

Notes on pseudopotential generation

Paolo Giannozzi

Università di Udine and

Democritos National Simulation Center, Trieste

URL: <http://www.fisica.uniud.it/~giannozz>

November 15, 2009

1 Introduction

When I started to do my first first-principle calculation (that is, my first²-principle calculation) with S. Baroni on CsI under pressure (1985), it became quickly evident that available pseudopotentials (PP's) couldn't do the job. So we generated our own PP's. Since that first experience I have generated a large number of PP's and people keep asking me new PP's from time to time. I am happy that "my" PP's are appreciated and used by other people. I don't think however that the generation of PP's is such a hard task that it requires an official (or unofficial) PP wizard to do this. For this reason I want to share here my (little) experience.

These notes were originally written having in mind my version of the PP generation code and of the various related utilities (still available on the web from my home page, but no longer maintained). I am in the process of adapting them to cover the extended capabilities of the improved `atomic` code, included in the QUANTUM ESPRESSO distribution (<http://www.quantum-espresso.org>). If you remark any inconsistencies, please let me know.

The `atomic` code, written in large part by A. Dal Corso (Democritos and Sissa Trieste) can generate both Norm-Conserving (NC) [1] and Ultrasoft (US) [2] PP's, plus Projector Augmented Waves (PAW) [3] sets. It allows for multiple projectors, full relativistic calculations, spin-split PP's for spin-orbit calculations.

1.1 Who needs to generate a pseudopotential?

There are at least three well-known published sets of NC-PP's: those of Bachelet, Hamann, and Schlüter [4], those of Gonze, Stumpf, and Scheffler [5], and those of Goedecker, Teter, and Hutter [6]. Moreover, all major packages for electronic-structure calculations include a downloadable table of PP's. One could then wonder what a PP generation code is useful for. The problem is that sometimes available PP's will not suit your needs. For instance, you may want:

- a better accuracy;

- PP’s generated with some exotic or new exchange-correlation functional;
- a different partition of electrons into valence and core;
- “softer” PP’s (i.e. PP that require a smaller cutoff in plane-wave calculations);
- PP’s with a core-hole for calculations of X-ray Adsorption spectra;
- all-electron wavefunctions reconstruction (requires the knowledge of all-electron and pseudo orbitals used in the generation of PP’s);

or you may simply want to know what is a PP, how to produce PP’s, how reliable they are.

1.2 About similar work

There are other PP generation packages available on-line. Those I am aware of include:

- the code by José-Luís Martins *et al.*[7]:
<http://bohr.inesc-mn.pt/~jlm/pseudo.html>
- the fhi98PP package[8]:
<http://www.fhi-berlin.mpg.de/th/fhi98md/fhi98PP>
- the OPIUM code by Andrew Rappe *et al.*[9]:
<http://opium.sourceforge.net/>
- David Vanderbilt’s US-PP package [2]:
<http://www.physics.rutgers.edu/~dhv/uspp/index.html>.

Other codes may be available upon request from the authors.

Years ago, it occurred to me that a web-based PP generation tool would have been nice. Being too lazy and too ignorant in web-based applications, I did nothing. I recently discovered that Miguel Marques *et al.* have implemented something like this: see <http://www.tddft.org/programs/octopus/pseudo.php>.

2 Pseudopotential generation, in general

In the following I am assuming that the basic PP theory is known to the reader. Otherwise, see Refs.[1, 4, 7, 8, 9] and references quoted therein for NC-PP’s; Refs.[2, 3] for US-PP’s and PAWsets. I am also assuming that the generated PP’s are to be used in separable form [10] with a plane-wave (PW) basis set.

The PP generation is a three-step process. First, one generates atomic levels and wavefunctions with Density-functional theory (DFT). Second, from atomic results one generates the PP. Third, one checks whether what he got is actually working. If not, one tries again in a different way.

The first step is invariably done assuming a spherically symmetric self-consistent Hamiltonian, so that all elementary quantum mechanics results for the atom apply. The atomic state is defined by the ”electronic configuration”, one-electron states are defined

by a principal quantum number and by the angular momentum and are obtained by solving a self-consistent radial Schrödinger-like (Kohn-Sham) equation.

The second step exists in many variants. One can generate “traditional” single-projector NC-PP’s; multiple-projector NC-PP’s; or US-PP’s. In the following we will consider mostly the case of “traditional” NC-PP’s. The crucial step is the generation of smooth, nodeless “pseudo-orbitals” from atomic all-electron (AE) orbitals. Two popular methods are presently implemented: Troullier-Martins [7] and Rappe-Rabe-Kaxiras-Joannopoulos [9] (RRKJ).

The third step is closer to cooking than to science. There is a large arbitrariness in the preceding step that one would like to exploit in order to get the “best” PP, but there is no well-defined way to do this. Moreover one is often forced to strike a compromise between accuracy and computer time. This step is the main focus of these notes.

3 Step-by-step Pseudopotential generation

If you want to generate a PP for a given atom, the checklist is the following:

1. choose the exchange-correlation functional
2. choose the valence-core partition
3. choose the electronic reference configuration
4. choose which reference states to pseudize, and at which energies
5. choose the type of pseudization
6. choose the matching radii
7. choose the parameters for the “nonlinear core correction” [11]
8. choose the local potential
9. generate the pseudopotential
10. check for transferability

3.1 Choosing the exchange-correlation functional

A large number of exchange-correlation functionals, both in the Local-Density Approximation (LDA) or in the Generalized Gradient Approximation (GGA), are implemented. Most of them have been extensively tested, but beware: some exotic or seldom-used functionals might contain bugs.

PP’s must be generated with the SAME functional that will be later used in calculations. The use of, for instance, GGA functionals with LDA PP’s is inconsistent. This is why the PP file contains information on the DFT level used in their generation: if you or your code ignore it, you do it at your own risk.

Note that functionals may present numerical problems when the charge density goes to zero. For instance, the Becke gradient correction to the exchange may diverge for $\rho \rightarrow 0$. This does not happen in a free atom if the charge density behaves as it should, that is, as $\rho(r) \rightarrow \exp(-\alpha r)$ for $r \rightarrow \infty$. In a pseudoatom, however, a weird behavior may arise around the core region, $r \rightarrow 0$, because the pseudocharge in that region is very small or sometimes vanishing (if there are no filled s states). As a consequence, nasty-looking “spikes” appear in the unscreened pseudopotential very close to the nucleus. This is not nice at all but it is usually harmless, because the interested region is really very small. However in some unfortunate cases there can be convergence problems. If you do not want to see those horrible spikes, or if you experience problems, you have the following choices:

- Use a better-behaved GGA, such as PBE
- Use the nonlinear core correction, which ensures the presence of some charge close to the nucleus.

3.2 Choosing the valence-core partition

This seems to be a trivial step, and often it is: valence states are those that contribute to bonding, core states are those that do not contribute. Things may sometimes be more complicated than this. For instance:

- in transition metals, whose typical outer electronic configuration is: $nd^i(n+1)s^j(n+1)p^k$ (n =main quantum number), it is not always evident that the ns and np states (“semicore states”) can be safely put into the core. The problem is that nd states are localized in the same spatial region as ns and np states, deeper than $(n+1)s$ and $(n+1)p$ states. This may lead to poor transferability. Typically, PP’s with semicore states in the core work well in solids with weak or metallic bonding, but perform poorly in compounds with a stronger (chemical) type of bonding.
- Heavy alkali metals (Rb, Cs, maybe also K) have a large polarizable core. PP’s with just one electron may not always give satisfactory results.
- In some II-VI and III-V semiconductors, such as ZnSe and GaN, the contribution of the d states of the cation to the bonding is not negligible and may require explicit inclusion of those d states into the valence.

In all these cases, promoting the highest core states ns and np , or nd into valence may be a computationally expensive but obliged way to improve poor transferability. .

Note that including semicore states into valence could make your PP harder, will increase the number of electrons, and may require more than one projector per angular momentum, or lead to slightly worse results for those cases in which such inclusion is not needed. Include semicore states into valence only if it is really needed.

3.3 Choosing the electronic reference configuration

This may be any reasonable configuration not too far away from the expected configuration in solids or molecules. As a first choice, use the atomic ground state, unless you have a reason to do otherwise, such as for instance:

- You do not want to deal with unbound states. Very often states with highest angular momentum l are not bound in the atom (an example: the $3d$ state in Si is not bound on the ground state $3s^23p^2$, at least with LDA or GGA). In such a case one has the choice between
 - using one configuration for s and p , another, more ionic one, for d , as in Refs.[4, 5];
 - choosing a single, more ionic configuration for which all desired states are bound;
 - generate PP's on unbound states: requires to choose a suitable reference energy.
- The results of your PP are very sensitive to the chosen configuration. This is something that in principle should not happen, but I am aware of at least one case in which it does. In III-V zincblende semiconductors, the equilibrium lattice parameter is rather sensitive to the form of the d potential of the cation (due to the presence of $p - d$ coupling between anion p states and cation d states [12]). By varying the reference configuration, one can change the equilibrium lattice parameter by as much as 1 – 2%. The problem arises if you want to calculate accurate dynamical properties of GaAs/AlAs alloys and superlattices: you need to get a good theoretical lattice matching between GaAs and AlAs, or otherwise unpleasant spurious effects may arise. When I was confronted with this problem, I didn't find any better solution than to tweak the $4d$ reference configuration for Ga until I got the observed lattice-matching.
- You know that for the system you are interested in, the atom will be in a given configuration and you try to stay close to it. This is not very elegant but sometimes it is needed. For instance, in transition metals described by a PP with semicore states in the core, it is probably wise to chose an electronic configuration for d states that is close to what you expect in your system (as a hand-waiving argument, consider that the $(n + 1)s$ and $(n + 1)p$ PP have a hard time in reproducing the true potential if the nd state, which is much more localized, changes a lot with respect to the starting configuration). In Rare-Earth compounds, leaving the $4f$ electrons in the core with the correct occupancy (if known) may be a quick and dirty way to avoid the well-known problems of DFT yielding the wrong occupancy in highly correlated materials.
- You don't manage to build a decent PP with the ground state configuration, for whatever reason.

NOTE 1: you can calculate PP for a l as high as you want, but you are not obliged to use all of them in PW calculations. The general rule is that if your atom has states

up to $l = l_c$ in the core, you need a PP with angular momenta up to $l = l_c + 1$. Angular momenta $l > l_c + 1$ will feel the same potential as $l = l_c + 1$, because for all of them there is no orthogonalization to core states. As a consequence a PP should have projectors on angular momenta up to l_c ; $l = l_c + 1$ should be the local reference state for PW calculations. This rule is not very strict and may be relaxed: high angular momenta are seldom important (but be careful if they are). Moreover separable PP pose serious constraints on local reference l (see below) and the choice is sometimes obliged. Note also that the highest the l in the PP, the more expensive the PW calculation will be.

NOTE 2: a completely empty configuration ($s^0 p^0 d^0$) or a configuration with fractional occupation numbers are both acceptable. Even if fractional occupation numbers do not correspond to a physical state, they correspond to a well-defined mathematical object.

NOTE 3: if you generate a single-projector PP using a configuration with semicore states in the valence, remember that for each l only the state with lowest n can be used to generate the PP. It is not necessary that the state with same l and higher n is empty, but you have to specify the correct configuration for unscreening.

NOTE 4: PP could in principle be generated on a spin-polarized configuration, but a spin-unpolarized one is typically used. Since PP are constructed to be transferrable, they can describe spin-polarized configurations as well. The nonlinear core correction is typically needed if you plan to use PP in spin-polarized (magnetic) systems.

3.4 Choosing reference states to pseudize, reference energies

With single-projector PP's (one potential per angular momentum l , i.e. one projector per l in the separable form), the choice of the electronic configuration automatically determines the reference states to pseudize: for each l , the bound valence eigenstate is pseudized at the corresponding eigenvalue. It is however possible to generate PP's by pseudizing atomic waves, i.e. regular solutions of the radial Kohn-Sham equation, at any energy. More than one such atomic waves of different energy can be pseudized for the same l , resulting in a PP with more than one projector per l . This possibility considerably extends the number of "degrees of freedom" in the generation of a PP. As a rule of thumb: start first with one projector per l , at the energy of the bound state. For atoms having semicore states in the valence, the obvious choice is to include two projectors, using both bound states. Unfortunately this is currently not possible with the `atomic` code, because the state pseudized at the upper bound-state energy will necessarily have a node (in order to be orthogonal to the one with lower energy). Presently, what you can do is to choose the semicore energy, plus a higher energy that is not too far away the former. You should experiment a bit with more projectors, different pseudization energies, etc.

3.5 Choosing the type of pseudization

Two possible types of pseudization are implemented, both claiming to yield optimally smooth PP's:

- Troullier-Martins [7] (TM)

- Rappe-Rabe-Kaxiras-Joannopoulos [9] (RRKJ).

Both pseudizations replace atomic orbitals in the core region with smooth nodeless pseudo-orbitals. The TM method uses an exponential of a polynomial (see Appendix B); the RRKJ method uses three or four Bessel functions. The former is very robust. The latter may occasionally fail to produce the required nodeless pseudo-orbital. If this happens, use the option to set a small nonzero value of the charge density at the origin: this forces the use of four Bessel functions.

3.6 Choosing the matching radii

Beyond the matching radius r_c , the AE orbital and the corresponding PP orbital match, with continuous first derivative at $r = r_c$. For bound states, r_c is typically at the outermost peak or somewhat larger. The larger the r_c , the softer the potential (less PW needed in the calculations), but also the less transferable. In most cases one has to strike a compromise between softness and transferability.

The choice of r_c is very important, especially in “hard” atoms i.e. second-row elements N, O, F, 3*d* transition metals, rare earths. The frequently-asked question is “how much should I push r_c outwards in order to have reasonable results with a reasonable PW cutoff for unreasonably hard atoms”. There is no well-defined answer. Hard atoms have 2*p* (N, O, F), or 3*d* (transition metals), or 4*f* (rare earths) valence states with no orthogonalization to core states of the same l and no nodes. The choice of r_c at the outermost maximum (typically 0.7-0.8 a.u., even less for 4*f* electrons) yields unacceptably hard PP’s. With a little bit of experience one can say that for second-row (2*p*) elements, $r_c = 1.1 - 1.2$ will yield reasonably good results for 50-70 Ry PW kinetic energy cutoff; for 3*d* transition metals, the same r_c will require > 80 Ry cutoff (highest l have slower convergence for the same r_c). The above holds for TM pseudization. RRKJ pseudization will yield an estimate of the required cutoff.

My advice: use US-PP’s for transition metals and rare earths (for the latter, remember that the problem of DFT reliability preempts the problem of generating a PP). With US-PP’s one can push the r_c outwards quite a bit, at the price of losing norm conservation and introducing an augmentation charge that compensates for the missing charge. The code `ld1.x` can generate US-PP’s starting from a “hard” NC-PP.

Note that it is the hardest atom that determines the PW cutoff in a solid or molecule. Do not waste time trying to find optimally soft PP’s if you have harder atoms around. Also note that one should try to have not too different r_c ’s for different angular momenta, but it is not always possible. The r_c cannot be smaller than the outermost node.

3.7 Choosing the parameters for the nonlinear core correction

The core correction accounts at least partially for the nonlinearity in the exchange-correlation potential. In the generation of a PP one first produces a potential with the desired pseudowavefunctions and pseudoenergies. In order to extract a “bare” PP that can be used in a self-consistent DFT calculation, one subtracts out the screening (Hartree and exchange-correlation) potential generated by the valence charge only.

This introduces an error because the XC potential is not linear in the charge density. With the core correction one keeps a smoothed core charge to be added to the valence charge both at the unscreening step and when using the PP.

The core correction is a must for alkali atoms (especially in ionic compounds) and for PP's to be used in spin-polarized (magnetic) systems. It is recommended whenever there is a large overlap between valence and core charge: for instance, in transition metals if the semicore states are kept into the core. It is never harmful but sometimes it may be of little help.

The smoothing works by replacing the true core charge with a fake, smoother, core charge for $r < r_{cc}$. The parameter r_{cc} is provided on input. If not, it is chosen as the point at which the core charge $\rho_c(r_{cc})$ is twice as big as the valence charge $\rho_v(r_{cc})$. In fact the effect of nonlinearity is important only in regions where $\rho_c(r) \sim \rho_v(r)$. Note that the smaller r_{cc} , the more accurate the core correction, but also the harder the smoothed core charge, and vice versa.

3.8 Choosing the local potential

For single-projector PP's, see Note 1 in Sec. 3.2. If one uses the semilocal form, the choice of the local (l -independent) potential is in most cases natural, and it would affect only PW components with $l > l_c$, that are seldom important. In most PW calculations, however, a separable, fully nonlocal form – one in which the PP's is written as a local potential plus projectors – is used. An arbitrary function can be added to the local potential and subtracted to all l components. Generally one exploits this arbitrariness to remove one l component using it as local potential. The separable form can be either obtained by the Kleinman-Bylander projection [10] applied to single-projector PP's, or directly produced using Vanderbilt's procedure [2] (for single-projector PP's the two approaches are equivalent).

Unfortunately the separable form is not guaranteed to have the correct ground state (the semilocal form is guaranteed, by construction): “ghost” states, having the wrong number of nodes, can appear among the occupied states or close to them, making the PP completely useless. This problem may show up with multiple-projectors and US PP's as well.

The freedom in choosing the “local part” can (and usually must) be used in order to avoid the appearance of ghosts. For PW calculations it is convenient to choose as local part the highest l , because this removes more projectors ($2l + 1$ per atom) than for low l . According to Murphy's law, this is also the choice that more often gives raise to problems, and one is forced to use a different l . Another possibility is to generate a local potential by pseudizing the AE potential.

Note that ghosts are invisible to atomic codes like `ld1.x`, because the algorithm used in the integration of radial orbitals discards states with the wrong number of nodes (they may actually show up under the form of difficult convergence or mysterious errors). A simple and safe way to check for the presence of a ghost is to diagonalize the Kohn-Sham hamiltonian in a basis set of spherical Bessel functions. This kind of test can be performed during transferability tests (See 3.10)

3.9 Generating the pseudopotential

As a first step, one can generate AE wavefunctions and one-electron levels for the reference configuration. This is done by using program `ld1.x`. You must specify in the input data: atomic symbol, what you choose as exchange-correlation functional (not needed if you stick to LDA), electronic reference configuration. A complete description of the input is contained in file `INPUT_LD1`. If you want accurate AE results for heavy atoms, you may want to specify a denser grid in r -space than the default one. The defaults one should be good enough for PP generation, though.

Before you proceed, it is a good idea to verify that the atomic data you just produced actually make sense. Some kind souls have posted on the web a complete set of reference atomic data :

<http://physics.nist.gov/PhysRefData/DFTdata/>

These data have been obtained with the Vosko-Wilk-Nusair functional, that for the unpolarized case is very similar to the Perdew-Zunger functional.

The generation step is also done by program `ld1.x`. One has to supply, in addition to AE data: a list of orbitals to be used in the pseudization (in increasing order of angular momentum), the pseudization energies, the matching radii, the filename where the newly generated PP, is written, plus a number of other optional parameters, fully described in file `INPUT_LD1`.

3.10 Checking for transferability

A simple way to check for correctness and to get a feeling for the transferability of a PP, with little effort, is to test the results of PP and AE atomic calculations on atomic configurations differing from the starting one. The error on total energy differences between PP and AE results gives a feeling on how good the PP is. Just to give an idea: an error ~ 0.001 Ry is very good, ~ 0.01 Ry may still be acceptable. The code `ld1.x` has a “testing” mode in which it does exactly the above operation. You provide the input PP file and a number of test configurations.

You are advised to perform also the test with a basis set of spherical Bessel functions $j_l(qr)$. In addition to revealing the presence of “ghosts”, this test also gives an idea of the smoothness of the potential: the dependence of energy levels upon the cutoff in the kinetic energy is basically the same for the pseudo-atom in the basis of $j_l(qr)$ ’s and for the same pseudo-atom in a solid-state calculation using PW’s.

Another way to check for transferability is to compare AE and PS logarithmic derivatives, also calculated by `ld1.x`. Typically this comparison is done on the reference configuration, but not necessarily. You should supply on input:

- the radius r_d at which logarithmic derivatives are calculated (r_d should be of the order of the ionic or covalent radius, and larger than any of the r_c ’s)
- the energy range E_{min}, E_{max} and the number of points for the plot. The energy range should cover the typical valence one-electron energy range expected in the targeted application of the PP.
- output file names (one for AE, one for PP) where results are written.

The file containing logarithmic derivatives can be easily read and plotted using for instance the plotting program `xmgrace`. Sizable discrepancies between AE and PS logarithmic derivatives are a sign of trouble (unless your energy range is too large or not centered around the range of pseudization energies, of course).

Note that the above checks, based on atomic calculations only, do not replace the usual checks (convergence tests, bond lengths, etc) one has to perform in at least some simple solid-state or molecular systems before starting a serious calculation.

A Atomic Calculations

A.1 Nonrelativistic case

Let us assume that the charge density $n(r)$ and the potential $V(r)$ are spherically symmetric. The Kohn-Sham (KS) equation:

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V(r) - \epsilon\right)\psi(\mathbf{r}) = 0 \quad (1)$$

can be written in spherical coordinates. We write the wavefunctions as

$$\psi(\mathbf{r}) = \left(\frac{R_{nl}(r)}{r}\right)Y_{lm}(\hat{\mathbf{r}}), \quad (2)$$

where n is the main quantum number $l = n - 1, n - 2, \dots, 0$ is angular momentum, $m = l, l - 1, \dots, -l + 1, -l$ is the projection of the angular momentum on some axis. The radial KS equation becomes:

$$\begin{aligned} &\left(-\frac{\hbar^2}{2m}\frac{1}{r}\frac{d^2R_{nl}(r)}{dr^2} + (V(r) - \epsilon)\frac{1}{r}R_{nl}(r)\right)Y_{lm}(\hat{\mathbf{r}}) \\ &- \frac{\hbar^2}{2m}\left(\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial Y_{lm}(\hat{\mathbf{r}})}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2 Y_{lm}(\hat{\mathbf{r}})}{\partial\phi^2}\right)\frac{1}{r^3}R_{nl}(r) = 0. \end{aligned} \quad (3)$$

This yields an angular equation for the spherical harmonics $Y_{lm}(\hat{\mathbf{r}})$:

$$-\left(\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial Y_{lm}(\hat{\mathbf{r}})}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2 Y_{lm}(\hat{\mathbf{r}})}{\partial\phi^2}\right) = l(l+1)Y_{lm}(\hat{\mathbf{r}}) \quad (4)$$

and a radial equation for the radial part $R_{nl}(r)$:

$$-\frac{\hbar^2}{2m}\frac{d^2R_{nl}(r)}{dr^2} + \left(\frac{\hbar^2}{2m}\frac{l(l+1)}{r^2} + V(r) - \epsilon\right)R_{nl}(r) = 0. \quad (5)$$

The charge density is given by

$$n(r) = \sum_{nlm}\Theta_{nl}\left|\frac{R_{nl}(r)}{r}Y_{lm}(\hat{\mathbf{r}})\right|^2 = \sum_{nl}\Theta_{nl}\frac{R_{nl}^2(r)}{4\pi r^2} \quad (6)$$

where Θ_{nl} are the occupancies ($\Theta_{nl} \leq 2l + 1$) and it is assumed that the occupancies of m are such as to yield a spherically symmetric charge density (which is true only for closed shell atoms).

A.1.1 Useful formulae

Gradient in spherical coordinates (r, θ, ϕ) :

$$\nabla\psi = \left(\frac{\partial\psi}{\partial r}, \frac{1}{r} \frac{\partial\psi}{\partial\theta}, \frac{1}{r\sin\theta} \frac{\partial\psi}{\partial\phi} \right) \quad (7)$$

Laplacian in spherical coordinates:

$$\nabla^2\psi = \frac{1}{r} \frac{\partial^2}{\partial r^2}(r\psi) + \frac{1}{r^2\sin\theta} \frac{\partial}{\partial\theta} \left(\sin\theta \frac{\partial\psi}{\partial\theta} \right) + \frac{1}{r^2\sin^2\theta} \frac{\partial^2\psi}{\partial\phi^2} \quad (8)$$

A.2 Fully relativistic case

The relativistic KS equations are Dirac-like equations for a spinor with a “large” $R_{nlj}(r)$ and a “small” $S_{nlj}(r)$ component:

$$c \left(\frac{d}{dr} + \frac{\kappa}{r} \right) R_{nlj}(r) = (2mc^2 - V(r) + \epsilon) S_{nlj}(r) \quad (9)$$

$$c \left(\frac{d}{dr} - \frac{\kappa}{r} \right) S_{nlj}(r) = (V(r) + \epsilon) R_{nlj}(r) \quad (10)$$

where j is the total angular momentum ($j = 1/2$ if $l = 0$, $j = l + 1/2, l - 1/2$ otherwise); $\kappa = -2(j - l)(j + 1/2)$ is the Dirac quantum number ($\kappa = -1$ is $l = 0$, $\kappa = -l - 1, l$ otherwise); and the charge density is given by

$$n(r) = \sum_{nlj} \Theta_{nlj} \frac{R_{nlj}^2(r) + S_{nlj}^2(r)}{4\pi r^2}. \quad (11)$$

A.3 Scalar-relativistic case

The full relativistic KS equations is be transformed into an equation for the large component only and averaged over spin-orbit components. In atomic units (Rydberg: $\hbar = 1, m = 1/2, e^2 = 2$):

$$\begin{aligned} -\frac{d^2 R_{nl}(r)}{dr^2} + \left(\frac{l(l+1)}{r^2} + M(r)(V(r) - \epsilon) \right) R_{nl}(r) \\ - \frac{\alpha^2}{4M(r)} \frac{dV(r)}{dr} \left(\frac{dR_{nl}(r)}{dr} + \langle\kappa\rangle \frac{R_{nl}(r)}{r} \right) = 0, \end{aligned} \quad (12)$$

where $\alpha = 1/137.036$ is the fine-structure constant, $\langle\kappa\rangle = -1$ is the degeneracy-weighted average value of the Dirac’s κ for the two spin-orbit-split levels, $M(r)$ is defined as

$$M(r) = 1 - \frac{\alpha^2}{4} (V(r) - \epsilon). \quad (13)$$

The charge density is defined as in the nonrelativistic case:

$$n(r) = \sum_{nl} \Theta_{nl} \frac{R_{nl}^2(r)}{4\pi r^2}. \quad (14)$$

A.4 Numerical solution

The radial (scalar-relativistic) KS equation is integrated on a radial grid. It is convenient to have a denser grid close to the nucleus and a coarser one far away. Traditionally a logarithmic grid is used: $r_i = r_0 \exp(i\Delta x)$. With this grid, one has

$$\int_0^\infty f(r) dr = \int_0^\infty f(x) r(x) dx \quad (15)$$

and

$$\frac{df(r)}{dr} = \frac{1}{r} \frac{df(x)}{dx}, \quad \frac{d^2 f(r)}{dr^2} = -\frac{1}{r^2} \frac{df(x)}{dx} + \frac{1}{r^2} \frac{d^2 f(x)}{dx^2}. \quad (16)$$

We start with a given self-consistent potential V and a trial eigenvalue ϵ . The equation is integrated from $r = 0$ outwards to r_t , the outermost classical (nonrelativistic for simplicity) turning point, defined by $l(l+1)/r_t^2 + (V(r_t) - \epsilon) = 0$. In a logarithmic grid (see above) the equation to solve becomes:

$$\begin{aligned} \frac{1}{r^2} \frac{d^2 R_{nl}(x)}{dx^2} &= \frac{1}{r^2} \frac{dR_{nl}(x)}{dx} + \left(\frac{l(l+1)}{r^2} + M(r)(V(r) - \epsilon) \right) R_{nl}(r) \\ &\quad - \frac{\alpha^2}{4M(r)} \frac{dV(r)}{dr} \left(\frac{1}{r} \frac{dR_{nl}(x)}{dx} + \langle \kappa \rangle \frac{R_{nl}(r)}{r} \right). \end{aligned} \quad (17)$$

This determines $d^2 R_{nl}(x)/dx^2$ which is used to determine $dR_{nl}(x)/dx$ which in turn is used to determine $R_{nl}(r)$, using predictor-corrector or whatever classical integration method. $dV(r)/dr$ is evaluated numerically from any finite difference method. The series is started using the known (?) asymptotic behavior of $R_{nl}(r)$ close to the nucleus (with ionic charge Z)

$$R_{nl}(r) \simeq r^\gamma, \quad \gamma = \frac{l\sqrt{l^2 - \alpha^2 Z^2} + (l+1)\sqrt{(l+1)^2 - \alpha^2 Z^2}}{2l+1}. \quad (18)$$

The number of nodes is counted. If there are too few (many) nodes, the trial eigenvalue is increased (decreased) and the procedure is restarted until the correct number $n-l-1$ of nodes is reached. Then a second integration is started inward, starting from a suitably large $r \sim 10r_t$ down to r_t , using as a starting point the asymptotic behavior of $R_{nl}(r)$ at large r :

$$R_{nl}(r) \simeq e^{-k(r)r}, \quad k(r) = \sqrt{\frac{l(l+1)}{r^2} + (V(r) - \epsilon)}. \quad (19)$$

The two pieces are continuously joined at r_t and a correction to the trial eigenvalue is estimated using perturbation theory (see below). The procedure is iterated to self-consistency.

The perturbative estimate of correction to trial eigenvalues is described in the following for the nonrelativistic case (it is not worth to make relativistic corrections on top of a correction). The trial eigenvector $R_{nl}(r)$ will have a cusp at r_t if the trial eigenvalue is not a true eigenvalue:

$$A = \frac{dR_{nl}(r_t^+)}{dr} - \frac{dR_{nl}(r_t^-)}{dr} \neq 0. \quad (20)$$

Such discontinuity in the first derivative translates into a $\delta(r_t)$ in the second derivative:

$$\frac{d^2 R_{nl}(r)}{dr^2} = \frac{d^2 \tilde{R}_{nl}(r)}{dr^2} + A\delta(r - r_t) \quad (21)$$

where the tilde denotes the function obtained by matching the second derivatives in the $r < r_t$ and $r > r_t$ regions. This means that we are actually solving a different problem in which $V(r)$ is replaced by $V(r) + \Delta V(r)$, given by

$$\Delta V(r) = -\frac{\hbar^2}{2m} \frac{A}{R_{nl}(r_t)} \delta(r - r_t). \quad (22)$$

The energy difference between the solution to such fictitious potential and the solution to the real potential can be estimated from perturbation theory:

$$\Delta \epsilon_{nl} = -\langle \psi | \Delta V | \psi \rangle = \frac{\hbar^2}{2m} R_{nl}(r_t) A. \quad (23)$$

B Equations for the Troullier-Martins method

We assume a pseudowavefunction R^{ps} having the following form:

$$R^{ps}(r) = r^{l+1} e^{p(r)} \quad r \leq r_c \quad (24)$$

$$R^{ps}(r) = R(r) \quad r \geq r_c \quad (25)$$

where

$$p(r) = c_0 + c_2 r^2 + c_4 r^4 + c_6 r^6 + c_8 r^8 + c_{10} r^{10} + c_{12} r^{12}. \quad (26)$$

On this pseudowavefunction we impose the norm conservation condition:

$$\int_{r < r_c} (R^{ps}(r))^2 dr = \int_{r < r_c} (R(r))^2 dr \quad (27)$$

and continuity conditions on the wavefunction and its derivatives up to order four at the matching point:

$$\frac{d^n R^{ps}(r_c)}{dr^n} = \frac{d^n R(r_c)}{dr^n}, \quad n = 0, \dots, 4 \quad (28)$$

- Continuity of the wavefunction:

$$R^{ps}(r_c) = r_c^{l+1} e^{p(r_c)} = R(r_c) \quad (29)$$

$$p(r_c) = \log \frac{R(r_c)}{r_c^{l+1}} \quad (30)$$

- Continuity of the first derivative of the wavefunction:

$$\frac{dR^{ps}(r)}{dr} = (l+1)r^l e^{p(r)} + r^{l+1} e^{p(r)} p'(r) = \frac{l+1}{r} R^{ps}(r) + p'(r) R^{ps}(r) \quad (31)$$

that is

$$p'(r_c) = \frac{dR(r_c)}{dr} \frac{1}{R^{ps}(r_c)} - \frac{l+1}{r_c}. \quad (32)$$

- Continuity of the second derivative of the wavefunction:

$$\begin{aligned}
\frac{d^2 R^{ps}(r)}{d^2 r} &= \frac{d}{dr} \left((l+1)r^l e^{p(r)} + r^{l+1} e^{p(r)} p'(r) \right) \\
&= l(l+1)r^{l-1} e^{p(r)} + 2(l+1)r^l e^{p(r)} p'(r) + r^{l+1} e^{p(r)} [p'(r)]^2 + r^{l+1} e^{p(r)} p''(r) \\
&= \left(\frac{l(l+1)}{r^2} + \frac{2(l+1)}{r} p'(r) + [p'(r)]^2 + p''(r) \right) r^{l+1} e^{p(r)}. \tag{33}
\end{aligned}$$

From the radial Schrödinger equation:

$$\frac{d^2 R^{ps}(r)}{dr^2} = \left(\frac{l(l+1)}{r^2} + \frac{2m}{\hbar^2} (V(r) - \epsilon) \right) R^{ps}(r) \tag{34}$$

that is

$$p''(r_c) = \frac{2m}{\hbar^2} (V(r_c) - \epsilon) - 2 \frac{l+1}{r_c} p'(r_c) - [p'(r_c)]^2 \tag{35}$$

- Continuity of the third and fourth derivatives of the wavefunction. This is assured if the third and fourth derivatives of $p(r)$ are continuous. By direct derivation of the expression of $p''(r)$:

$$p'''(r_c) = \frac{2m}{\hbar^2} V'(r_c) + 2 \frac{l+1}{r_c^2} p'(r_c) - 2 \frac{l+1}{r_c} p''(r_c) - 2 p'(r_c) p''(r_c) \tag{36}$$

$$\begin{aligned}
p''''(r_c) &= \frac{2m}{\hbar^2} V''(r_c) - 4 \frac{l+1}{r_c^3} p'(r_c) + 4 \frac{l+1}{r_c^2} p''(r_c) \\
&\quad - 2 \frac{l+1}{r_c} p'''(r_c) - 2 [p''(r_c) p''(r_c)]^2 - 2 p'(r_c) p'''(r_c) \tag{37}
\end{aligned}$$

The additional condition: $V''(0) = 0$ is imposed. The screened potential is

$$V(r) = \frac{\hbar^2}{2m} \left(\frac{1}{R^{ps}(r)} \frac{d^2 R^{ps}(r)}{dr^2} - \frac{l(l+1)}{r^2} \right) + \epsilon \tag{38}$$

$$= \frac{\hbar^2}{2m} \left(2 \frac{l+1}{r} p'(r) + [p(r)]^2 + p''(r) \right) + \epsilon \tag{39}$$

Keeping only lower-order terms in r :

$$V(r) \simeq \frac{\hbar^2}{2m} \left(2 \frac{l+1}{r} (2c_2 r + 4c_4 r^3) + 4c_2^2 r^2 + 2c_2 + 12c_4 r^2 \right) + \epsilon \tag{40}$$

$$= \frac{\hbar^2}{2m} \left(2c_2(2l+3) + ((2l+5)c_4 + c_2^2) r^2 \right) + \epsilon. \tag{41}$$

The additional constraint is:

$$(2l+5)c_4 + c_2^2 = 0. \tag{42}$$

References

- [1] D.R. Hamann, M. Schlüter, and C. Chiang, Phys. Rev. Lett. **43**, 1494 (1979).
- [2] D. Vanderbilt, Phys. Rev. B **47**, 10142 (1993).
- [3] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).
- [4] G.B. Bachelet, D.R. Hamann and M. Schlüter, Phys. Rev. B **26**, 4199 (1982).
- [5] X. Gonze, R. Stumpf, and M. Scheffler, Phys. Rev. B **44**, 8503 (1991).
- [6] S. Goedecker, M. Teter, and J. Hutter, Phys. Rev. B **54**, 1703 (1996).
- [7] N. Troullier and J.L. Martins, Phys. Rev. B **43**, 1993 (1991).
- [8] M.Fuchs and M. Scheffler, Comput. Phys. Commun. **119**, 67 (1999).
- [9] A. M. Rappe, K. M. Rabe, E. Kaxiras, and J. D. Joannopoulos, Phys. Rev. B **41**, 1227 (1990) (erratum: Phys. Rev. B **44**, 13175 (1991)).
- [10] L. Kleinman and D.M. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).
- [11] S.G. Louie, S. Froyen, and M.L. Cohen, Phys. Rev. B **26**, 1738 (1982).
- [12] S.H. Wei and A. Zunger, Phys. Rev. B **37**, 8958 (1987).