

Chapter 3

(L)APW+lo

From the HK theorems, the electron density determines all system properties. Unfortunately, there is no direct way to obtain the electron density, and it is hard to express the kinetic energy as a functional of the density as well. Therefore, we have to integrate Kohn-Sham wave functions to generate ρ (Eq. 2.12), and insert wave functions into the kinetic potential to get the kinetic energy. In order to convert the corresponding wave function equations into suitably numerically accessible matrix equations, the wave functions and electron densities are expanded into a basis set. Based on the type of basis functions a variety of DFT formulations in numerical codes exist. In this chapter, we will introduce one of the most accurate methods, (L)APW+lo, which is used in the present work.

3.1 Bloch's Theorem

The Bloch theorem [20] states that *Any eigenfunction $\Psi(\mathbf{r})$ can be written as a product of a function $\phi_{\mathbf{g}}(\mathbf{r})$ that has the periodicity of the lattice, and a plane wave $e^{i\mathbf{g}\cdot\mathbf{r}}$ with \mathbf{g} any vector in reciprocal space,*

$$\Psi(\mathbf{r}) = \phi_{\mathbf{g}}(\mathbf{r})e^{i\mathbf{g}\cdot\mathbf{r}} \quad . \quad (3.1)$$

If the reciprocal vector \mathbf{g} is written as the sum of a vector (\mathbf{k}) in the first Brillouin zone and a reciprocal lattice vector \mathbf{K} , $\mathbf{g}=\mathbf{k}+\mathbf{K}$, the Bloch theorem is rewritten as,

$$\begin{aligned} \Psi(\mathbf{r}) &= \Psi_{\mathbf{k}_n}(\mathbf{r}) = \{\phi_{\mathbf{g}}(\mathbf{r})e^{i\mathbf{K}\cdot\mathbf{r}}\}e^{i\mathbf{k}\cdot\mathbf{r}} \\ &= \phi_{\mathbf{k}_n}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}} \quad , \end{aligned} \quad (3.2)$$

where, n indicates the number of the Brillouin zone where \mathbf{g} is in, and is called *band index*.

The first term in Eq. 3.2 $\phi_{\mathbf{k}_n}(\mathbf{r})$, determines the eigenfunction, $\Psi(\mathbf{r})$. Due to the periodicity of the lattice, the straight idea to get $\phi_{\mathbf{k}_n}(\mathbf{r})$ is to sum over plane waves that have the same periodicity of the lattice. Then,

$$\phi_{\mathbf{k}_n}(\mathbf{r}) = \sum_{\mathbf{K}} c_{\mathbf{K}}^{n,\mathbf{k}} e^{i\mathbf{K}\cdot\mathbf{r}} \quad . \quad (3.3)$$

The expansion of $\Psi_{\mathbf{k}_n}(\mathbf{r})$ becomes,

$$\Psi_{\mathbf{k}_n}(\mathbf{r}) = \sum_{\mathbf{K}} c_{\mathbf{K}}^{n,\mathbf{k}} e^{i(\mathbf{K}+\mathbf{k})\cdot\mathbf{r}} \quad . \quad (3.4)$$

Popular DFT codes like VASP or CASTEP [21] are based on such a plane wave basis sets. Describing the steep potential close to the nuclei (and correspondingly highly oscillating wave functions) is demanding with a pure plane wave basis set, and requires to go to high \mathbf{K} values in Eq. 3.3. When aiming to maintain a full-potential description a suitable approach followed in the (L)APW+lo idea is thus to augment the plane wave basis functions with more localized functions.

3.2 Augmented Plane Wave (APW)

Before embarking on the (L)APW+lo method, it is advantageous to discuss the APW method introduced by Slater [22]. Considering the behavior of electrons in space, when electrons are far away from the nuclei, they show the behavior of free electrons, and are then suitably described by plane waves. While close to the nuclei, electrons bind strongly to their nuclei, their behavior is quite as in a free atom and they could be described more efficiently by atomic like functions. Therefore, the whole space can be technically divided into two regions, non-overlapping atomic spheres (so-called Muffin-tin (MT) regions) and interstitial (I) region (Fig. 3.1). Correspondingly, the potential in the whole space can be defined as

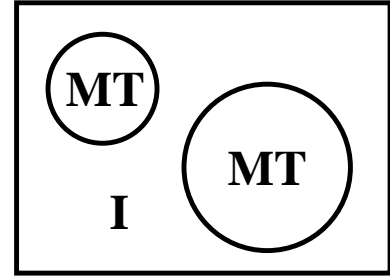


Figure 3.1: Schematic division of space into atomic sphere region (I) and interstitial region (II).

$$V(\mathbf{r}) = \begin{cases} V(\mathbf{r}) & (\mathbf{r} \in \text{MT}) \\ \text{constant} & (\mathbf{r} \in \text{I}) \end{cases} \quad , \quad (3.5)$$

and, two types of basis sets are used in the two different regions,

$$\phi_{\mathbf{k}_n}^{\text{APW}}(\mathbf{r}, \epsilon_l) = \begin{cases} \sum_{lm} A_{lm,\mathbf{k}_n} u_l(r, \epsilon_l) Y_{lm}(\hat{\mathbf{r}}) & (\mathbf{r} \in \text{MT}) \\ \frac{1}{\sqrt{\Omega}} e^{i\mathbf{k}_n \cdot \mathbf{r}} & (\mathbf{r} \in \text{I}) \end{cases} \quad . \quad (3.6)$$

In the atomic spheres (MT), the wave functions are expanded by radial functions times spherical harmonics. u_l is the solution of the radial Schrödinger equation for a spherical potential $[V(r)]$ for energy ϵ_l ,

$$-\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{du_l}{dr} \right) + \left[\frac{l(l+1)}{r^2} + V(r) - \epsilon_l \right] r u_l = 0 \quad . \quad (3.7)$$

In the interstitial region (I) plane waves are instead used to build the wave function. The coefficients A_{lm} in the atomic sphere expansion are determined by requiring that

the wave functions in the MT and the interstitial regions match each other at the MT boundary [24]. Thus, each plane wave is augmented by an atomic-like function in every atomic sphere and constitutes thus the basis set used to expand the wave function,

$$\Psi(\mathbf{r}) = \sum_n c_n \phi_{\mathbf{k}_n}(\mathbf{r}) \quad . \quad (3.8)$$

The biggest disadvantage in the APW method is that it can not get the eigenvalues from a single diagonalization due to the unknown parameter ϵ_l in Eq. 3.6. The exact ϵ_l value, which is what we want to know, is needed to describe the eigenstate $\Psi_{\mathbf{k}_n}(\mathbf{r})$ accurately. Since this energy depends on the function $u_l(r, \epsilon_l)$, the resulting eigenvalue problem is non-linear in energy. One has to set a trial energy for ϵ_l , solve Eq. 3.7 to obtain the APW basis, set up the matrix elements, and compute the determinant $|H - ES|$. If the eigenenergy does not equal ϵ_l , another trial energy must be chosen until the eigenenergy equals ϵ_l . This makes the APW method extremely inefficient.

3.3 LAPW

In order to overcome the non-linearity problem in the APW method, Anderson developed the linearized augmented plane wave method (LAPW) [23, 24]. In his idea, the radial function u_l is expanded by a Taylor expansion around ϵ_l ,

$$u_l(r, \epsilon_l) = u_l(r, \epsilon_l^1) + (\epsilon_l - \epsilon_l^1) \dot{u}_l(r, \epsilon_l^1) + O((\epsilon_l - \epsilon_l^1)^2) \quad , \quad (3.9)$$

where $\dot{u}_l = \frac{\partial u_l}{\partial \epsilon_l}$. In this case the radial function error is second order, and the energy error is of fourth order [25]. When ϵ_l^1 is set near ϵ_l , the radial function and energy errors are negligible. Substituting Eq. 3.9 into Eq. 3.6, we get the formulation of the LAPW basis set,

$$\phi_{\mathbf{k}_n}^{\text{LAPW}}(\mathbf{r}) = \begin{cases} \sum_{lm} [A_{lm, \mathbf{k}_n} u_l(r, \epsilon_l^1) + B_{lm, \mathbf{k}_n} \dot{u}_l(r, \epsilon_l^1)] Y_{lm}(\hat{\mathbf{r}}) & (\mathbf{r} \in \text{MT}) \\ \frac{1}{\sqrt{\Omega}} e^{i\mathbf{k}_n \cdot \mathbf{r}} & (\mathbf{r} \in \text{I}) \end{cases} \quad . \quad (3.10)$$

In the interstitial region, the basis set is the same as in the APW method, but in the MT spheres, the basis functions not only depend on u_l , but also on its energy derivative, \dot{u}_l . It is very clear that the LAPW method is thus more flexible than the APW in the MT spheres. To know the exact value for ϵ_l as in the APW is not important anymore. For a fixed value of ϵ_l^1 , the modified basis functions (Eq. 3.9) provide the extra flexibility to cover a large energy region around this linearization energy. In order to determine both A_{lm, \mathbf{k}_n} and B_{lm, \mathbf{k}_n} , the functions in the MT spheres are required to match the plane wave function both in value and in slope at the sphere boundary. However, the continuous derivatives require higher plane wave cutoffs to achieve a given level of convergence.

3.4 LAPW with Local Orbital (LAPW+LO)

Based on whether or not electrons in an atom participate in the chemical bonding with other atoms, the electrons can be divided into two types. One type of electrons are *core electrons*, which are extremely bound to their nucleus and are thus entirely localized in the MT sphere. The corresponding states are called *core states*. The other type of electrons are *valence electrons*, who are leaking out of the MT sphere and bond with other atoms. However, for many elements, the electrons cannot be clearly distinguished like that. Some states are neither constrained in the core states, nor lie in the valence states and are correspondingly termed *semi-core states*. They have the same angular quantum number l as the valence states but with lower principal quantum number n . When applying LAPW on these states, it is thus hard to use one ϵ_l^1 to determine the two *same* l in Eq. 3.10. The dilemma is solved by introducing *local orbitals* (LO), which are defined as

$$\phi_{lm}^{\text{LO}}(r) = \begin{cases} [A_{lm}u_l(r, \epsilon_l^1) + B_{lm}\dot{u}_l(r, \epsilon_l^1) + C_{lm}u_l(r, \epsilon_l^2)]Y_{lm}(\hat{\mathbf{r}}) & (\mathbf{r} \in \text{MT}) \\ 0 & (\mathbf{r} \in \text{I}) \end{cases} \quad (3.11)$$

Each local orbital is zero in the interstitial region and other atoms' MT sphere. The three coefficients A_{lm} , B_{lm} and C_{lm} can be determined by requiring the LO to have both zero value and zero slope at the MT boundary and be normalized.

3.5 APW+lo

It has been realized that the standard LAPW method is not the most efficient way to linearize Slater's APW method [27]. Instead, the basis set of the introduced APW+lo [26, 27] method is also energy independent and still has the same basis size as the original APW method. In order to achieve that a new local orbital (lo) is added, which is different from the LOs used to describe semicore states to gain enough variational flexibility in the radial basis functions. The lo definition is,

$$\phi_{lm}^{\text{lo}}(\mathbf{r}) = \begin{cases} [A_{lm}u_l(r, \epsilon_l^1) + B_{lm}\dot{u}_l(r, \epsilon_l^1)]Y_{lm}(\hat{\mathbf{r}}) & (\mathbf{r} \in \text{MT}) \\ 0 & (\mathbf{r} \in \text{I}) \end{cases} \quad (3.12)$$

The two coefficients A_{lm} and B_{lm} are determined by normalization, and by requiring that the local orbital has zero value at the Muffin-tin boundary. The advantage of the APW+lo method is that it has the same small basis set size as the APW method, and has the same accuracy compared to the LAPW method.

As mentioned by Madsen *et al* [27], it is also possible to use a hybrid basis set, LAPW in combination with APW [(L)APW+lo], and treat the physically important orbitals by the APW+lo method, but the polarization l -quantum numbers with LAPW. All the data presented in our work are using this hybrid basis set (L)APW+lo.

3.6 Full Potential (L)APW+lo Method

In Eq. 3.5, we supposed the potential is constant in the interstitial region and spherical in the MT region. The accuracy of (L)APW+lo method can be further improved by considering the full potential (FP), and expand it similar to the wave functions,

$$V(\mathbf{r}) = \begin{cases} \sum_{lm} V_{lm}(r) Y_{lm}(\hat{\mathbf{r}}) & (\mathbf{r} \in \text{MT}) \\ \sum_{\mathbf{G}} V_{\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}} & (\mathbf{r} \in \text{I}) \end{cases}, \quad (3.13)$$

This is also called *non-muffin-tin* correction. In this case, the radial function u_l in Eq. 3.7 is not the exact solution inside the MT sphere. It should be evaluated for the true MT potential.

3.7 Two Important Basis Set Parameters: Energy Cutoff and K-mesh

In all DFT codes for solid state calculations the energy cutoff (E_{cut}) and k-mesh (or \mathbf{k} -points) are important basis set parameters to determine the accuracy of the computational results. Both parameters must be tested in DFT calculations in order to find the optimum compromise between accuracy and implied computational burden.

When we use plane waves to construct wave functions, in principle the more plane waves, the better the results are. However, it is not necessary to use infinitely many plane waves to construct the wave function. We can limit the energy cutoff, which describes the number of plane waves used, to an optimum value. The relationship between energy cutoff and plane wave coefficient is,

$$E_{\text{cut}} = \frac{\hbar^2 K_{\text{max}}^2}{2m_e}. \quad (3.14)$$

And the wave function based on K_{max} is,

$$\psi_{\mathbf{k}_n}(\mathbf{r}) = \sum_{\mathbf{K}}^{K_{\text{max}}} C_{\mathbf{K},\mathbf{k}_n} e^{i(\mathbf{k}+\mathbf{K})\cdot\mathbf{r}}. \quad (3.15)$$

In the FP-(L)APW+lo method this energy cutoff is employed in the interstitial region. Turning to the MT sphere, the spherical harmonics should also be terminated at a maximum lm for the same reason.

The evaluation of the Kohn-Sham equations in a periodic boundary calculation requires many system quantities like the charge density to be integrated over the Brillouin Zone (BZ). Exploring the symmetry of the system, it is more specifically only necessary to integrate over the irreducible part of the Brillouin Zone (IBZ). The integrals are dealt with numerically by summing over a finite number of \mathbf{k} -points in the IBZ,

$$\int_{\text{IBZ}} \frac{1}{\Omega_{\text{IBZ}}} d\mathbf{k} \rightarrow \sum_{\mathbf{k}} \omega_{\mathbf{k}}. \quad (3.16)$$

Therefore, the denser the k -mesh, the more accurate the integrand is. Similar to E_{cut} , we should also find an optimum k -mesh at which the quantities of interest are converged.

There are two common methods to get \mathbf{k} -points: The tetrahedron method [28, 29] and the special \mathbf{k} -points method [30–32]. In our work we use the special \mathbf{k} -point method according to Monkhorst and Pack [31] throughout. The method generates a uniform \mathbf{k} -point grid in the full BZ, then uses the space group of the system to rotate the \mathbf{k} -points into the IBZ, and determine the proper weights, $\omega(\mathbf{k})$, of each special \mathbf{k} -point by dividing the number of equilibrium \mathbf{k} -points of a special \mathbf{k} -point in the BZ by the total number of points in the grid.

In metals, energy bands intersect the Fermi energy. This leads to discontinuities in the occupation and in the integrand on the Fermi surface, and can cause problems with reaching self-consistency due to charge sloshing. We can replace the step function occupation at the Fermi energy with a smoother function, like a Fermi distribution at a finite temperature to solve the problem. After the integration the resulting free energy has then to be extrapolated to $T = 0$ K.

3.8 WIEN2k Code

The **WIEN2k** code [33] is based on the FP-(L)APW+lo method. The code can only be applied to systems with periodic boundary conditions and is therefore mainly used for crystal calculations.

The **WIEN2k** code has two main parts (Fig. 3.2). One is the initialization (left part in Fig. 3.2). It is used to check if MT spheres overlap, generate a new structure file according to its space group, detect its symmetry operations, generate a k -mesh in its BZ, and get the input trial density. The other one is the self-consistency cycle (right part in Fig. 3.2). It calculates the potential used in the KS equation, diagonalizes the Hamiltonian and overlap matrices and generates eigenvalues and eigenvectors, integrates all valence states and obtains the valence electron density (ρ_{val}), solves the atomic calculation and gets the core electron density (ρ_{core}), mixes [34] the two electron densities with the old total electron density (ρ_{old}) and gets the new total electron density (ρ_{new}). Thereafter it checks if the properties (ρ_{new} , or E_{tot} , or F ...) of the system are converged, and either stops the self-consistency cycle or starts anew.

Apart from two the main parts, **WIEN2k** has lots of additional packages to evaluate a variety of system properties. This comprises geometry optimization, plotting the density of states (DOS), band structure, electron density, X-ray spectra ...

Two kinds of parallelization modes are implemented in the **WIEN2k** code to increase the calculation efficiency. One is \mathbf{k} -point parallelization, which distributes the computations for the different irreducible \mathbf{k} -points over several CPUs. This method is only useful for small calculations and a low communication bandwidth between the CPUs. If we go to big systems with few \mathbf{k} -points, fine grained parallelization can be additional applied. It diagonalizes the Hamiltonian and overlap matrices for

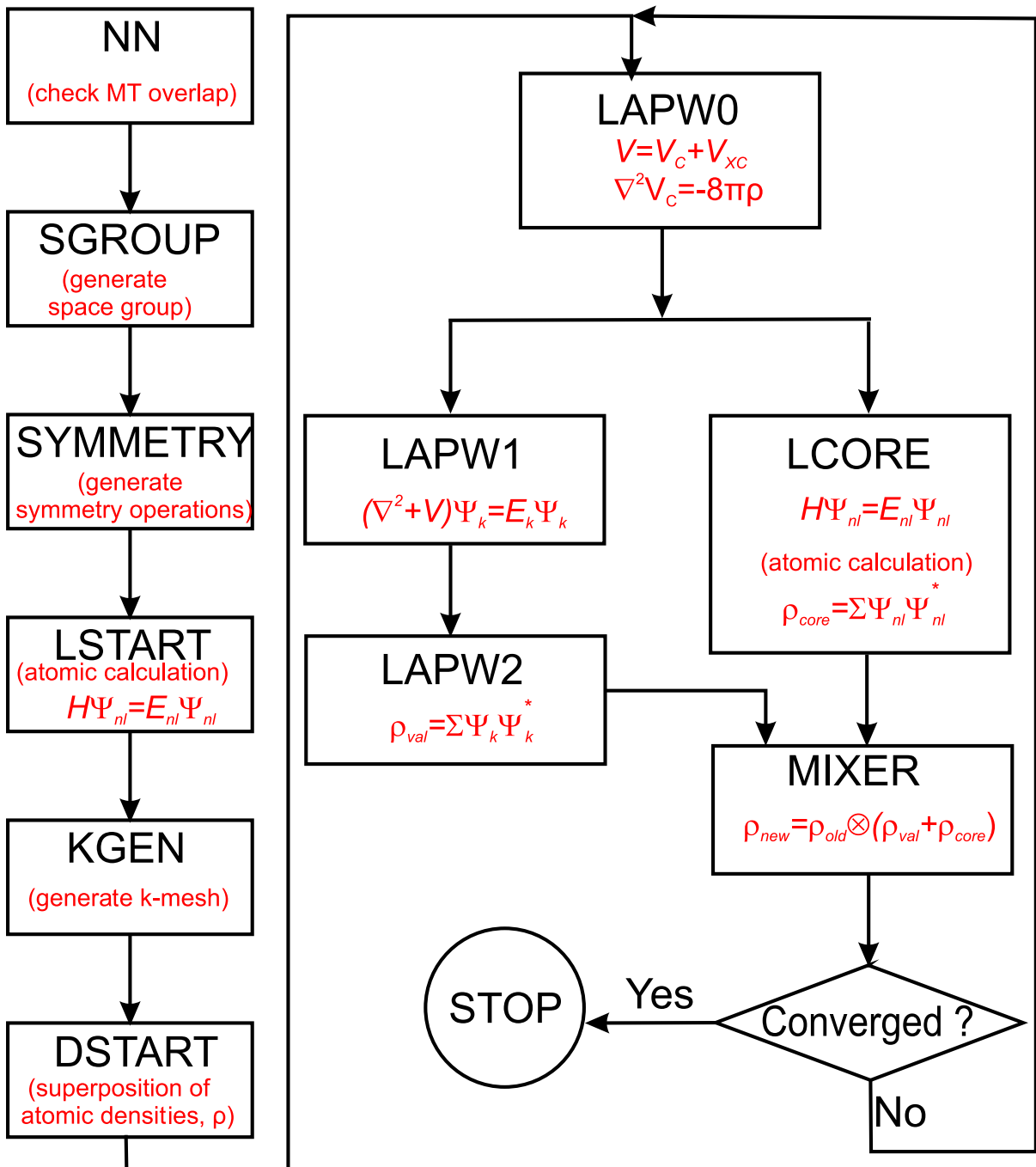


Figure 3.2: Flow chart of the **WIEN2k** code. The left part is the initialization, which is used to get the trial input electron density for starting the self-consistency cycle. The right part is the self-consistency cycle, which is used to converge the electron density.

each \mathbf{k} -point on different CPUs. The speed of this method strongly depends on the communication between the CPUs, on the number of CPUs and on the matrix size. In order to run the code more efficiently, the **WIEN2k** code can combine the two methods, which separates number of \mathbf{k} -points on different block of CPUs, and applies the fine grained parallelization on each block of CPUs.