Lett.* **67** 2179
[35]  Fedders P A, Drabold D A and Klemm S 1992 *Phys. Rev.* B **45** 4048
[36]  Fedders P A, Fu Y and Drabold D A 1992 *Phys. Rev. Lett.* **68** 1888
[37]  Fedders P A and Drabold D A 1993 *Phys. Rev.* B **47** 13 277
[38]  Kilian K A, Drabold D A and Adams J B 1993 *Phys. Rev.* B **48** 17 393
[39]  Fedders P A 1995 *Phys. Rev.* B **52** 1729
[40]  Sankey O F, Niklewski D J, Drabold D A and Dow J D 1990 *Phys. Rev.* B **41** 12 750
[41]  Adams G B and Sankey O F 1991 *Phys. Rev. Lett.* **67** 867
[42]  Feil H, Zandvliet J W, Tsai M-H, Dow J D and Tsong I S T 1992 *Phys. Rev. Lett.* **69** 3076

[43] Adams G B, Page J B, Sankey O F, Sinha K, Menendez J and Huffman D R 1991 *Phys. Rev.* B **44** 4052
[44] O'Keeffe M, Adams G B and Sankey O F 1992 *Phys. Rev. Lett.* **68** 2325
[45] Drabold D A, Wang R, Klemm S, Sankey O F and Dow J D 1991 *Phys. Rev.* B **43** 5132
[46] Yang S H, Drabold D A and Adams J B 1993 *Phys. Rev.* B **48** 5261
[47] Alfonso D R, Yang S H and Drabold D A 1994 *Phys. Rev.* B **50** 15 369
[48] Drabold D A, Fedders P A and Stumm P 1994 *Phys. Rev.* B **49** 16 415
[49] Adams G B, Page J B, Sankey O F and O'Keeffe M 1994 *Phys. Rev.* B **50** 17 471
[50] Alfonso D R, Drabold D A and Ulloa S E 1995 *Phys. Rev.* B **51** 1989
[51] Alfonso D R, Drabold D A and Ulloa S E 1995 *Phys. Rev.* B **51** 14 669
[52] Cappelletti R L, Cobb M, Drabold D A and Kamitakahara W A 1995 *Phys. Rev.* B **52** 9133
[53] Demkov A A and Sankey O F 1993 *Phys. Rev.* B **48** 2207
[54] Phillips R, Drabold D A, Lenosky T, Adams G B and Sankey O F 1992 *Phys. Rev.* B **46** 1941
[55] Tsai M-H, Sankey O F and Dow J D 1992 *Phys. Rev.* B **46** 10 464
[56] Yang S H, Drabold D A, Adams J B and Sachdev A 1993 *Phys. Rev.* B **47** 1567
[57] Adams G B, O'Keeffe M, Demkov A A, Sankey O F and Huang Y-M 1994 *Phys. Rev.* B **49** 8048
[58] Caro A, Drabold D A and Sankey O F 1994 *Phys. Rev.* B **49** 6647
[59] Ortega J, Lewis J P and Sankey O F 1994 *Phys. Rev.* B **50** 10 516
[60] Demkov A A, Sankey O F, Schmidt K E, Adams G B and O'Keeffe M 1994 *Phys. Rev.* B **50** 17 001
[61] Demkov A A, Ortega J, Sankey O F and Grumbeth M P 1995 *Phys. Rev.* B **52** 1618
[62] Huang Y M, Spence J C H and Sankey O F 1995 *Phys. Rev. Lett.* **74** 3392
[63] Landman J I, Morgan C G and Schick J T 1995 *Phys. Rev. Lett.* **74** 4007
[64] Park Y K, Estreicher S K, Myles C W and Fedders P A 1995 *Phys. Rev.* B **52** 1718
[65] Demkov A A, Windl W and Sankey O F 1996 *Phys. Rev.* B **53** 11 288
[66] Demkov A A and Sankey O F 1997 *Phys. Rev.* B **56** 10 497
[67] Lewis J P, Ordejon P and Sankey O F 1997 *Phys. Rev.* B **55** 6880
[68] Demkov A A, Sankey O F, Gryko J and McMillan P F 1997 *Phys. Rev.* B **55** 6904
[69] Fritsch J, Sankey O F and Schmidt K E 1998 *Phys. Rev.* B **57** 15 360
[70] Windl W, Sankey O F and Menendez J 1998 *Phys. Rev.* B **57** 2431
[71] Delley B 1990 *J. Chem. Phys.* **92** 508
[72] Seifert G, Porezag D and Frauenheim T 1996 *Int. J. Quantum Chem.* **58** 185
[73] Eschrig H 1979 *Phys. Status Solidi* b **96** 329
[74] Seifert G, Eschrig H and Bieger W 1986 *Z. Phys. Chem.* **267** 529
[75] Frauenheim Th, Weich F, Köhler Th, Uhlmann S, Porezag D and Seifert G 1995 *Phys. Rev.* B **52** 11 492
[76] Widany J, Frauenheim Th, Porezag D, Köhler Th and Seifert G 1996 *Phys. Rev.* B **53** 4443
[77] Sitch P and Frauenheim Th 1996 *J. Phys.: Condens. Matter* **8** 6873
[78] Haugk M, Elsner J and Frauenheim Th 1997 *J. Phys.: Condens. Matter* **9** 7305
[79] Jackson K A, Jungnickel G and Frauenheim Th 1998 *Chem. Phys. Lett.* **292** 235
[80] Elsner J, Jones R, Heggie M I, Sitch P K, Haugk M, Frauenheim Th, Öberg S and Briddon P R 1998 *Phys. Rev.* B **58** 12 571
[81] Ziman J 1965 *Principles in the Theory of Solids* (Cambridge: Cambridge University Press)
[82] Andersen O K 1975 *Phys. Rev.* B **12** 3060
[83] Andersen O K, Pawlowska Z and Jepsen O 1986 *Phys. Rev.* B **34** 5253
     Andersen O K and O Jepsen 1984 *Phys. Rev. Lett.* **53** 2571
[84] Haydock R, Heine V and Kelly M J 1975 *J. Phys. C: Solid State Phys.* **8** 2591
[85] Heine V, Bullett D W, Haydock R and Kelly M J 1980 *Solid State Physics* vol 35, ed F Seitz and D Turnbull
     (New York: Academic)
[86] Nowak H J, Andersen O K, Fujiwara T and Jepsen O 1991 *Phys. Rev.* B **44** 3577
[87] Bratkovsky A M and Smirnov A V 1993 *Phys. Rev.* B **48** 9606
[88] Bratkovsky A M, Smirnov A V, Nguyen Manh D and Pasturel A 1995 *Phys. Rev.* B **52** 3056
[89] Smirnov A V and Bratkovsky A M 1996 *Phys. Rev.* B **53** 8515
[90] Smirnov A V and Bratkovsky A M 1996 *Phys. Rev.* B **54** R17 371
[91] Smirnov A V and Bratkovsky A M 1997 *Phys. Rev.* B **55** 14 434
[92] Stokbro K, Chetty N, Jacobsen K W and Norskov J K 1994 *Phys. Rev.* B **50** 10 727
[93] Hansen L B, Stokbro K, Lundqvist B I, Jacobsen K W and Deaven D M 1995 *Phys. Rev. Lett.* **75** 4444
[94] Stokbro K, Jacobsen K W, Norskov J K, Deaven D M, Wang C Z and Ho K M 1996 *Surf. Sci.* **360** 221
[95] Hansen L B, Stokbro K, Lundqvist B I and Jacobsen K W 1996 *Mater. Sci. Eng.* B **37** 185
[96] Elstner M, Porezag D, Jungnickel G, Elsner J, Haugk M, Frauenheim Th, Suhai S and Seifert G 1998 *Phys. Rev.* B **58** 7260

[97]  Horsfield A P, Dunham S and Fujitani H 1998 *MRS Proc. (Fall 1998)* (Warrendale, PA: Materials Research Society) Symposium J

[98]  Löwdin P 1950 *J. Chem. Phys.* **18** 365

[99]  Lin Z and Harris J 1992 *J. Phys.: Condens. Matter* **4** 1055

[100]  Ordejon P, Artacho E and Soler J M 1996 *Phys. Rev.* B **53** R10 431

[101]  Bowler D, Fearn M, Goringe C, Horsfield A and Pettifor D 1998 *J. Phys.: Condens. Matter* **10** 3719

[102]  Mercer J L, Nelson J S, Wright A F and Stechel E B 1998 *Modell. Simul. Mater. Sci. Eng.* **6** 1

[103]  Varshalovich D A, Moskalev A N and Khersonskii V K 1989 *Quantum Theory of Angular Momentum* (Singapore: World Scientific)

[104]  Landau L D and Lifshitz E M 1977 *Quantum Mechanics, Non-Relativistic Theory* 3rd edn (Oxford: Pergamon)

[105]  Bratkovsky A M and Savrasov S Yu 1990 *J. Comput. Phys.* **88** 243

[106]  Andersen O K, Jepsen O and Sob M 1987 *Electronic Band Structure and its Applications (Springer Lecture Notes in Physics, vol 283)* ed M Yussouff (Berlin: Springer)

[107]  Savrasov S Yu and Savrasov D Yu 1992 *Phys. Rev.* B **46** 12 181
        Savrasov S Yu 1992 *Phys. Rev. Lett.* **69** 2819
        Savrasov S Yu 1998 *Phys. Rev. Lett.* **81** 2570

[108]  Andersen O K, Jepsen O and Krier G 1994 *Preprint*
        Andersen O K, Arcangeli C, Tank R W, Saha-Dasgupta T, Krier G, Jepsen O and Dasgupta I 1998 *Proc. Mater. Res. Soc.* vol 491 (Warrendale, PA: Materials Research Society) p 3