# 4. Practical implementation of DFT and HF calculation schemes

## 4.1 Basis sets: an overview

Historically, the invention of the HF method made possible numerical solution, albeit approximative, of many-electron problems too difficult to be treated analytically. During 1930s and 40s, HF equations were looked at as a system of (integro-)differential equations, and corresponding methods of solution were applied. In many cases it turned out too difficult to solve these equations for HF-functions $\varphi(\mathbf{r})$ of general form. A further progress, at least in what regards numerical accuracy, was related with the use of *basis sets* in which the unknown functions can be expanded. Certain shape approximations on the shape of one-electron wavefunctions and/or potential helped further on to simplify the problem. An example of this development is e.g. the *muffin-tin* approximation for the potential proposed by Slater which gave rise to powerful computational methods like APW, KKR and their derivatives – LAPW and LMTO – to be discussed in some detail below. Finally, with the increase of computer power the trend of solving Kohn-Sham equations still numerically but "directly", i.e. without relation to any particular basis system, becomes noticeable and promising.

The idea to expand one-electron functions over a limited basis set and thus to reduce integro-differential equation system to a matrix eigenvalue problem is probably due to Roothaan.[27] While originally giving rise to what is referred to as the Hartree-Fock-Roothaan method, the same approach will work for the Kohn-Sham equations. Let us look at it in more detail. Eq. (2.17), with $j \neq \alpha$ dropped in both sums, reads:

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + u(x) + \int \frac{\rho(x')\,dx'}{|\mathbf{r}-\mathbf{r}'|}\right]\varphi_\alpha(x) - \int \frac{\rho(x',x)\,\varphi_\alpha(x')}{|\mathbf{r}-\mathbf{r}'|}dx' = \varepsilon_\alpha\varphi_\alpha(x) \qquad (4.1)$$

in the HF formalism, and similarly, substituting

$$-\int \frac{\rho(x',x)\,\varphi_\alpha(x')}{|\mathbf{r}-\mathbf{r}'|}dx' \quad \Rightarrow \quad V_{\mathrm{XC}}(x)\varphi_\alpha(x)$$

in the Kohn–Sham equations. We expand $\varphi_\alpha(x)$ over a basis set $\chi$:

$$\varphi_\alpha(x) = \sum_{p=1}^{Q} C_{\alpha p}\chi_p(x)\,, \qquad (4.2)$$

where the dimension of basis $Q$ is reasonably larger than the number of occupied states $N$, and $\chi_p(x)$ may retain the dependence on both spatial and spin variables. Eq. (4.1) then reduces to

$$\sum_p C_{\alpha p}\mathcal{H}\chi_p(x) = \varepsilon_\alpha \sum_p C_{\alpha p}\chi_p(x)\,,$$

[27] C. C. J. Roothaan, *New Developments in Molecular Orbital Theory*, Rev. Mod. Phys. **23**, 69 (1951).

with $\mathcal{H}$ being the Fock operator defined by Eq. (2.25) in the HF method, or its counterpart in the Kohn–Sham approach. We shall refer to it as "Hamiltonian", since it is formally indeed a Hamiltonian of the reference system of independent particles, which produce the one-electron density equal to the correct one. Multiplying the above equation on the left by $\chi_q^*(x)$ and integrating in $x$, we get

$$\sum_p C_{\alpha p} \left[ \int \chi_q^*(x) \mathcal{H} \chi_p(x) dx - \varepsilon_\alpha \int \chi_q^*(x)\, \chi_p(x)\, dx \right] = 0 \tag{4.3}$$

That is a system of algebraic equations on the expansion coefficients $C_{\alpha p}$, or a generalized (i.e., with overlap) diagonalization problem. The overlap matrix elements

$$S_{qp} \equiv \int \chi_q^*(x)\, \chi_p(x)\, dx \tag{4.4}$$

need to be calculated only once (for basis functions fixed in advance). The matrix elements of the Hamiltonian

$$H_{qp} \equiv \int \chi_q^*(x) \left[ -\frac{\hbar^2}{2m} \nabla^2 + u(x) + \int \frac{\rho(x')dx'}{|\mathbf{r} - \mathbf{r}'|} + V_{\mathrm{XC}}(x) \right] \chi_p(x) dx$$

are actually dependent (via $\rho$ and $V_{\mathrm{XC}}$) on yet unknown coefficients $C_{\alpha p}$. Specifically, only the eigenvectors corresponding to $N$ lowest eigenvalues $\varepsilon_\alpha$ must be included in the construction of the ground-state density (similarly to the *aufbau pronciple* in the HF formalism), so the density is

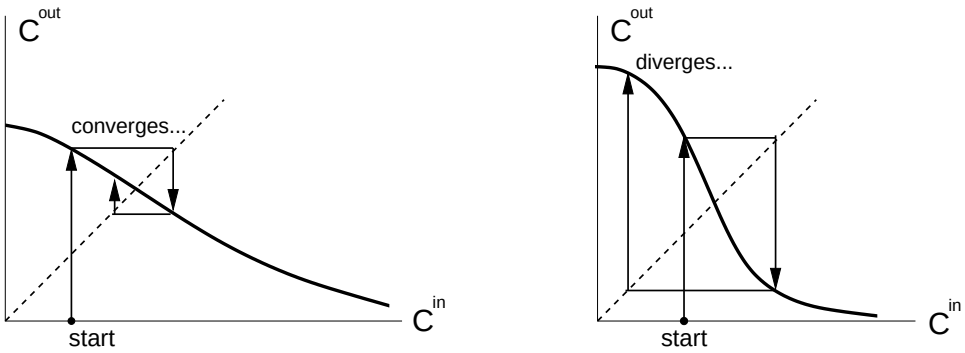
$$\rho(x) = \sum_{\alpha=1}^{N} \varphi_\alpha^*(x)\, \varphi_\alpha(x) = \sum_{pq} \underbrace{\left[ \sum_{\alpha=1}^{N} C_{\alpha q}^*\, C_{\alpha p} \right]}_{\equiv D_{pq},\ \text{density matrix}} \chi_q^*(x)\, \chi_p(x)\,. \tag{4.5}$$

The density can be either diagonal in spin variables, if basis functions for both spin components remain completely decoupled, or it may retain a general form, if spin components interact due to spin-orbit interaction and/or spin noin-collinearity effects being included. The internal dependency of $H_{pq}$ on $C_{\alpha q}$ via the density matrix prevents the solution of Eq. (4.3) in a single matrix manipulation. However, an iterative solution is possible: one fixes a trial set of $C_{\alpha q}$ in the density matrix and constructs the Hamiltonian, then (4.3) becomes a system of linear equations solvable by a single diagonalization. Among the $C_{\alpha q}$ ($q = 1, \ldots Q > N$) found by diagonalization one selects those corresponding to the $N$ lowest eigenvalues, uses them to update the density matrix and proceeds till everything converges.

In practical calculations, one would typically use a dampfing parameter $\beta < 1$,

$$C^{\mathrm{next}} = \beta C^{\mathrm{new}} + (1 - \beta) C^{\mathrm{old}}$$

or a more spphisticated mixing scheme, in order to prevent instability of expansion coefficients $C$ in the course of iterations, as is illustrated by the following figure.




A schematically shown dependency of the output solution vector on an input one. The self-consistency condition is shown by a dashed line. Increased dempfing (smaller $\beta$) scales the dependency towards the type of behaviour shown on the left.

In so doing, one has to solve in each iteration a system of linear equations. While $S_{qp}$ remains unchanged if the basis functions do not change (that is the case, at least, in solving a static electronic-structure problem with fixed positions of atoms), $H_{qp}$ has to be re-calculated and includes the following terms:

$$\begin{aligned} H_{qp} &= \int \chi_q^*(x) \left[ -\frac{\hbar^2}{2m} \nabla^2 \right] \chi_p(x) dx && (\leq \text{2-center integrals}) \\ &+ \int \chi_q^*(x)\, u(x)\, \chi_p(x) dx && (\leq \text{3-center integrals}) \\ &+ \sum_{q'p'} \underbrace{\left[ \sum_{\alpha=1}^{N} C_{\alpha q'}^* \, C_{\alpha p'} \right]}_{\text{density matrix}} \int \frac{\chi_q^*(x)\, \chi_{q'}^*(x')\, \chi_{p'}(x')\, \chi_p(x)}{|\mathbf{r} - \mathbf{r}'|} \, dx\, dx' && (\leq \text{4-center integrals}) \\ &+ \int \chi_q^*(x)\, V_{\mathrm{XC}}(x)\, \chi_q(x) dx \,. && (4.6) \end{aligned}$$

The term "$n$-center integral" refers to the situation when the basis functions are centered somewhere in space, usually at atom sites. This is often the case – but now always, as the site-independent basis functions (e.g., plane waves) may be also used. For site-centered bases, one usually takes into account the fact that the external potential $u(x)$ contains, along with other possible contributions, the Coulomb potential from atom nuclei, either screened by core electrons or not:

$$u(x) = \sum_{\mu} \frac{eZ_\mu}{|\mathbf{r} - \mathbf{R}_\mu|} + ...$$

Then, depending on whether both basis functions and the corresponding atom-centered component relate to the same center or not, the contributions of the form

$$\int \chi_q^{\alpha *}(\mathbf{r} - \mathbf{R}_\alpha) \, \frac{1}{|\mathbf{r} - \mathbf{R}_\mu|} \, \chi_p^{\beta}(\mathbf{r} - \mathbf{R}_\beta) \, d\mathbf{r}$$

may be one- , two- or three-center integral. Similarly, the Hartree contribution to the Hamiltonian (the 3d line of Eq. (4.6) may contain one- to four-center terms. The exchange-correlation part depends on $\rho(x)$ and may also give rise to up to four-center integrals, as

its particular form in the HF-formalism, the exchange term, would obviously do. The calculation of one-center integrals is usually trivial; two-center integrals may be simplified by an appropriate coordinate transformation (e.g., using elliptic coordinates); the three- and four-center integrals however may present a problem, and their evaluation deviates between different calculation methods.

After eigenvectors $C_{\alpha q}$ are finally converged and the density matrix $D_{pq}$ thus determined, the total energy can be recovered as follows from Eq. (3.14):

$$\begin{aligned} E_{\rm tot} &= -\frac{\hbar^2}{2m}\sum_{pq} D_{pq}\sum_i \int \chi_q^*(x)\nabla_i^2\chi_p(x)dx + \\ &+ \frac{e^2}{2}\sum_{pqrs} D_{pq}\,D_{rs}\int \chi_q^*(x)\,\chi_p(x)\,\chi_r^*(x')\,\chi_s(x')\frac{dx\,dx'}{|\mathbf{r}-\mathbf{r}'|} + \\ &+ e\sum_{pq} D_{pq}\int \chi_q^*(x)\,\chi_p(x)\,u(x)\,dx + E_{\rm XC}[\rho]\,, \end{aligned} \tag{4.7}$$

or, similarly to Eq. (3.17):

$$\begin{aligned} E_{\rm tot} &= \sum_{i=1}^{N}\varepsilon_i - \frac{e^2}{2}\sum_{pqrs} D_{ps}\,D_{rs}\int \chi_q^*(x)\,\chi_p(x)\,\chi_r^*(x')\,\chi_s(x')\,\frac{dx\,dx'}{|\mathbf{r}-\mathbf{r}'|} - \\ &- e\sum_{pq} D_{rq}\int \chi_q^*(x)\,\chi_p(x)\,V_{\rm XC}(x)\,dx + E_{\rm XC}[\rho]\,. \end{aligned} \tag{4.8}$$

So far, we haven't yet specified the form of basis functions $\chi$. As already mentioned, a quite common choise is some atom-centered functions localized in real space, but the functions localized in reciprocal space (plane waves) could be a convenient option, or some hybride bases. The basis of choice is often related to the nature of boundary conditions used. Thus, for a simulation of a periodic solid with Bloch – von Kármán boundary conditions the use of plane wave basis may seem a natural choice, whereas non-periodic systems like molecules or clusters may seem to be more naturally treated with localized atom-centered bases. However, in both these cases an alternative choice of basis is equally possible and in effect broadly used: molucules are simulated in a repeated "simulation box" with periodic boundary conditions thus imposed and plane waves used as basis set; for periodic solids, the (**k**-dependent Bloch sums can be constructed by lattice summation of localized functions, as:

$$\tilde{\chi}^{\alpha}_{p\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{N}}\sum_{\nu}\chi_p^{\alpha}(\mathbf{r}-\mathbf{u}_\alpha-\mathbf{R}_\nu)\exp[i\mathbf{k}(\mathbf{R}_\nu+\mathbf{u}_\alpha)]\,, \tag{4.9}$$

where $\mathbf{R}_\nu$ runs over lattice vectors and $\mathbf{u}_\alpha$ is the basis vector of atom $\alpha$ in the unit cell. Between the two extremities of plane waves and atom-centered functions as candidates for a basis set definition in a workable method, large variety of combined methods has been developed and is in use. The need for combined methods is due to the fact that none of both "pure" choices is fully satisfactory. The practical difficulty is that the electron density in real materials behave very differently near atomic cores (large fluctuations with short spatial period) and far from the cores (smooth distribution of density with possibly elevated values on the chemical bonds, i.e. away from atoms). In discussing a

compromise between computational efficiency and accuracy, let us sort out advantages and disadvantages of plane waves and atom-centered basis functions.

**Plane waves**
(+) Ultimately complete basis; systematic augmentation of accuracy is controlled by a single parameter (planewave cutoff);
(+) Easy analytical manipulation, e.g., when calculating matrix elements of different observables, derivatives of the total energy etc.
(−) Boundary conditions can be only periodic, therefore in the course of simulating finite fragments "in the box" the spurious interactions across the box boundary are built in, and their suppression may demand for a large box size.
(−) The number of plane waves necessary to describe fluctuations of all-electron charge density is usually beyond the reasonable computational resources. The use of pseudopotentials is a typical solution, but this is a certain approximation.
(−) For a given cutoff (i.e., the largest wavevector used in the planewave expansion), the size of basis grows very rapidly with the size of simulation cell, irrespectively on whether it contains extra atoms or not. This is a serious handicap in a simulation of large systems, but also of open systems, like surfaces or molecules.

**Atom-centered functions**
(+) Since all charge is physically delivered by one-electron functions centered on atoms, the basis size scales linearly with the number of atoms, irrespectively of empty space in the system. This is especially important for simulations of open systems.
(+) Boundary conditions can be either periodic or strictly "vacuum-like", with no spurious interaction between repeated fragments.
(−) The lack of systematics in gradually enhancing the completeness of basis; additional basis functions are added *ad hoc*, and no asymptotic completeness of the basis is guaranteed.
(−) Difficulties in calculating matrix elements. This is probably the most serious drawback that can be overcome by the following tricks:
– if possible, calculate in advance and store in tables for subsequent fast interpolation (good, e.g., for dynamical simulations);
– use efficient statistical scheme rather than straightforward integration;
– use localized basis functions which allow analytical (or otherwise easy) integration (Gaussian basis sets).

Among "analytical" localized basis functions, important are so-called Slater-type and Gaussian orbitals. The Slater-type orbitals (STO) imitate the asymptotics of one-electron wavefunctions, known for the hydrogen atom:

$$\chi_{\mathrm{STO}}(\mathbf{r}) \sim r^{n-1} e^{-\zeta r}\, Y_l^m(\theta,\phi)\,. \tag{4.10}$$

The STOs form a complete basis of solutions for the central-field problem, while the textbook solutions for the hydrogen atom are only complete if taken together with the eigenfunctions corresponding to the continuous energy spectrum. $n$ in STOs has the meaning of principal quantum numberm and $\zeta$ is "effective nuclear charge", but essentially it is an adjustable (free to vary) parameter. A peculiar feature of STOs is that, differently

from the hydrogen-atom solutions, they have no node structure, but a combination of several STOs makes it possible to imitate the nodes – and hence fluctuations of charge density near the atomic core. One speaks of "minimal basis", "double-zeta", "triple-zeta" etc., the latter combining several exponents with different $\zeta$ values. This terminology is also transferred to the classification of Gaussian bases discussed below. The number of Slater-type basis functions per atoms of different periods is shown in the following table:

| | Minimal basis | | Double-zeta | |
|---|---|---|---|---|
| H, He | $1s$ | [1] | $1s1s'$ | [2] |
| Li $\rightarrow$ Ne | $1s2s2p$ | [5] | $1s1s'2s2s'2p2p'$ | [10] |
| Na $\rightarrow$ Ar | $1s, \cdots, 3p3d$ | [14] | $\cdots$ | [28] |
| K $\rightarrow$ Kr | $1s, \cdots, 4s4p$ | [18] | $\cdots$ | [36] |
| Rb $\rightarrow$ Xe | $1s, \cdots, 5s5p$ | [27] | $1s1s', \cdots, 5p5p'$ | [54] |

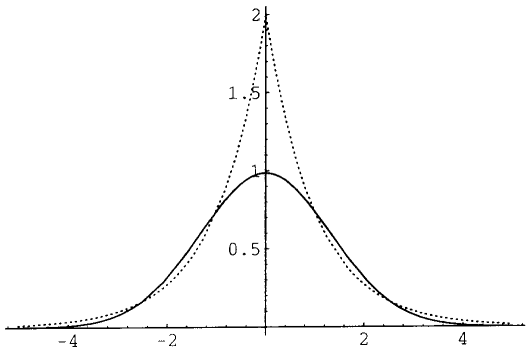
STO bases are known to be quite efficient in a sense that convergency of results is achieved fast with the augmentation of basis. However, the evaluation of integrals is as difficult as with any other (numerical) functions.

## 4.2 Gaussian basis sets

Gaussian-type orbitals (GTO) have been proposed by Boys in 1950[28]:

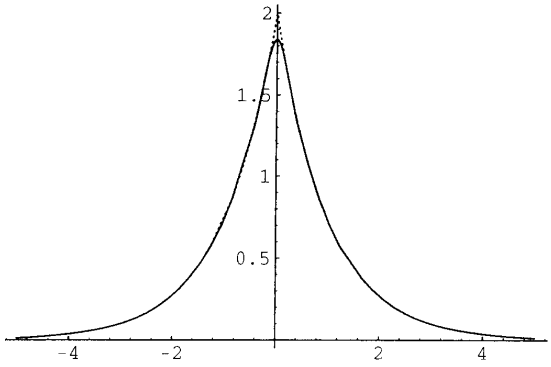
$$\chi_{\mathrm{GTO}} \sim r^{2(n-1)}\, e^{-\alpha r^2}\, Y_l^m(\theta, \phi)\,. \tag{4.11}$$

Their big advantage is the possibility to calculate the integrals (2- to 4-centered in the construction of Hamiltonian, as well as in many cases matrix elements of other operators of interest) analytically. In GTO, only functions with $l = n - 1$ are explicitly used, i.e. $1s$, $2p$, $3d$ etc. but not $2s$, $3p$. Thus, the node structure of one-electron wavefunctions is not included in any single GTO, but can be imitated by a linear combination of several basis functions. The drawback of gaussian bases is their slower convergency as compared to STO. This follows from the fact that individual GTOs are relatively poor representatives of true one-electron wavefunctions: GTOs have wrong asymptotics in the infinity (fall down too fast) and wrong behaviour near the nucleus. The following figure shows the "optimum" GTO obtained for the H$1s$ state by least-square fit, preserving the normalization.




[28] S. F. Boys, *Electronic wave functions. I. A general method of calculation for the stationary states of any molecular system*, Proc. Roy. Soc. (London) **A200**, 542 (1950).

The performance can be improved by using several GTO (3 – 4) to imitate a single STO. If keeping all them independent, the bottleneck is the storage of many integrals. A possible solution is to use *fixed* linear combinations of GTO in a calculation rather than individual GTOs; such combinations are referred to as *contracted* GTOs. One can fit contracted GTOs either to STOs, or to numerical solutions for single atoms or ions. The following figure shows again the H1$s$ wavefunction, now approximated by a contracted 4-GTO set.




Let us discuss now the evaluation of integrals. The GTOs may be rewritten in Cartesian form,

$$G_{ijk}(\mathbf{r},\alpha,\mathbf{A}) \equiv x_A^i\, y_A^j\, z_A^k\, \exp(-\alpha\, r_A^2) \tag{4.12}$$

where $\mathbf{A}$ is the origin and $\mathbf{r}_A \equiv \mathbf{r} - \mathbf{A}$; the Cartesian Gaussians can be factorized:

$$G_{ijk}(\mathbf{r},\alpha,\mathbf{A}) = G_i(x,\alpha,A_x)\, G_j(y,\alpha,A_y)\, G_k(z,\alpha,A_z) \tag{4.13}$$

with $G_i(x,\alpha,A_x) = x_A^i \exp(-\alpha x_A^2)$ etc. The differentiation of Gaussians (necessary e.g. for calculating forces) yields

$$\frac{\partial G_n}{\partial A_x} = -\frac{\partial G_n}{\partial x} = 2\alpha G_{n+1} - n\, G_{n-1}\,. \tag{4.14}$$

For higher-order differentiation, one can use the recursion

$$\frac{\partial^{\,q+1} G_n}{\partial A_x^{q+1}} = \left(\frac{\partial}{\partial A_x}\right)^q [\,2\alpha G_{n+1} - n\, G_{n-1}] = 2\alpha \frac{\partial^{\,q} G_{n+1}}{\partial A_x^q} - n \frac{\partial^{\,q} G_{n-1}}{\partial A_x^q}\,,$$

or more compact:

$$G_n^{(q+1)} = 2\alpha\, G_{n+1}^{(q)} - n\, G_{n-1}^{(q)}\,. \tag{4.15}$$

Of central importance for analytical evaluation of integrals is the *Gaussian product rule*, according to which a product of two Gaussians is again a Gaussian, centered at some point on the line connecting original centers. For $s$-Gaussian (without polynomial in $\mathbf{r}$) in one dimension the relation is straightforward:

$$\exp(-\alpha x_A^2)\, \exp(-\beta x_B^2) = \underbrace{\exp(-q\, Q_x^2)}_{\begin{array}{c}\text{constant}\\ \equiv K_{AB}\end{array}} \times \underbrace{\exp(-p\, x_P^2)}_{\begin{array}{c}\text{a new}\\ \text{Gaussian}\end{array}} \tag{4.16}$$

$$\text{with} \qquad x_P \equiv x - P_x\,, \quad p = \alpha + \beta\,, \quad q = \frac{\alpha\beta}{\alpha+\beta}\,, \quad pP_x = \alpha A_x + \beta B_x\,, \quad Q_x = A_x - B_x\,.$$

Along $y$ (or $z$)-direction, the product of two original exponents would yield $\exp[-(\alpha+\beta)y^2]$ with the same exponent prefactor $p$ as in the $x$-dependence, therefore the above product formula is valid in 3-dimensional case:

$$\exp(-\alpha \mathbf{r}_A^2)\, \exp(-\beta \mathbf{r}_B^2) = K_{AB}\, \exp(-p \mathbf{r}_P^2)\,.$$

In the product of two non-$s$-Gaussians that contain powers of $(\mathbf{r}-\mathbf{A})$ and $(\mathbf{r}-\mathbf{B})$ as well, the polynomial terms can be rearranged in another polynomial of $(\mathbf{r}-\mathbf{P})$. So a product of two arbitrary Gaussians yields

$$G_{ijk}(\mathbf{r},\alpha,\mathbf{A})\,G_{i'j'k'}(\mathbf{r},\beta,\mathbf{B}) = \sum_{(\alpha)} C^{(\alpha)} G_{(\alpha)}(\mathbf{r},p,\mathbf{P})$$

where $(\alpha)$ incorporates different possible combinations of powers of $x$, $y$ and $z$, up to the maximal power $i+j+k+i'+j'+k'$. After such decomposition, the overlap matrix elements $S_{qp}$ (4.4) reduce to the sum of integrals

$$S_{qp} = \sum_{(\alpha)} C^{(\alpha)} \int d\,\Omega Y_l^m(\theta,\phi) \int_0^\infty r^{n(\alpha)} e^{-\gamma r^2} r^2 dr$$

that are taken using (with $\gamma > 0$, $n = 0, 1, \ldots$):

$$\int_0^\infty x^{2n} e^{-\gamma x^2} dx = \frac{(2n-1)!!}{2(2\gamma)^n}\sqrt{\frac{\pi}{\gamma}}\,; \qquad \int_0^\infty x^{2n+1} e^{-\gamma x^2} dx = \frac{n!}{2\gamma^{n+1}}\,.$$

Analyzing different terms in Eq. (4.6), one can sort them out into

• multipole moments, $\langle G_A|\, x_C^e\, y_C^f\, z_C^g\, |G_B\rangle$ with $G_A = x_A^i\, y_A^j\, z_A^k \exp(-\alpha r_A^2)$ etc. and integration done over coordinates of one electron. Such integrals describe, for example, the effect of the external field;

• integrals $\langle G_A|\, e^{ikCxC}\, |G_B\rangle$ that appear, e.g., in the calculation of exchange-correlation term, when the Fourier transformation of a general-form density is used. These two types of integrals can be analytically taken by parts and are available in closed form;

• kinetic energy. Since differential operators acting on Cartesian Gaussians are separable,

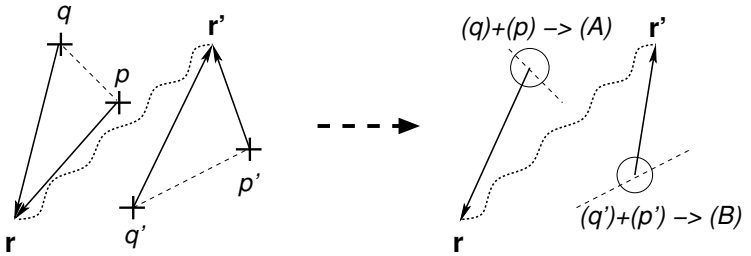
$$\langle G_A|\left(\frac{\partial}{\partial x}\right)^q\left(\frac{\partial}{\partial y}\right)^r\left(\frac{\partial}{\partial z}\right)^s|G_B\rangle = \langle G_i|\frac{\partial^q}{\partial x^q}|G_j\rangle\langle G_k|\frac{\partial^q}{\partial x^q}|G_l\rangle\langle G_m|\frac{\partial^q}{\partial x^q}|G_n\rangle\,.$$

By the virtue of recursion (4.15), these matrix elements are related to overlap integrals between GTOs of reduced and increased order, up to $\pm 2$.

More involved is calculation of (up to) four-center Coulomb integrals,

$$\int \chi_q^*(\mathbf{r})\,\chi_p(\mathbf{r})\,\chi_{q'}^*(\mathbf{r}')\,\chi_{p'}(\mathbf{r}')\,\frac{d\mathbf{r}\,d\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}\,.$$

For simplicity, we discuss in the following the case of all four GTOs being of $s$-type, i.e. without polynomial prefactor. For general-form Gaussians analytical integration is also possible but much more lengthy.



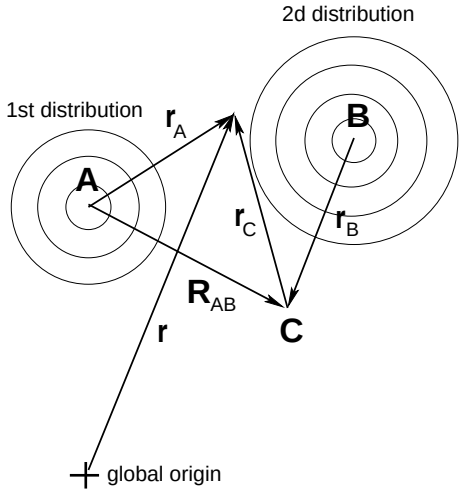

The figure above explains the geometry of our problem: the origins of four basis functions, if all different, are at $q$, $p$, $q'$ and $p'$; two spatial variables $\mathbf{r}$ and $\mathbf{r}'$ are involved.

The Gaussian product rule allows to reduce this to, at most, 2-center problem, with new centers **A** and **B** that can be straightforwardly found. Now, the integral to be evaluated is formally identical to a Coulomb interaction between two Gaussian distributions of charge:

$$V_{\mathbf{AB}} = \int d\mathbf{r}_1 \int d\mathbf{r}_2 \frac{\rho_{\mathbf{A}}(\mathbf{r}_A)\,\rho_{\mathbf{B}}(\mathbf{r}_B)}{|\mathbf{r}_A - \mathbf{r}_B|}\,. \tag{4.17}$$

In order to evaluate this, we specify first the electrostatic potential (at arbitrary point **C**) of such distribution centered at **A**, then the electrostatic energy of the second charge distribution (centered at **B**) in this potential. For the following, we assume the charge distribution around **A** to be normalized to unity,

$$\begin{aligned} \rho_{\mathbf{A}}(\mathbf{r}_A) &= \left(\frac{\alpha}{\pi}\right)^{3/2} e^{-\alpha r_A^2}\,; \\ V_{\mathbf{A}}(\mathbf{C}) &= \int \frac{\rho_{\mathbf{A}}(\mathbf{r}-\mathbf{A})}{|\mathbf{r}-\mathbf{C}|} d\mathbf{r}\,. \end{aligned}$$



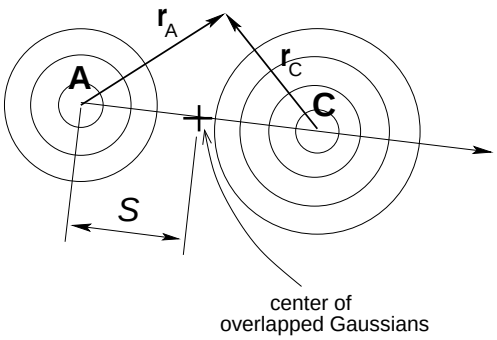

For $1/|\mathbf{r}-\mathbf{C}| \equiv 1/\mathbf{r}_C$ we make a formal substitution using

$$\int_{-\infty}^{\infty} e^{-R^2 t^2} dt = \frac{\sqrt{\pi}}{R}\,;$$

$$\frac{1}{|\mathbf{r}-\mathbf{C}|} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-r_C^2\, t^2} dt\,.$$

Applying the Gaussian product rule (4.16) to $\exp(-\alpha R_A^2)\exp(-r_C^2 t^2)$, we transform

$$\begin{aligned} V_{\mathbf{A}}(\mathbf{C}) &= \frac{\alpha^{3/2}}{\pi^2} \int d\mathbf{r}\, e^{-\alpha r_A^2} \left[\int_{-\infty}^{\infty} e^{-r_C^2\, t^2} dt\right] = \\ &= \frac{\alpha^{3/2}}{\pi^2} \int_{-\infty}^{\infty} dt\, e^{-\frac{\alpha t^2}{\alpha+t^2} R_{AC}^2} \left[\int d\mathbf{r}\, e^{-(\alpha+t^2)(\mathbf{r}-\mathbf{S})^2}\right]. \end{aligned}$$




Note that $t$ is a dimensionless integration variable, and **S** marks the position of the resulting (overlapped) Gaussian along the **A**–**C** line (different for each $t$), as shown on the left figure:

$$\mathbf{S} = \frac{t^2\,\mathbf{R}_{AC}}{\alpha + t^2}\,,$$

Thus the origin of integration in **r** is simply displaced to **S**, with the result unchanged:

$$\int d\mathbf{r}\, e^{-(\alpha+t^2)\,r^2} = \frac{\pi^{3/2}}{(\alpha+t^2)^{3/2}}\,,$$

hence

$$V_{\mathbf{A}}(\mathbf{C}) = \frac{\alpha^{3/2}}{\sqrt{\pi}} \int\limits_{-\infty}^{\infty} \frac{\exp(-\frac{\alpha\, t^2}{\alpha+t^2}\, R_{AC}^2)}{(\alpha+t^2)^{3/2}}\, dt\,.$$

A variable substitution

$$\frac{t^2}{\alpha+t^2} \to u^2\,; \quad dt = \frac{(\alpha+t^2)^{3/2}}{\alpha}\, du$$

results in

$$V_{\mathbf{A}}(\mathbf{C}) = \sqrt{\frac{4\alpha}{\pi}} \int\limits_0^1 e^{-\alpha R_{AC}^2 u^2} du\,. \tag{4.18}$$

This integral can be expressed via error function:

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int\limits_0^x e^{-t^2} dt\,,$$

hence

$$\int\limits_0^1 e^{-Ax^2} dx = \frac{1}{\sqrt{A}} \int\limits_0^{\sqrt{A}} e^{-t^2} dt = \sqrt{\frac{\pi}{4A}} \mathrm{erf}(\sqrt{A})$$

and

$$V_{\mathbf{A}}(\mathbf{C}) = \frac{1}{R_{AC}} \mathrm{erf}(R_{AC}\sqrt{\alpha})\,.$$

Now we are ready to calculate the Coulomb integral (4.17). Assuming that the charge distribution centered at $\mathbf{B}$ has an exponent $\beta$ and is normalized to unity, $\rho_{\mathbf{B}}(\mathbf{r}_B) = (\beta/\pi)^{3/2} e^{(-\beta r_B^2)}$, similarly to how it was for $\rho_{\mathbf{A}}(\mathbf{r}_A)$, (4.17) becomes:

$$V_{\mathbf{AB}} = \int d\mathbf{r}\, V_{\mathbf{A}}(\mathbf{r}) \left(\frac{\beta}{\pi}\right)^{3/2} e^{-\beta(\mathbf{r}-\mathbf{B})} = \sqrt{\frac{4\alpha}{\pi}} \left(\frac{\beta}{\pi}\right)^{3/2} \int d\mathbf{r} \int\limits_0^1 du\, e^{-\alpha r_A^2 u^2} e^{-\beta r_B^2}\,.$$

Once more applying the product rule yields:

$$\exp(-\alpha r_A^2 u^2) \exp(-\beta r_B^2) = \exp\Big[-(\alpha u^2+\beta) r_S^2\Big] \exp\left(-\frac{\alpha u^2 \beta}{\alpha u^2+\beta} R_{AB}^2\right)$$

where $\mathbf{S}$, from which $r_S$ is measured, is the center position of overlapped Gaussians somewhere on the $\mathbf{A}$–$\mathbf{B}$ line.

$$\begin{aligned} V_{\mathbf{AB}} &= \sqrt{\frac{4\alpha}{\pi}} \left(\frac{\beta}{\pi}\right)^{3/2} \int\limits_0^1 du\, \exp\left(-\frac{\alpha u^2 \beta}{\alpha u^2+\beta}\, R_{\mathbf{AB}}^2\right) \underbrace{\int d\mathbf{r} \exp[-(\alpha u^2+\beta) r_S^2]}_{\left(\frac{\pi}{\alpha u^2+\beta}\right)^{3/2}} \\ &= \sqrt{\frac{4\alpha\beta}{\pi}} \int\limits_0^1 \frac{\beta}{(\alpha u^2+\beta)^{3/2}}\, \exp\left(-\frac{\alpha u^2 \beta}{\alpha u^2+\beta} R_{AB}^2\right) du\,. \end{aligned}$$

Another variable substitution

$$\frac{\alpha+\beta}{\alpha u^2+\beta}\,u^2 \to t^2\,; \qquad (\alpha+\beta)^{1/2}\frac{\beta du}{(\alpha u^2+\beta)^{3/2}} \to dt$$

leads to an expression completely symmetric in (**A**,$\alpha$) and (**B**,$\beta$), as it of course should be, with $\gamma \equiv \alpha\beta/(\alpha+\beta)$:

$$V_{\mathbf{AB}} = \sqrt{\frac{4\alpha\beta}{\pi}}\,\frac{1}{\sqrt{\alpha+\beta}}\int_0^1 \exp\left(-\frac{\alpha\beta}{\alpha+\beta}R_{AB}^2 t^2\right)dt = \sqrt{\frac{4\gamma}{\pi}}\int_0^1 e^{-\gamma R_{AB}^2 t^2}dt\,. \tag{4.19}$$

The above derivation helps to calculate the values of (up to) 4-center Coulomb integrals, provided they do not contain polynomials (i.e., only $s$-Gaussians enter). For GTO of general form, closed analytical expressions, albeit of much more complicated form, have also been obtained. For a review, one can consult, e.g., an article by Helgaker and Taylor[29]

Whereas in HF calculations all terms in the Hamiltonian reduce to 2-, 3- or 4-center integrals, in a DFT formalism the exchange-correlation potential $V_{\rm XC}$ is present that is a nonlinear function (or, beyond the LDA, a functional) of density. Moreover, the Hartree potential, although reducible to Coulomb interaction between individual Gaussians, is expensive (in terms of computation time and storage) due to the necessity to re-calculate and store multicenter integrals. It is more efficient to fit the total resulting density to a superposition of support functions, probably the same Gaussians (typically taken to a higher order than for the representation of basis functions). Therefore, it may be both advantageous and unavoidable to deal with potential and charge density of spatially unconstrained, i.e., general form. Practically used solutions are:

• expand everything in Gaussians centered at the atoms, or in the interstitial (i.e. at chemical bonds).

• single out slowly varying interstitial density (by subtracting atom-like contributions) and represent it on a discrete mesh (e.g., by splines, or by Fourier transformation) with the aim to solve the Poisson equation easily.

A typical structure of the latter decomposition of density could be:

$$\begin{aligned}\rho(\mathbf{r}) &= \rho_{\rm loc}(\mathbf{r}) + \rho_{\text{non-loc}}(\mathbf{r}) + \rho_{\rm nucl}(\mathbf{r}) \\ &= [\rho_{\rm loc}(\mathbf{r}) + \rho_{\rm nucl}(\mathbf{r}) + \xi(\mathbf{r})] + [\rho_{\text{non-loc}}(\mathbf{r}) - \xi(\mathbf{r})]\,,\end{aligned}$$

where $\rho_{\rm nucl}(\mathbf{r})$ is a well localized near nuclei contribution of deep core states, $\rho_{\rm loc}(\mathbf{r})$ is a valence charge density near atoms, that can be typically expanded over spherical harmonics inside a sphere of appropriate radius around each site. $\rho_{\text{non-loc}}(\mathbf{r})$ is a smoothly varying charge density in the interstitial region. The separation density $\xi(\mathbf{r})$ is chosen in such way that each square bracket remains neutral and as smooth as possible. For $\xi(\mathbf{r})$, a dual representation is used: by spherical harmonic expansion around centers in the first bracket and e.g. by plane wave expansion in the second bracket. Then the contribution to the potential follows by multipole expansion from the first bracket and by Fourier transformation from the second bracket.

[29] T. Helgaker and P. R. Taylor, "Caussian Basis Sets and Molecular Integrals", in: "Modern Electronic Structure Theory", Ed. D. R. Yarkony, World Scientific, Singapore 1995.

## 4.3 Planewaves and pseudopotentials

Earlier, we already dealt with planewave basis functions in order to calculate different integrals related to a system of non-interacting electrons. When dealing with planewave bases in crystals, we'll write them down as

$$\chi_{\mathbf{k}}^{\mathbf{G}} = \frac{1}{\sqrt{\Omega}} e^{i(\mathbf{k}+\mathbf{G})\mathbf{r}} .$$

Here, $\mathbf{k}$ stands for a point in the Brillouin zone labeling a certain irreducible representation of the translation group. In other words, this is an "artefact" of the Born – von Kármán periodic boundary conditions. Infinite number of cells in a crystal maps onto infinite number $\mathbf{k}$ points, effectively a continuous distribution, over the Brillouin zone. As we scan the Brilluoin zone, we study the energy dispersion. As the repeated unit cell gets larger (in one dimension, as for slabs, or in all dimensions, as for a "molecule in a box"), the Brilluoin zone shrinks correspondingly, and the dispersion effectively disappears – one gets instead a sequence of discrete energy levels. In this case, it may be sufficient to use just one $\mathbf{k}$-point, $\mathbf{k} = 0$.

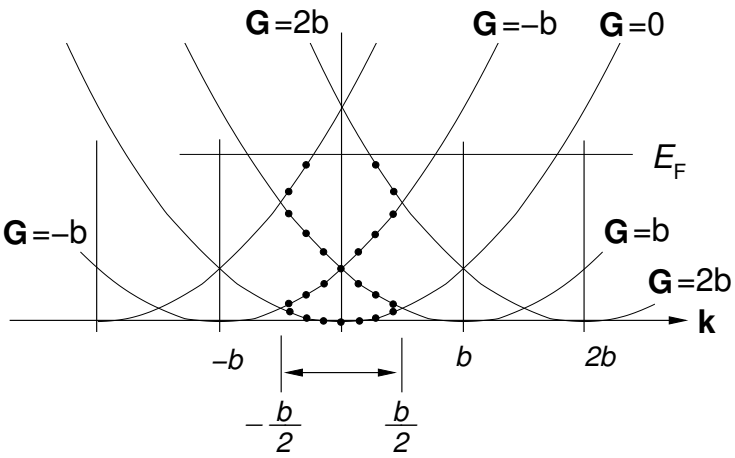
For each $\mathbf{k}$ (the results for different $\mathbf{k}$ do not mix other than via the construction of density), $\mathbf{G}$ runs over plane waves actually used as basis functions. $\mathbf{G}$ labels the sites of the reciprocal lattice, usually within some cutoff, $|\mathbf{k} + \mathbf{G}| \leq \mathbf{G}_{\max}$. (The actual set of $\mathbf{G}$ can be therefore slightly different for different $\mathbf{k}$). Further on, $\mathbf{G}$ will number the matrix elements in the planewave representation, e.g.,

$$S_{\mathbf{G}\mathbf{G}'} = \frac{1}{\Omega} \int_{\Omega} d\mathbf{r}\, e^{i(\mathbf{G}'-\mathbf{G})\,\mathbf{r}} = \delta_{\mathbf{G}\mathbf{G}'} .$$

The big advantage of the planewave basis is the easiness with which some matrix elements can be calculated. The momentum operator and hence the kinetic energy are diagonal in the planewave basis, yielding

$$T_{\mathbf{G}\mathbf{G}'}^{\mathbf{k}} = \frac{\hbar^2}{2m} (\mathbf{G} + \mathbf{k})^2 \, \delta_{\mathbf{G}\mathbf{G}'} .$$

This solves analytically an important (for numerical control of accuracy of a computational scheme in question) case of empty lattice. The figure on the right shows several lowest bands (labeled by $\mathbf{G}$) for one-dimensional empty lattice. The occupied eigenstates (for several $\mathbf{k}$ within the first Brillouin zone) are indicated by dots.



The presence of non-constant potential disturbs this band structure, and eigenfunctions are not "pure" planewaves anymore, but their linear combinations. In order to calculate them by diagonalization, one needs the matrix elements of the Hamiltonian that read

$$\begin{aligned} H^{\mathbf{k}}_{\mathbf{G}\mathbf{G}'} &= \frac{\hbar^2}{2m}\left(\mathbf{k}+\mathbf{G}\right)^2 \delta_{\mathbf{G}\mathbf{G}'} + \frac{1}{\Omega}\int e^{i(\mathbf{G}'-\mathbf{G})\mathbf{r}} u(\mathbf{r}) d\mathbf{r} + \\ &+ \frac{e^2}{\Omega^2}\sum_{\mathbf{G}_1\mathbf{G}_2} D^{\mathbf{k}}_{\mathbf{G}_1\mathbf{G}_2} \int\int \frac{\exp[i(\mathbf{G}'-\mathbf{G})\mathbf{r}]\,\exp[i(\mathbf{G}_2-\mathbf{G}_1)\mathbf{r}']}{|\mathbf{r}-\mathbf{r}'|}\, d\mathbf{r}\, d\mathbf{r}' + \\ &+ \frac{1}{\Omega}\int e^{i(\mathbf{G}'-\mathbf{G})\mathbf{r}}\, V_{\mathrm{XC}}(\mathbf{r}) d\mathbf{r}\,, \end{aligned} \tag{4.20}$$

where $D^{\mathbf{k}}_{\mathbf{G}_1\mathbf{G}_2}$ is the density matrix as introduced in Eq. (4.5).

The Hartree term in the second line reduces to the Fourier transform of the Coulomb interaction,

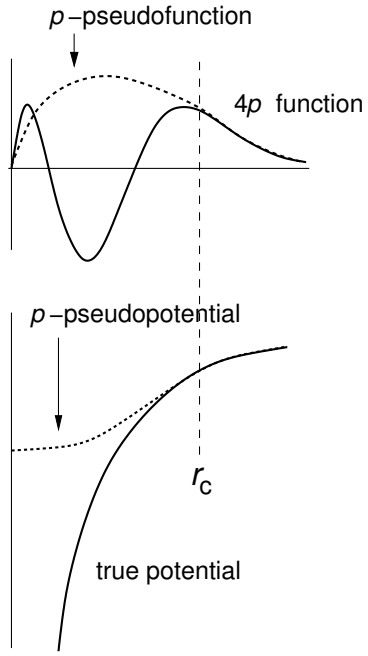
$$\underbrace{\int d\mathbf{r} e^{i(\mathbf{G}'-\mathbf{G}+\mathbf{G}_2-\mathbf{G}_1)\mathbf{r}}}_{\Omega\,\delta(\mathbf{G}'-\mathbf{G}+\mathbf{G}_2-\mathbf{G}_1)} \int d(\mathbf{r}'-\mathbf{r}) \frac{\exp[i(\mathbf{G}_2-\mathbf{G}_1)(\mathbf{r}'-\mathbf{r})]}{|\mathbf{r}'-\mathbf{r}|}\,,$$

whereas external potential and $V_{\mathrm{XC}}(\mathbf{r})$ contribute via their Fourier transforms:

$$\langle \mathbf{G}|V_{\mathrm{XC}}|\mathbf{G}'\rangle = \sum_{\mathbf{Q}} V_{\mathrm{XC}}(\mathbf{Q})\,\delta(\mathbf{Q}+\mathbf{G}'-\mathbf{G})\,.$$

Only these last terms are actually problematic in a planewave calculation, because they may be strongly localized in real space, so that their Fourier expansions do not converge up to a prohibitively large $\mathbf{G}$-cutoff. This can be understood as a consequence of the fact that these terms provide a strong perturbation to the empty lattice case and hence mix the basis states far separated in energy. Or, a deep potential in the vicinity of atomic cores results in rapidly fluctuating density and hence high $\mathbf{G}$ cutoff for its proper description. Such problems did not arise in tight-binding-type schemes, where the basis functions could be tailored to account for such fluctuations.

The only possibility to save a purely planewave basis for practical calculations is to discard core states, considering them as "frozen". In other words, their spatial form can be obtained from a single-atom calculation and further on to be believed to remain the same in a molecule or a solid. The justification for such treatment could be, that the core states do not participate in chemical bonding. The exclusion of core states from the direct consideration means that we have to cope in the following not with the effect of bare nuclei potentials, but rather with the potential screened by core electrons. This potential is more smooth and shallow than the "true" one. This simplification gives rise to a family of pseudopotential methods, which are contrasted to "all-electron methods" (i.e., those where all electrons, valence and core, are treated on equal footing). gives rise to a family of pseudopotential methods, which are contrasted to "all-electron methods" (i.e., those where all electrons, valence and core, are treated on equal footing).




The most straightforward procedure of screening the "true" potential with fixed core density is not practically useful. For one thing, the Coulomb field of a not fully compensated bare charge remains singular at the nuclei. Moreover, a true valence wavefunction must have nodes in the intraatomic region for ensuring its orthogonality to the core states. The description of these nodes by plane waves needs high cutoffs. In reality, one works with smooth node-free pseudofunctions generated in a shallow pseudopotential. In historical perspective, pseudopotentials have often been "model" ones, constructed *ad hoc* and tunable with just several parameters. Nowadays, pseudopotentials in use are usually of *ab initio* type. They are "cooked" (with the use of certain approximations and criteria, to be discussed further on) from "true" (all-electron; in a sense that all electrons participate in the solution of the Kohn–Sham equations) solutions for free atoms or ions.

The construction of a pseudopotential typically starts with the choice of an appropriate reference configuration (e.g., Fe$3d^7 4s^1$) and "pseudoization radii" $r_{\rm C}$ (see the figure) that can be different for different $l$-channels. As a rule, the following conditions are imposed:
• the pseudofunction must have no nodes (in order to avoid wiggles that would demand for higher cutoff);
• the pseudofunction matches the all-electron one beyond the cutoff radius;
• norm conservation, meaning that the charge contained within the pseudoization radius is the same for the pseudofunction and the all-electron one. (This condition however can be relaxed; deviations from this rule give rise to ultrasoft pseudopotentials, to be specially discussed below);
• the eigenvalues corresponding to pseudofunctions must be equal to those of the all-electron solution – at least for the reference configuration, and in the ideal case over a range of reasonable configurations, meaning the pseudopotential is *transferable*.

The pseudopotential approach is related to the orthogonalized planewave method (OPW), in which the basis set consists of plane waves, orthogonalized to lower-lying core states $\chi^\alpha_{\rm core}$:

$$\chi_{\mathbf{k}+\mathbf{G}}(\mathbf{r}) = e^{i(\mathbf{k}+\mathbf{G})\mathbf{r}} - \sum_\alpha \sum_{\rm core} \langle \chi^\alpha_{\rm core} | e^{i(\mathbf{k}+\mathbf{G})\mathbf{r}} \rangle \, \chi^\alpha_{\rm core}(\mathbf{r}) \,. \tag{4.21}$$

We now consider the Phillips–Kleinmann construction of pseudopotential which, if formally appearing in the Kohn–Sham equations, leads, as a solution, to the pseudofunction $\varphi^{\rm PS}_{\rm c}$. Let $\mathcal{H}$ be the Hamiltonian whose core and valence wavefunctions are $\varphi_{\rm c}$ and $\varphi_{\rm v}$, correspondingly:

$$\begin{aligned} \mathcal{H}\varphi^\alpha_{\rm c} &= \varepsilon^\alpha_{\rm c}\varphi^\alpha_{\rm c} \,; \\ \mathcal{H}\varphi^{\mathbf{k}}_{\rm v} &= \varepsilon^{\mathbf{k}}_{\rm v}\varphi^{\mathbf{k}}_{\rm v} \,. \end{aligned}$$

In analogy with Eq. (4.21), we construct pseudofunctions from valence and core states (we remove wiggles in valence states by cancelling them by those in core states):

$$\varphi_{\rm v}^{\rm PS} = \varphi_{\rm v} + \sum_{\alpha\,{\rm c}} a_{\rm vc}^{\alpha}\,\varphi_{\rm c}^{\alpha}\,. \tag{4.22}$$

By multiplying with $\varphi_{\rm c'}$ and integrating, one gets the expansion coefficients:

$$\langle\varphi_{\rm c'}^{\beta}|\varphi_{\rm v}^{\rm PS}\rangle \;=\; \underbrace{\langle\varphi_{\rm c'}^{\beta}|\varphi_{\rm v}\rangle}_{\substack{=0,\text{ since core and}\\ \text{valence states}\\ \text{correspond to the}\\ \text{same Hamiltonian}}} + \sum_{\alpha{\rm c}} a_{\rm vc}^{\alpha}\,\underbrace{\langle\varphi_{\rm c'}^{\beta}|\varphi_{\rm c}^{\alpha}\rangle}_{\substack{=\delta_{\alpha\beta}\delta_{\rm cc'};\\ \text{core states at different}\\ \text{centers do not overlap}}}\;;$$

$$\text{whence}\quad a_{\rm vc}^{\alpha} \;=\; \langle\varphi_{\rm c'}^{\alpha}|\varphi_{\rm v}^{\rm PS}\rangle\,. \tag{4.23}$$

Now we act on $\varphi_{\rm v}^{\rm PS}$ of Eq. (4.22) by $\mathcal{H}$:

$$\begin{aligned}
\mathcal{H}\varphi_{\rm v}^{\rm PS} &= \underbrace{\mathcal{H}\varphi_{\rm v}}_{\varepsilon_{\rm v}\varphi_{\rm v}} + \sum_{\alpha\,{\rm c}} a_{\rm vc}^{\alpha}\,(\mathcal{H}\varphi_{\rm c}^{\alpha}) \\
&= \varepsilon_{\rm v}\left[\varphi_{\rm v}^{\rm PS} - \sum_{\alpha\,{\rm c}} a_{\rm vc}^{\alpha}\varphi_{\rm c}^{\alpha}\right] + \sum_{\alpha\,{\rm c}} a_{\rm vc}^{\alpha}\,\varepsilon_{\rm c}^{\alpha}\varphi_{\rm c}^{\alpha} \\
&= \varepsilon_{\rm v}\varphi_{\rm v}^{\rm PS} + \sum_{\alpha\,{\rm c}} a_{\rm vc}^{\alpha}(\varepsilon_{\rm c}^{\alpha} - \varepsilon_{\rm v})\varphi_{\rm c}^{\alpha}\,.
\end{aligned}$$

$$\Big[\mathcal{H} + \sum_{\alpha\,{\rm c}}(\varepsilon_{\rm c}^{\alpha} - \varepsilon_{\rm v})\,\underbrace{|\varphi_{\rm c}^{\alpha}\rangle\langle\varphi_{\rm c}^{\alpha}|}_{\substack{\text{projection}\\ \text{operator}}}\Big]|\varphi_{\rm v}^{\rm PS}\rangle = \varepsilon_{\rm v}\,|\varphi_{\rm v}^{\rm PS}\rangle\,. \tag{4.24}$$

It should be emphasized here that $\mathcal{H}$ in Eq. (4.24) is the conventional Kohn–Sham Hamiltonian, that includes contributions from ion cores, all electrons in the Hartree term and exchange-correlation potential depending on the total density. The transformation leading to (4.24) is merely a technical regrouping of terms. The additional term in the Hamiltonian yielding the pseudofunction, as compared to the Hamiltonian $\mathcal{H}$ yielding the "true" wavefunction, can be added to the conventional potential $V(\mathbf{r})$:

$$V(\mathbf{r}) \equiv u(\mathbf{r}) + \frac{e^2}{2}\int\frac{\rho(\mathbf{r})\,\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}\,d\mathbf{r}' + V_{\rm XC}(\mathbf{r})$$

and gives rise to the **Phillips – Kleinman pseudopotential**:

$$V^{\rm PK} = V(\mathbf{r}) + \sum_{\alpha\,{\rm c}}(\varepsilon_{\rm v}^{\alpha} - \varepsilon_{\rm c})|\varphi_{\rm c}^{\alpha}\rangle\langle\varphi_{\rm c}^{\alpha}|\,. \tag{4.25}$$

It is *energy-dependent* (via $\varepsilon_{\rm v}$); outside the core region it coincides with $V(\mathbf{r})$. Since $\varepsilon_{\rm v} > \varepsilon_{\rm v}$, the additional term is **repulsive** and hence the pseudopotential is softer than the original potential. But the valence eigenvalues are by construction identical with those of the corresponding all-electron problem.

Actually, the construction of pseudopotential in reality is seldom done this way – because the main idea is to eliminate core wavefunctions from calculation, rather than calculate projections of pseudofunctions on them. In practice, one typically starts from a smooth pseudofunction – there is a certain freedom in choosing it – and then *inverts* the radial Schrödinger equation to get the corresponding pseudopotential:

$$V_l^{\text{pseudo}}{}_{\text{screened}}(\mathbf{r}) = \varepsilon_l - \frac{l(l+1)}{r^2} + \frac{1}{2r R_l^{\text{PS}}(r)} \frac{d^2}{dr^2}\left[r\, R_l^{\text{PS}}(r)\right] . \tag{4.26}$$

This gives a *screened* pseudopotential, which incorporates the Hartree and exchange-correlation terms related to the pseudocharge only, i.e. to the charge distribution in a pseudo-wavefunction. When using the pseudopotential in the following, we would need to re-calculate these terms according to how the pseudo-WF evolves; so we subtract the corresponding terms and get an *unscreened* pseudopotential,

$$V_l^{\text{pseudo}}{}_{\text{unscreened}}(\mathbf{r}) = V_l^{\text{pseudo}}{}_{\text{screened}}(\mathbf{r}) - V_l^{\text{pseudo-H}}(\mathbf{r}) - V_l^{\text{pseudo-XC}}(\mathbf{r}) .$$

A further progress in the use of pseudopotential methods was achieved along the line of implementing the following priorities:
• the PP should be as soft as possible;
• it should be as transferable as possible (i.e., the PP generated for a given atomic configuration should reproduce others accurately), that is essential for applications dealing with simulations of chemical bonding;
• The pseudo-charge density should reproduce the "true" valence charge density as accurately as possible.

These goals are essentially conflicting, for the reasons to be discussed below. In the search for a compromise, the concept of **norm conservation** remained for a long time generally adopted. The condition of norm-conservation reads

$$\int_0^{r_{\text{c}}} r^2\, dr \left|\varphi_{\text{v}}^{\text{PS}}(\mathbf{r})\right|^2 = \int_0^{r_{\text{c}}} r^2\, dr \left|\varphi_{\text{v}}^{\text{all-electron}}(\mathbf{r})\right|^2 . \tag{4.27}$$

that must hold for all valence pseudofunctions. Since all-electron (single-atom) wavefunctions and eigenvalues $\varepsilon_{\text{v}}$ are different for different angular momentum values $l$, so would be cutoff radii and pseudopotentials. Such ($l$-dependent) pseudopotentials are called *semi-local*.

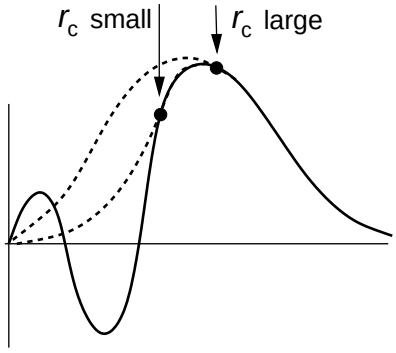
It is good for the transferability of pseudopotential if the slope of the pseudofunction equals that of the true wavefunction at the pseudoization radius. Taken together with the norm-conservation, this formulates the condition on the logarithmic derivatives:

$$\frac{1}{\varphi^{\text{PS}}(r_{\text{c}}, E)} \frac{d\,\varphi^{\text{PS}}(r_{\text{c}}, E)}{d\,r}\bigg|_{r=r_{\text{c}}} = \frac{1}{\varphi^{\text{all-el.}}(r_{\text{c}}, E)} \frac{d\,\varphi^{\text{all-el.}}(r_{\text{c}}, E)}{d\,r}\bigg|_{r=r_{\text{c}}} . \tag{4.28}$$

The larger the energy "window" within which the logarithmic derivatives match approximately, the better the transferability.

In 1982, Bachelet, Hamann and Schluter proposed a scheme to construct transferable semi-local pseudopotentials. The scheme consists of several steps:

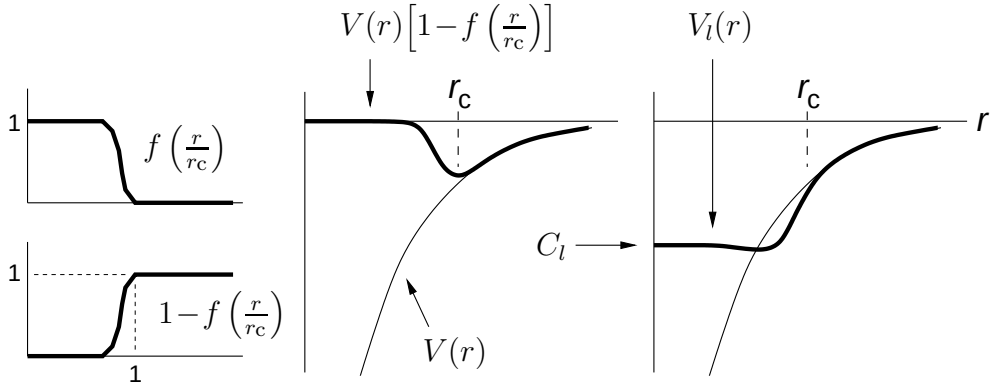
(1) Choose reference configuration, calculate all-electron eigenvalues and eigenfunctions of a free atom.
(2) Choose pseudoization radii for each $l$-channel. The value of choice is slightly beyond, or inside, the outmost maxima of the all-electron wavefunction. Larger $r_c$ results in softer, but less transferable, pseudopotential, smaller $r_c$ in hard but more transferable one.




(3) Construct "first-step pseudopotential" $V_l(r$ by cutting off the singularity in the all-atomic potential:

$$V_l(r) = V(r)\left[1 - f\left(\frac{r}{r_c}\right)\right] + C_l\, f\left(\frac{r}{r_c}\right) .$$

Here, $f(x)$ is essentially an arbitrary monotonic function such that $f(0) = 1$, $f(\infty) = 0$, cutting off rapidly about $x = 1$.




(4) Norm conservation is imposed by adding a correction to the core region, e.g.

$$\varphi^{\rm PS}(r) = \gamma_l \left[\varphi^{(1)}(r) + \delta_l\, g_l(r)\right] ,$$

where $\delta_l$ is a variable parameter that enforces norm conservation, $\gamma_l$ is a scaling factor $\gamma_l = \varphi^{\text{all-el.}}/\varphi^{(1)}$ (in the outer region $r > r_c$ these functions may differ at most by a constant); $g_l(r)$ is an arbitrary, in principle, function, that must however have correct behaviour ($\sim r^{l+1}$) near the origin and fall down at $r_c$; $g_l(r) = r^{l+1} f(r/r_c)$ is a possible candidate.
(5) The *screened* pseudopotential (the one that would result in $\varphi^{\rm PS}(r)$) is obtained by numerical inversion of the Schrödinger equation.
(6) This pseudopiotential is then unscreened by removing Hartree and exchange-correlation contributions related to valence pseudocharge, as discussed above.

One should note that in the course of constructing the pseudopotential, a particular charge distribution (dependent on a reference atomic configuration) was used, that gave rise to Hartree and exchange-correlation parts. Later on, the pseudo-$\varphi$ would vary in the course of iterations, and one should take into account how Hartree and XC contributions

will change. The Hartree term presents no problem since it is linear in charge. The exchange-correlation potential is not linear in $\rho$, and the subtraction of just valence contribution for its "unscreening" is often not satisfactory. What is instead done in reality – a fixed smooth function imitating the core density (just for the sake of calculating $V_{\rm XC}$ is *removed* at the unscreening step and then *added* on each iteration to the actual pseudo-density, before calculating $V_{\rm XC}$. [30] (One could of course remove a true core density, but it is strongly fluctuating, and we would like to avoid the wiggles in the pseudopotential. Usually a more smooth analytical function does the job well.)

As a result of the above scheme, a semi-local pseudopotential is generated that has the following form:

$$V^{\rm PS}(\mathbf{r}) = V^{\rm loc}(\mathbf{r}) + \sum_{l=0}^{l_{\rm max}} V_l^{\rm PS}(r)\hat{P}_l\,, \tag{4.29}$$

where $V^{\rm loc}$ is a local pseudopotential, $\hat{P}_l$ is the angular momentum projection operator; the corresponding expansion coefficients of the pseudopotential $V_l^{\rm PS}$ converge fast with $l$, so that $l_{\rm max} = 1-2$ usually suffices.

The matrix elements in the planewave basis reduce to

$$\begin{aligned}\frac{1}{\Omega}\int d\mathbf{r}\, e^{-i(\mathbf{k}+\mathbf{G})\mathbf{r}}\, V_l^{\rm PS}(r)\,\hat{P}_l\, e^{i(\mathbf{k}+\mathbf{G}')\mathbf{r}} &= \\ = \;\; 4\pi\,\frac{2l+1}{\Omega}P_l(\cos\gamma)\int_0^\infty V_l^{\rm PS}(r)\, j_l\left(|\mathbf{k}+\mathbf{G}|\, r\right)\, j_l\left(|\mathbf{k}+\mathbf{G}'|\, r\right) r^2 dr\,, & \qquad (4.30)\\ \text{with}\quad \cos\gamma = \frac{(\mathbf{k}+\mathbf{G})(\mathbf{k}+\mathbf{G}')}{|\mathbf{k}+\mathbf{G}|\,|\mathbf{k}+\mathbf{G}'|}\,.\end{aligned}$$

Here, $P_l$ are Legendre polynomials. An obvious disagvantage is that matrix elements depend not on $(\mathbf{k}+\mathbf{G})-(\mathbf{k}-\mathbf{G}') = \mathbf{G}-\mathbf{G}'$, but on $(\mathbf{k}+\mathbf{G})$ and $(\mathbf{k}+\mathbf{G}')$ separately. Thus, for $n$ planewaves in basis, the number of matrix elements of a semi-local pseudopotential is $\sim n(n+1)/2$, that may rapidly become prohibitively large for their storage or calculation. Note that the matrix elements of local (Phillips – Kleinmann) pseudopotential of Eq. (4.25) obviously depend on $(\mathbf{G}-\mathbf{G}')$ only and hence their number scales $\sim 8n$ (number of reciprocal latttice sites within the sphere of the radius $2G_{\rm max}$, where $G$ is the planewave cutoff and $2G_{\rm max}$ – the maximal value of $\mathbf{G}-\mathbf{G}'$).

This problem can be solved if *semi-local* pseudopotential $V_l^{\rm PS}(r)$ is transformed to a fully *non-local* but separable form, as proposed by Kleinmann and Bylander.[31]

$$V_l^{\rm PN}(r) = \frac{\left|V_l^{\rm PS}(r)\,\varphi_l^{\rm P}\right\rangle\left\langle\varphi_l^{\rm P}\,V_l^{\rm PS}(r)\right|}{\langle\varphi_l^{\rm P}|V_l^{\rm PS}(r)|\varphi_l^{\rm P}\rangle}\,. \tag{4.31}$$

One can see that

$$V_l^{\rm PN}(r)\left|\varphi_l^{\rm P}\right\rangle = V_l^{\rm PS}(r)\left|\varphi_l^{\rm P}\right\rangle\,. \tag{4.32}$$

[30] Steven G. Louie, Sverre Froyen and Marvin L. Cohen, *Nonlinear ionic pseudopotentials in spin-density-functional calculations*, Phys. Rev. B **26**, 1738–1742 (1982).

[31] Kleinmann and Bylander, Phys. Rev. Lett. **48**, 1425 (1982).

The advantage of such separation is that the pseudopotential aquires a form similar to

$$V^{\mathrm{PN}} = \sum_j |V\rangle \langle V_j| \ ,$$

and its matrix elements in the planewave basis

$$\langle \mathbf{G}| \, V^{\mathrm{PN}} \, |\mathbf{G}'\rangle = \sum_j \langle \mathbf{G}|V_j\rangle\langle V_j|\mathbf{G}'\rangle$$

become convenient for calculation.

Let us emphasize the origin of the problem with semi-local pseudopotential once more. The $\hat{P}_l$ projection in Eq. (4.29) acts on $e^{i(\mathbf{k}+\mathbf{G}')}$ and produces results depending on the direction of $(\mathbf{k}+\mathbf{G})$. If then $(\mathbf{k}+\mathbf{G}')$ has another direction, the angle between them needs to be treated specially, as in Eq. (4.30). In the Kleinmann–Bylander construction, a pseudo-wavefunction is built in that "absorbs" the effect of projection angle – the result is merely a function of $r$. So one can decouple the effects of $(\mathbf{k}+\mathbf{G})$ and $(\mathbf{k}+\mathbf{G}')$. The price we pay is the acccuracy: (4.32) holds exactly only for the reference function, i.e. the one used for the construction of pseudopotential. But in reality, the separation is used over a certain range of energies and eigenstates in the valence band, that is not exactly justified – but may acceptable for well transferable pseudopotentials.

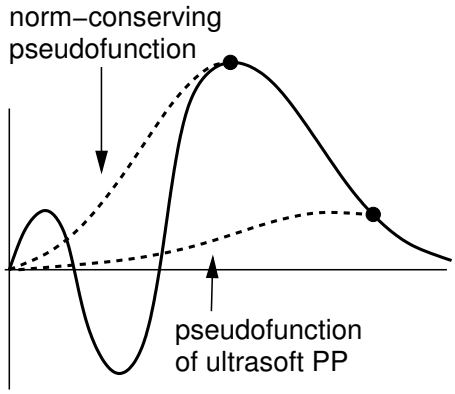
## 4.4 Ultrasoft pseudopotentials

Vanderbilt[32] suggested to abandon the norm-conservation condition (4.27), that would allow to make the pseudoization radius $r_{\mathrm{c}}$ large – essentially, limited only by the condition that the spheres of this radius centered on different atoms must not overlap in a simulation. A big advantage would be that pseudopotentials generated with larger $r_{\mathrm{c}}$ are much softer and hence much lower planewave cutoff would be needed. There are also disadvantages:

• the formalism becomes less transparent and more mathematically involved, because the functions get a non-trivial overlap; charge density is not $\sum |\varphi^{\mathrm{PS}}(\mathbf{r})|^2$ anymore but includes contributions from "local functions";

• a pseudopotential becomes less transferable.

The latter disadvantage is not very serious because one can typically afford generating the pseudopotential anew several times, as the calculation converges and the final charge configuration becomes apparent. The mathematical complications are briefly discussed below, following the original paper by Vanderbilt.




[32] David Vanderbilt, *Soft self-consistent pseudopotentials in a generalized eigenvalue formalism*, Phys. Rev. B **41**, 7892–7895 (1990)

We label all-electron wavefunctions $\psi_i(\mathbf{r})$, where $i$ is composite index, $i = \{\varepsilon_i lm\}$, whereby one allows in principle more than one reference energy $\varepsilon_i$ per $l$-channel. $\psi_i(\mathbf{r})$ is hence a radial solution of the Schrödinger equation at such fixed energy:

$$\left[\hat{T} + V^{\mathrm{AE}}(\mathbf{r})\right] \psi_i(\mathbf{r}) = \varepsilon_i \psi_i(\mathbf{r}) \,. \tag{4.33}$$

One introduces the following cutoff radii:
$r_{cl}$ for wavefunction,
$r_{\mathrm{c}}^{\mathrm{loc}}$ for local pseudopotential,
$R$ – diagnostic radius, beyond which all all-electron and pseudo properties (wavefunctions, potential) agree for all $l$ channels. In the following the bra-ket notation stands for integrals up to $R$:

$$\langle \psi_i | \psi_j \rangle \equiv \int Y^*_{lm}(\Omega) Y_{l'm'}(\Omega) d\Omega \int_0^R r^2 \psi_i(r)\, \psi_j(r) \,.$$

Next we introduce pseudofunctions $\phi_i$ that match corresponding $\psi_i$ smoothly at $r_{cl}$ *and* is norm-conserving:

$$\langle \phi_i | \phi_i \rangle = \langle \psi_i | \psi_i \rangle \,.$$

It will be a solution of the Schrödinger equation with an, in principle, nonlocal pseudopotential:

$$\left[\hat{T} + V^{\mathrm{loc}}(\mathbf{r}) + V^{\mathrm{NL}}(\mathbf{r})\right] \phi_i(\mathbf{r}) = \varepsilon_i \phi_i(\mathbf{r}) \,. \tag{4.34}$$

that is exact for the reference energy $\varepsilon_i$. This allows to construct the "difference" function $\chi_i$,

$$|\chi_i\rangle = \left[\varepsilon_i - \hat{T} - V^{\mathrm{loc}}(\mathbf{r})\right] |\phi_i\rangle \,. \tag{4.35}$$

From (4.34) and (4.35),

$$\begin{aligned} V^{\mathrm{NL}} |\phi_i\rangle &= |\chi_i\rangle \,, \quad \text{hence} \\ V^{\mathrm{NL}} &= \frac{|\chi_i\rangle \langle \chi_i|}{\langle \chi_i | \phi_i \rangle} \,. \end{aligned} \tag{4.36}$$

This bears some similarity to the Phillips – Kleinmann pseudopotential (4.5). Indeed, $|\chi_i\rangle$ has similarity to core states $|\phi_{\mathrm{c}}^{\alpha}\rangle$ in (4.25) in a sense that it incorporates the wiggles to be suppressed in a pseudofunction.

An important extension in the Vanderbilt scheme is the possibility to allow two or more reference energies $\varepsilon_i$ at which the scattering properties of the pseudopotential will be correct. The reference functions $\phi_i$ taken at different $\varepsilon_i$ within the same $l$-channel are of course not orthogonal. these non-orthogonalities will be accounted for by *augmentation coefficients* $Q_{ij}$:

$$Q_{ij} = \langle \psi_i | \psi_j \rangle - \langle \phi_i | \phi_j \rangle \tag{4.37}$$

which are $Q_{ij} = 0$, if norm-conservation condition is imposed – but this is not necessary. Further on, we define

$$B_{ij} = \langle \phi_i | \chi_j \rangle \tag{4.38}$$

and

$$|\beta_i\rangle = \sum_j \left(B^{-1}\right)_{ji} |\chi_j\rangle \qquad \langle \beta_i| = \sum_j \langle \chi_j| \left(B^{-1}\right)_{ij} \,. \tag{4.39}$$

The meaning of $|\beta_i\rangle$ is that they are *localized* (since they disappear outside $R$, as $|\chi_i\rangle$) projectors which are dual to pseudofunctions $|\phi_i\rangle$, i.e.,

$$\langle\beta_i|\phi_i\rangle = \delta_{ij}\,. \tag{4.40}$$

Indeed,

$$\langle\beta_i|\phi_i\rangle = \sum_l \left(B^{-1}\right)_{il} \underbrace{\langle\chi_l|\phi_j\rangle}_{B_{jl}} = \delta_{ij}\,.$$

With this, we arrive at

$$V^{\rm NL} = \sum_{mn} B_{mn}\,|\beta_m\rangle\,\langle\beta_n|\ , \tag{4.41}$$

because from (4.36)

$$V^{\rm NL}\,|\phi_i\rangle = \frac{|\chi_i\rangle\langle\chi_i|\phi_i\rangle}{\langle\chi_i|\phi_i\rangle} = |\chi_i\rangle\,.$$

and from (4.41),

$$\begin{aligned} V^{\rm NL}\,|\phi_i\rangle &= \sum_{mn} B_{mn}|\beta_m\rangle \underbrace{\langle\beta_n|\phi_i\rangle}_{=\delta_{ni}} = \\ &= \sum_m B_{mi}|\beta_m\rangle = \sum_m B_{mi} \sum_l \left(B^{-1}\right)_{lm} |\chi_l\rangle = \\ &= \sum_l \underbrace{\sum_m \left(B^{-1}\right)_{lm} B_{mi}}_{=\delta_{li}}\,|\chi_l\rangle\,. \end{aligned}$$

The norm of all-electron wavefunction is $\langle\psi_i|\psi_i\rangle$. It can be recovered from pseudo-wavefunction if one introduces a non-trivial overlap,

$$\mathbf{S} = \mathbf{1} + \sum_{mn} Q_{mn}|\beta_m\rangle\langle\beta_n|\,, \tag{4.42}$$

so that

$$\begin{aligned} \langle\phi_i|\hat{S}|\phi_j\rangle = \langle\phi_i|\phi_j\rangle &= \langle\phi_i|\phi_j\rangle + \sum_{mn} Q_{mn} \underbrace{\langle\phi_i|\beta_m\rangle}_{\delta_{im}} \underbrace{\langle\beta_n|\phi_j\rangle}_{\delta_{jn}} = \\ &= \langle\phi_i|\phi_j\rangle + Q_{ij} = [\text{using (4.37)}] = \langle\psi_i|\psi_j\rangle\,. \end{aligned}$$

Then (non norm-conserving) functions $|\phi_i\rangle$ satisfy the equation:

$$\left[\hat{T} + V_{\rm loc} + \sum_{mn} D_{mn}|\beta_m\rangle\langle\beta_n|\right]|\phi_i\rangle = \varepsilon_i\hat{S}|\phi_i\rangle \tag{4.43}$$

with

$$D_{mn} = B_{mn} + \varepsilon_n Q_{mn}\,. \tag{4.44}$$

Indeed, the left side of Eq. (4.43) reduces to

$$\underbrace{\left(\hat{T} + V_{\rm loc}\right)|\phi_i\rangle + |\chi_i\rangle}_{\varepsilon_i|\phi_i\rangle} + \sum_m \varepsilon_i Q_{ni}|\beta_m\rangle\,,$$

that equals the right side.

As was discussed before for norm-conserving pseudopotentials, for the final construction of pseudopotential one has to "unscreen" its ion and non-local parts:

$$\begin{aligned} V^{\rm ion}_{\rm lov}(\mathbf{r}) &= V_{\rm loc}(\mathbf{r}) - \int d\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} - V_{\rm XC}(\mathbf{r})\,; \\ D^{(0)}_{mn} &= D_{mn} - \int d\mathbf{r}'\, V_{\rm loc}(\mathbf{r})\,\rho(\mathbf{r}')\,, \end{aligned} \tag{4.45}$$

where $\rho(\mathbf{r})$ is the valence density – a "pseudo" density, but corresponding to norm conservation.

Exactly as the integral over this density over the $R$-sphere can be recovered from Eq. (4.37), the density *at arbitrary* $r < R$ can be reconstructed in terms of *augmentation functions* $Q_{ij}(\mathbf{r})$:

$$Q_{ij}(\mathbf{r}) = \psi^*_i\psi_j(\mathbf{r}) - \phi^*_i\phi_j(\mathbf{r})\,. \tag{4.46}$$

the indices $i$, $j$ numbered so far reference energies. The equations like (4.43) hold exactly for such reference states. Our aim is, however, to solve the bandstructure problem in a whole interval of valence band energies, using plane waves as basis. Therefore the generalized gradient problem

$$(\mathcal{H} - \varepsilon\mathcal{S})\,\Phi_{\alpha\mathbf{k}} = 0\,,$$

of which Eq. (4.43) is an example, must be solved for each $\mathbf{k}$ point, yielding a sequence of bands numbered by $\alpha$. The explicit form of Hamiltonian and overlap includes summations over reference states $i, n$ etc. The (corrected) density includes summation over occupied band states, like in all other methods. Explicitly,

$$\rho(\mathbf{r}) = \sum_{\substack{\alpha,\,\mathbf{k} \\ \text{(occupied)}}} \phi^*_{\alpha\mathbf{k}}(\mathbf{r})\phi_{\alpha\mathbf{k}}(\mathbf{r}) + \sum_{ij} \rho_{ij}\, Q_{ji}(\mathbf{r})\,. \tag{4.47}$$

with $\rho_{ij}$ constructed from projections of occupied band states onto $|\beta_i\rangle$:

$$\rho_{ij} = \sum_{\substack{\alpha,\,\mathbf{k} \\ \text{(occupied)}}} \langle\beta_i|\phi_{\alpha\mathbf{k}}\rangle\langle\phi_{\alpha\mathbf{k}}|\beta_j\rangle\,. \tag{4.48}$$

These terms must be recalculated on each iteration.

While the pseudo-component of the density (4.47) is now smooth, the fluctuating part of it is shifted into augmentation functions. They can be represented by spherical harmonics expansion inside $R$,

$$Q_{ij}(\mathbf{r}) = \sum_{lm} C^{ij}_{lm}\, Y^m_l(\hat{\mathbf{r}})\, Q_{ij}(r)\,,$$

but still they have to be transformed to reciprocal space for integration (e.g., for solving the Poisson equation). Therefore, the fact that they are localized may result in high cutoff in their planewave *expansion*. But this is the price paid for keeping the dimension of of planewave *basis* compact. Moreover, it is not necessary to keep true augmentation

functions; in practice it suffices to imitate them by more smooth $l$-dependent functions which faithfully represent *moments* of true augmentation functions, up to a certain $l_{\max}$.

Let us summarize some important features of ultrasoft potentials scheme.

• The cutoff radius $R$ is only limited by next-neighbor distance (pseudoization spheres must not overlap).

• Within this $R$, individual $r_c$ can be adjusted in order to keep pseudofunction as smooth one can manage.

• The necessary planewave cutoff *for the planewave basis* is drastically reduced.

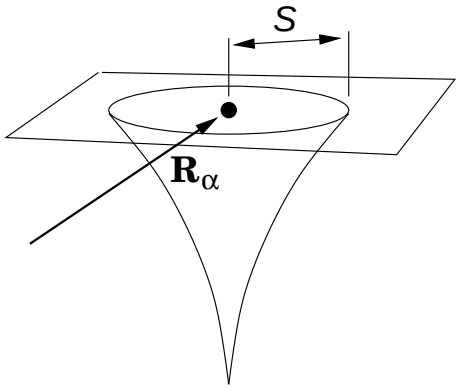
• The amount of computational work (in the generation of pseudopotential etc.) is increased. But most of these additional efforts need not to be repeated in the course of iterations.

• As augmentation charge varies in the course of iterations, so does the local potential in the sphere. It can be considered as part of pseudopotential, so the pseudopotential develops itself as the calculation proceeds (similar to all-electron methods).

• The transferability of pseudopotentials is worse than of conventional pseudopotentials generated with smaller $r_c$. But this is compensated by an option to use several reference energies instead of just one, and by the possibility to generate pseudopotential anew in the course of iterations, as final electron configuration emerges. The main area of application of ultrasoft pseudopotentials is large systems, where the relative cost of the pseudopotential generation is low, as compared to solving the electronic structure problem. For the latter, a low planewave cutoff is a major advantage.

## 4.5 All-electron methods using *muffin-tin* geometry

We discuss now methods tat allow to treat both valence and core states on the same footing, as it was already the case with the use of Gaussian orbitals. Differently from the latter, we consider now the methods where the basis is not fixed (i.e., it develops from iteration to iteration), or where there is no basis, in a sense that the solutions come out **not** from a single diagonalization. Historically, this branch of methods' development starts with the *muffin-tin approximation* and **augmented planewave method** as proposed by Slater in 1937 and 1964. The idea behind it was to use arbitrarily deep – but spherically symmetric – potential near nuclei and a flat one in the interstitial (I) region, with the following dual representation for the wavefunction:




$$\psi_{\mathbf{k}}(\mathbf{r}) = \begin{cases} \frac{1}{\sqrt{\Omega}} \sum\limits_{\mathbf{G}} C_{\mathbf{G}} e^{i(\mathbf{k}+\mathbf{G})\mathbf{r}}\,, & \mathbf{r} \in \mathrm{I} \\ \sum\limits_{lm} A_{lm}^{\alpha\mathbf{k}} u_l^\alpha(r) Y_l^m(\hat{\mathbf{r}})\,, & r < S \end{cases} \qquad (4.49)$$

$u_l(r)$ is a regular (at the origin) solution of radial Schrödinger equation in spherical symmetric potential:

$$\left[ -\frac{d^2}{dr^2} + \frac{l(l+1)}{r^2} + V(r) - \varepsilon_l \right] r u_l(r) = 0\,. \qquad (4.50)$$

$\varepsilon_l$ is a parameter, variable in order to achieve the continuity of the wavefunction at the sphere boundary. The continuity of the dual representation (4.49) is necessary in order to have well defined kinetic energy. The matching condition, relating $C_{\mathbf{G}}$ and $A_{lm}$, follows from the planewave expansion

$$e^{i\mathbf{K}\mathbf{r}} = 4\pi e^{i\mathbf{K}\mathbf{R}_\alpha} \sum_{lm} i^l j_l\left(K|\mathbf{r}-\mathbf{R}_\alpha|\right) Y_l^{m*}(\hat{\mathbf{K}}) Y_l^m(\hat{\mathbf{r}-\mathbf{R}_\alpha}) \tag{4.51}$$

around an arbitrary cite $\mathbf{R}_\alpha$.

$$\frac{1}{\sqrt{\Omega}} \sum_{\mathbf{G}} C_{\mathbf{G}} 4\pi e^{i(\mathbf{k}+\mathbf{G})\mathbf{R}_\alpha} \sum_{lm} i^l j_l\left(|\mathbf{k}+\mathbf{G}|r\right) Y_l^{m*}(\mathbf{k}+\mathbf{G}) Y_l^m(\hat{\mathbf{r}})$$

and

$$\sum_{lm} A_{lm}\, u_l(r)\, Y_l^m(\hat{\mathbf{r}})$$

must match at $r=S$ for all $\{l,m\}$, hence

$$A_{lm}^{\alpha\mathbf{k}} = \frac{4\pi i^l}{\sqrt{\Omega} u^{\alpha_l}(S)} \sum_{\mathbf{G}} e^{i(\mathbf{k}+\mathbf{G})\mathbf{R}_\alpha} j_l\left(|\mathbf{k}+\mathbf{G}|S\right) Y_l^{m*}(\mathbf{k}+\mathbf{G}) \tag{4.52}$$

with $\varepsilon_l$ fixed as external parameter. For a given composition of wavefunctions from plane waves in the interstitial, one can thus recover $A_{lm}$ which would guarantee the matching at $S$.

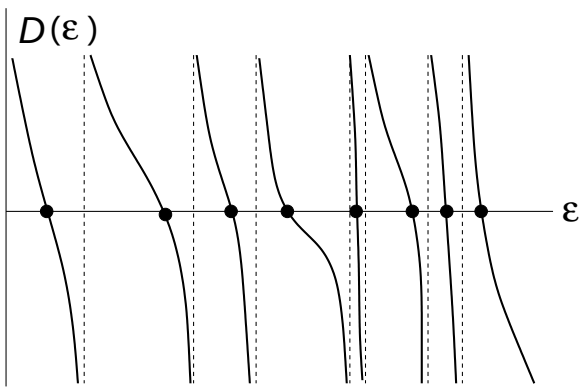
Since the matching can be achieved for every single $\mathbf{G}$-wave, one could exploit the idea of using a global basis set, labeled by $\mathbf{G}$, and inside each sphere there will be an augmentation to $\exp[i(\mathbf{k}+\mathbf{G})\mathbf{r}]$ by corresponding $\sum_{lm} A_{lm}^{\alpha} u_l(|\mathbf{r}-\mathbf{R}_\alpha|)$. There will be Hamiltonian matrix $H_{\mathbf{G}\mathbf{G}'}$ and overlap $S_{\mathbf{G}\mathbf{G}'}$,and one could search for the solutions of the secular equation

$$\|\langle \mathbf{G}_1|\mathcal{H}-\varepsilon\mathcal{S}|\mathbf{G}_2\rangle\| = 0\,.$$

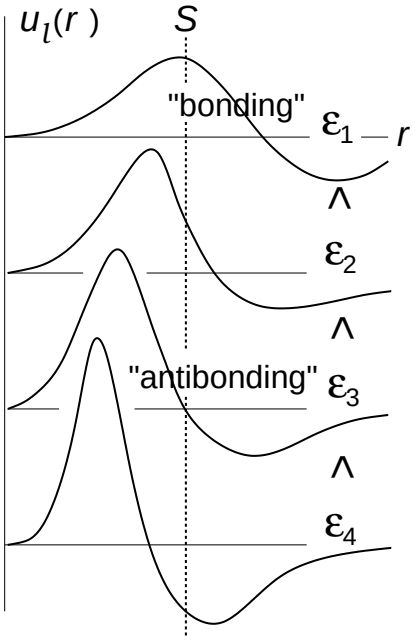
Note that $u_l$ are only solutions inside sphere for a certain energy. If $\varepsilon$ deviates from it, $u_l$ is not a solution anymore. One cannot simply fix the energy parameter $\varepsilon_l$ to any arbitrary value. Instead, one has to scan the values of this parameter that fulfil the condition

$$\|\langle \mathbf{G}_1,\varepsilon|\mathcal{H}-\varepsilon\mathcal{S}|\mathbf{G}_2,\varepsilon\rangle\| \equiv D(\varepsilon) = 0\,,$$

that is *not* a single diagonalization problem. In practice, one calculates the values of the determinant $D(\varepsilon)$ on a sufficiently fine energy mesh and tries to localize its roots. This has to be done for each $\mathbf{k}$ point. The determinant is an ever-decreasing function of energy consisting of many branches; the number of roots within a given energy interval is not known in advance.



An additional complication comes from the fact that for some energy values one or another of atomic solutions $u_l(r)$ may have a node at the sphere boundary, then (4.52) would diverge and augmentation of this particular radial solution to planewaves breaks down – an additional numerical problems, if there is a root nearby. This situation is almost unavoidable, because radial solutions of the Schrödinger equation for each given $l$ show a systematic behavior as the integration energy grows: the function gradually becomes more compact, and at certain energies a next node squeezes into the sphere of radius $S$. This is shown schematically in the following figure. Two special cases $u_l'(S)$=0 and $u_l(S)$=0 are labeled as "bonding" and "antibonding", correspondingly, for further reference.




A related calculation scheme, that is however derived from quite different starting point, is the **Korrings–Kohn–Rostoker** (KKR) method. It uses the fact that in the interstitial, i.e. in the region with constant potential $V_0$, the Schrödinger equation degenerates into the Helmholz equation

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V_0 - \epsilon\right)\psi = 0$$

with a solution

$$\psi = \exp\left[i\sqrt{\frac{2m}{\hbar^2}(\varepsilon - V_0)}\, r\right] \equiv e^{i\kappa r}$$

that can be expanded in spherical harmonics with $j_l(\kappa^2, r)$ – regular at the origin and $n_l(\kappa^2, r)$ – singular at the origin radial functions. Such functions to be used in the following are, up to a normalization, spherical Bessel and Neumann functions:

$$j_l(\kappa^2, r) \equiv j_l(\kappa r)\frac{(2l-1)!!}{2(\kappa S)^l}\,; \qquad n_l(\kappa^2, r) \equiv -n_l(\kappa r)\frac{(\kappa S)^{l+1}}{(2l-1)!!}\,. \tag{4.53}$$

These plane waves are *scattered* on individual atomic potentials, that gives rise to phase shifts $\eta$. The phase shift shows in which relation regular and singular solutions are present in the scattered wave. In the most direct way the relation between $J_l$ and $n_l$ is accounted for by the *potential function* of scattering problem $P_(\epsilon, \kappa^2)$. In terms of it, the wavefunction can be expanded over spherical harmonics inside and outside the sphere with scattering potential in the following way:

$$\psi(\varepsilon, \mathbf{r}) = \sum_{lm} Y_l^m(\hat{\mathbf{r}}) \times \begin{cases} u_l(\varepsilon, r)\,, & r < S \\ n_l(\kappa^2, r) - P_l(\varepsilon, \kappa^2)\, j_l(\kappa^2, r) & r > S \end{cases} \tag{4.54}$$

The potential function can be obtained from the condition that both the radial wavefunction $u_l(\varepsilon, r)$ and its first derivative match a combination of $n_l(\kappa^2, r)$ and $j_l(\kappa^2, r)$:

$$P(\varepsilon, \kappa^2) = \frac{\{u_l(\varepsilon), n_l(\kappa^2)\}}{\{u_l(\varepsilon), j_l(\kappa^2)\}}\,, \tag{4.55}$$

where the Wronskian of two radial functions, taken at the sphere boundary, is defined as

$$\{f, g\} \equiv S^2[f(S)g'(S) - f'(S)g(S)]$$

that gives specifically for the functions (4.53)

$$\{n_l, j_l\} = \frac{S}{2}\,.$$

The so calculated potential function relates to the phase shift $\eta_l$ in the following way:

$$P_l(\varepsilon, \kappa^2) = -\cot\eta_l(\varepsilon, \kappa)\frac{2(S\kappa)^{2l+1}}{[(2l-1)!!]^2}\,. \tag{4.56}$$

Indeed, for the case of no scattering (the potential equal to $V_0$ everywhere inside the sphere) $u_l = j_l$, $P(\varepsilon, \kappa^2) = \infty$ and $\eta = 0$. It follows from Eq. (4.56) that $P_l(\varepsilon, \kappa^2)$ is an always increasing function of $\varepsilon$. Adding $P_l(\varepsilon, \kappa^2)j_l(\kappa^2, r)$ to the wavefunction representation (4.52) results in *muffin-tin orbital* (MTO) $\chi_l$:

$$\chi_{lm}(\varepsilon, \kappa^2, \mathbf{r}) = Y_l^m(\hat{\mathbf{r}}) \times \begin{cases} u_l(\varepsilon, r) + P_l(\varepsilon, \kappa^2)j_l(\kappa^2, r)\,, & r < S \\ n_l(\kappa^2, r)\,. & r > S \end{cases} \tag{4.57}$$

One could in principle fix $\varepsilon$ and use muffin-tin orbitals as localized basis functions, but the quality of this basis would be quite bad for any general $\varepsilon$, for the same reasons as discussed above for augmanted plane waves. Here also, instead of performing a diagonalization one can, as a more practical scheme, scan the integral of energies of interest and search for those values of $\varepsilon$ at which the matching conditions are satisfied at all sphere boundaries simultaneously. This is achieved by taking into account the expansion of non-regular solution, taken at any center $\mathbf{R}_\alpha$, around any other center $\mathbf{R}_\beta$, that has the following form:

$$\begin{aligned} n_{lm}(\kappa^2, |\mathbf{r} - \mathbf{R}_\alpha|) &= \sum_{l'm'} j_{l'm'}(\kappa^2, |\mathbf{r} - \mathbf{R}_\beta|) \times \\ &\times \sum_{l''} 4\pi\, C_{l\ \ l'\ \ l''}^{m m' m''}\, i^{-l+l'-l''}\, \frac{2(\kappa R_\alpha)^{l'+l-l''}(2l''-1)!!}{(2l'-1)!!\,(2l-1)!!}\, n^*_{l''(m'-m)}(\kappa^2, |\mathbf{R}_\alpha - \mathbf{R}_\beta|) \\ &\equiv \sum_{l'm'} j_{l'm'}(\kappa^2, |\mathbf{r} - \mathbf{R}_\beta|)\, S_{\mathbf{R}_\beta l'm', \mathbf{R}_\alpha lm}(\kappa^2)\,. \end{aligned} \tag{4.58}$$

The condition that a linear combination of MTOs is the solution in all space reads

$$\sum_{\mathbf{R}lm} \left[P_{\mathbf{R}'l'}(\varepsilon, \kappa^2)\,\delta_{\mathbf{R}\mathbf{R}'}\,\delta_{ll'}\,\delta_{mm'} - S_{\mathbf{R}'l'm', \mathbf{R}lm}(\kappa^2)\right] C_{\mathbf{R}lm} = 0\,. \tag{4.59}$$

The zeros of the determinant of this equation

$$||\mathbf{P}(\varepsilon, \kappa^2) - \mathbf{S}(\kappa^2)|| = 0$$

determine the solutions of the scattering problem.

Comparing the performance of the KKR method with the augmented plane wave scheme one can note the following advantages:

• a much more compact basis set can be typically used. Whereas a high **G**-cutoff in the augmented plane wave method may typically result in $\sim 1000$ plane waves per atom, the size of the basis set in the KKR is essentially limited by maximal $l$ in the expansion of structure constants. The latter converge quite fast, with $l_{\text{max.}} \approx 2-3$ being typically acceptable, resulting in $(l+1)^2 \sim 16$ functions per atom (for the description of valence states);

• radial solutions and functions in the interstitial are allowed to match *along with their derivatives*.

• The structure constants do not depend on scattering potentials and they are scaled in a simple way as the lattice constant changes; therefore they do not need to be recalculated in the course of electronic-structure iterations and can be stored in advance (as functions of $\kappa^2$) for any given crystal structure.

Still the problem preventing to use a single diagonalization in order to obtain all eigenvalues at once is unsufficient variational freedom of numerical solutions inside the spheres, taken at certain fixed $\varepsilon$. This can be helped by *linearization*, due to an observation by Andersen[33] that $u_l(\varepsilon, r)$ as functions of *energy* is with good accuracy linear:

$$u_l(\varepsilon, r) \approx u_l(\varepsilon_\nu, r) + (\varepsilon - \varepsilon_\nu)\dot{u}_l(\varepsilon_\nu, r)\,. \tag{4.60}$$

The linearization of augmented plane waves leads to the LAPW method, the linearization of scattering-wave formalism – to the LMTO method. The advantage of the former is a relatively easy generalization over non-spherical potentials, the advantage of the latter – compact and highly efficient basis; the inclusion of potentials of general form is also possible but somehow more mathematically involved.

Differentiating Eq. (4.50) in energy yields:

$$\left[-\frac{d^2}{dr^2} + \frac{l(l+1)}{r^2} + V(r) - \varepsilon_l\right] r\,\dot{u}_l(r, \varepsilon_l) = r\,u_l(r, \varepsilon_l)\,, \tag{4.61}$$

with

$$\dot{u}_l \equiv \frac{d}{d\varepsilon} u_l(r, \varepsilon)\,.$$

In the LAPW formalism, both functions $u_l$ and $\dot{u}_l$ are used to match the plane wave on the sphere boundary, whereby the wavefunction expansion (4.49) takes a form:

$$\psi_{\mathbf{k}}(\mathbf{r}) = \begin{cases} \frac{1}{\sqrt{\Omega}} \sum\limits_{\mathbf{G}} C_{\mathbf{G}} e^{i(\mathbf{k}+\mathbf{G})\mathbf{r}}\,, & \mathbf{r} \in \mathrm{I} \\ \sum\limits_{lm} \left[A^{\mathbf{k}}_{lm} u_l(r) + B^{\mathbf{k}}_{lm} \dot{u}_l(r)\right] Y_l^m(\hat{\mathbf{r}})\,, & r < S\,. \end{cases} \tag{4.62}$$

[33] O. Krogh Andersen, *Linear methods in band theory, Phys. Rev. B* **12**, *3060–3083 (1975).*

This combination now has sufficient variational freedom: if band energy deviates from an in advance fixed value $\varepsilon_\nu$, the shape of the wavefunction inside the sphere can still be recovered using the approximation (4.60). One can then use as basis functions $\psi^{\mathbf{G}}(\mathbf{r})$, corresponding to different $\mathbf{G}$ in the dual representation (4.62), with the coefficients $A_{lm}$ and $B_{lm}$ obtained from matching conditions there. The basis functions are then *energy independent*, and the electronic-structure problem can be solved by matrix diagonalization, as in any method with fixed basis. The error in the band energy due to linearization is $\sim (\varepsilon - \varepsilon_\nu)^4$. Typically, it is negligible within an energy window of 1–1.5 Ry.

Two parameters $A$ and $B$ for each $(l, m)$ are fixed by *two* requirements, that the wavefunction and its radial derivative match the plane wave. Therefore the asymptote problem of $u_l(S)$=0 in the denominator, as in Eq. (4.52) does not arise: $u_l(S)$ and $u_l'(S)$ cannot be equal to zero simultaneously. $u_l(S) = 0$ corresponds to the "antibonding" solution (see preceding figure) and $u_l'(S) = 0$ – to the "bonding" solution; their corresponding energies lie apart by $\sim$ the valence band width.

With this, the secular equation becomes:

$$\left[H^{\mathbf{k}}_{\mathbf{G}\mathbf{G}'} - \varepsilon\, S_{\mathbf{G}\mathbf{G}'}\right] C^{\mathbf{k}}_{\mathbf{G}'} = 0\,. \tag{4.63}$$

The expression for the overlap matrix

$$S_{\mathbf{G}\mathbf{G}'} = \frac{1}{\Omega}\int d\mathbf{r} e^{i(\mathbf{G}-\mathbf{G}')\mathbf{r}}\Theta(\mathbf{r}) + \sum_{\mathbf{R}_\alpha} S_\alpha(\mathbf{G},\mathbf{G}') \tag{4.64}$$

contains a step function, $\Theta(\mathbf{r})$, that is = 0 inside any muffin-tin sphere, and $S_\alpha(\mathbf{G},\mathbf{G}')$, contributions to overlap from a sphere centered at $\mathbf{R}_\alpha$. The Hamiltonian matrix

$$\begin{aligned} H^{\mathbf{k}}_{\mathbf{G}\mathbf{G}'} &= \frac{1}{\Omega}\int d\mathbf{r} e^{-i(\mathbf{G}+\mathbf{k})\mathbf{r}}\,\Theta(\mathbf{r})\left[\hat{T} + \hat{V}_{\rm PW}\right] e^{i(\mathbf{G}'+\mathbf{k})\mathbf{r}} + \\ &+ \sum_{\mathbf{R}_\alpha}\left[H_\alpha(\mathbf{G},\mathbf{G}') + V^{\rm NS}_\alpha(\mathbf{G},\mathbf{G}')\right] \end{aligned} \tag{4.65}$$

has $H_\alpha(\mathbf{G},\mathbf{G}')$ – spherical and $V^{\rm NS}_\alpha(\mathbf{G},\mathbf{G}')$ – non-spherical (i.e., from $l \neq 0$) contributions from muffin-tin sphere at $R_\alpha$. The details of practical implementations of full-potential LAPW method are well described in a book by D. Singh.[34]

[34] David J. Singh, *Planewaves, pseudopotentials and the LAPW method*, Kluwer (1994), 115 pp.