Notes on pseudopotential generation

Paolo Giannozzi, Università di Udine
e-mail: `paolo.giannozzi@uniud.it`
URL: `http://www.fisica.uniud.it/~giannozz`

Many years ago

1 Introduction

When I started to do my first first-principle calculation (that is, my first^2-principle calculation) with Stefano Baroni on CsI under pressure (1985), it became quickly evident that the 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. However I don't think 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 and the (primitive) computer codes I am using.

1.1 Disclaimers, acknowledgments, etc. etc.

Please note that

– this PP program set is provided "as is". No warranty, no liability, don't blame me if the PP you generate does not work, etc. etc.. Use at your own risk and under your responsibility.

– Several people have contributed to the computer codes. Among them let me cite

Stefano Baroni, Stefano de Gironcoli, Andrea Dal Corso, Claudia Bungaro, Giovanni Cangiani, Antonio Perin, Mauro Boero, Juerg Hutter

and let me apologize with those I forgot to mention.

– Since the origin of parts of the code is unknown, I cannot guarantee that they are free from copyright/license/patent paranoia. If this is the case, please let me know. Please note that the routine `functionals.f` is derived from the CPMD package.

1.2 About similar work

There are at least three well-known published sets of PP's: those of Bachelet, Hamann, and Schlüter [1](BHS), those of Gonze, Stumpf, and Scheffler [2], and those of Goedecker, Teter, and Hutter [3]. These published tables are a very good and important piece of work. If a published PP suits your needs, you should definitely use it and cite it. One could then wonder what a PP generation code is useful for. The problem is that sometimes published PP's will not suit your needs. For instance:

– the alkali atoms from the BHS tables are ok if you calculate properties of alkali metals, but they yield poor equilibrium lattice parameters in alkali halides (this is why I started with the PP of Cs in CsI, by the way). In such compounds you definitely need the "core correction" of Froyen, Louie, and Cohen [4].

– there are many gradient-corrected functionals around, but only a few published gradient-corrected PP's.

– you may need a distinction between "valence" and "core" electrons that is different from the one chosen in the tables.

– you may need "softer" PP's than those of the tables.

There are other PP generation packages available on on-line (I am aware of the code by José-Luís Martins *et al.*: `http://bohr.inesc-mn.pt/~jlm/pseudo.html`, and of the `fhi98PP` package [5]), or upon request from the authors (I have heard of codes by Hamann and by Bachelet). If you are happy with a given code, use that one. If you are happy with this, use this.

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`.

Ultrasoft (Vanderbilt) PP's are a wholly different story. David Vanderbilt's package can be found on-line at: `http://www.physics.rutgers.edu/~dhv/uspp/index.html`.

2 Pseudopotential generation, in general

In the following I am assuming that the basic theory of PP is known to the reader. If not, a good starting point is ref.[5] and references quoted therein. I am also assuming that the generated PP's are to be used with a plane-wave (PW) basis set.

The PP generation is a three-step process. First, one generates DFT atomic levels and wavefunctions. From this, one generates the PP. Then, one checks whether what he just did is useful, and if it is not, 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. I have implemented two methods: Troullier and Martins [6] (TM) and what is usually referred to as "Von Barth and Car" (VBC). I tend to use the TM because it is quick and simple. The VBC approach works well only for light atoms and requires more work. It has the advantage of producing simple analytical PP's, but with the advent of the Kleinman-Bylander[7] (KB) projection one needs anyway wavefunctions, in numerical form.

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 a suitable density functional

2. choose valence and core states

3. choose a reference configuration

4. generate all-electron wavefunctions

5. choose the matching radii

6. choose if you want the core corrections

7. generate the pseudopotential

8. check its transferability

9. check the required cutoff

10. check its separable form

3.1   Choosing a suitable density functional

A large number of density functionals are implemented in the codes. Many of them have been extensively tested (and some bugs in the original version fixed) but beware: some never-used density functional could contain bugs.

PP's must be generated with the SAME functional that will be later used in calculations. The use of, for instance, gradient-corrected functionals with Local-Density Approximation (LDA) PP's is inconsistent.

NOTE THAT gradient corrections may present numerical problems when the charge density goes to zero. For instance, the Becke gradient correction to the exchange may diverge for $\rho \to 0$. This does not happen in a free atom if the charge density behaves as it should, that is, as $\rho(r) \to exp(-\alpha r)$ for $r \to \infty$. However in a pseudoatom a

weird behavior may arise around the core region, $r \to 0$, 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 better-behaved gradient corrections, such as PBE

– Use good old analytic PP's generation (VBC)

– Use the core corrections (so that there is enough charge close to the nucleus).

– cut out the gradient correction for small r (set variable `rcut` to ~ 0.001 or so).

3.2   Choosing valence and core states

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

– in transition metals, whose typical outer electronic configuration is $nd^i(n + 1)s^j(n + 1)p^k$, it is not always evident that the ns and np 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 intolerable loss of transferability.

– Heavy alkali metals (Rb, Cs, maybe also K) have a large polarizable core. PP's with just one electron may not work properly (even with the core correction, see below)

– $3d$ states in GaAs may safely be left in the core (maybe with a core correction for Ga), not so in GaN.

– In ZnSe and other II-VI the d state of the cation has a nonnegligible contribution to the bonding.

In all these cases, promoting the highest core states ns and np, or nd (the "semicore" states) into valence may be a computationally expensive but obliged way to improve poor transferability. It may happen that the same atom that works great with the "natural" valence in a solid with weak or metallic bonding does not work well at all in another compound with a different (stronger) type of bonding. This happens for many transition and noble metals. Note that including semicore states into valence could make your PP harder, will increase the number of electrons, and may also make your PP slightly worse for those cases in which such inclusion is not needed. Do it only if it is needed.

3.3 Choosing a reference configuration

This may be any reasonable configuration not too far away from the expected configuration in solids or molecules. Use the ground state if you do not have a good reason to do otherwise. The good reason may be one of the following:

– You want to use a single configuration for all angular momenta. 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 $3s^2 3p^2$, at least not in LDA). In such a case one has the choice between

  – using two different configurations, as it is done in the BHS paper;
  – choosing a single, more ionic configuration;
  – use Hamann's approach[8] for generating PP's on unbound states (unfortunately not implemented here).

  I used to choose the first option. Now I am becoming lazy and choose the second.

– your results are very sensitive to the chosen configuration. This is something that in principle should not happen, but I am aware of at least one very important case in which it happens. 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 [9]). By varying the reference configuration one can change the equilibrium lattice parameter by as much as $1 - 2\%$. This would not be a problem: it is still in the typical range of errors for the PW-PP-LDA method. Unfortunately if you want to accurately calculate properties of GaAs/AlAs alloys and superlattices, you have to force the lattice constant of your GaAs to match almost exactly that of AlAs (as it happens in real world), or otherwise unpleasant spurious effects may arise. When I was confronted with this problem, I did not find any better solution than to tweak the $4d$ reference configuration for Ga. If you have better ideas, please go ahead. The "core correction" partially cures the problem.

– you know that the atom will be in a given configuration in the system you are interested in and you try to stay close to it. This is not very elegant but sometimes it is useful: in transition metals, with semicore states in the core, it is probably better to chose a reasonable configuration for d states and not to use it for systems with very different d configuration. The problem is that the $(n+1)s$ and $(n+1)p$ PP have a hard time in reproducing the true potential if the nd state changes a lot with respect to the starting configuration. Consider this a "hand-waiving argument".

– you do not manage to get a separable (KB) PP that work. See later.

NOTE 1: you can calculate PP for as high l 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 and $l = l_c + 1$ 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.

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

NOTE 3: in the TM approach you generate the PP for every l on the valence state of lowest main quantum number n. If there are states with the same l and higher n, they MUST BE EMPTY.

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 "core correction" is typically needed if you plane to use PP in spin-polarized (magnetic) systems. In any event: the atomic code presently does not perform true spin-polarized calculations. It can calculate the energy perturbatively: up and down orbitals are the same, the total energy is calculated using a spin-polarized functional, assuming Hund's rule for spin occupations. As a matter of fact, the energy calculated in this way is quite a good approximation.

3.4 Generating all-electron results

You may now generate all-electron (AE) wavefunctions and one-electron levels for the reference configuration. This is done by using program `ld1`. You must specify in the input data: atomic symbol, what you chose as density functional (not needed if you stick to LDA), the name of a wavefunction file, electronic reference configuration. A complete description of the input is in the header of file `Src/ld1.f`. 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 is very similar (at least for the unpolarized case) to the Perdew-Zunger functional. You may want to have a look at them.

3.5 Choosing the matching radii

At the matching radius r_c the AE and PP wavefunction of angular momentum l match, with at least continuous first derivative. The choice of the r_c is very important and must be guided by the following criteria:

– the r_c must be larger than the outermost node (if any) of the wavefunction for any given l

– a typical r_c is the outermost peak, beyond if needed (see following point)

– the larger the r_c, the softer the potential (less PW needed), but also the less transferable

– usually there is one l that is harder than the others (in transition metals, the d state, in second-row elements N, O, F, the p state). One should concentrate on this one and push outwards its r_c as much as possible.

– this is not very important, and usually impossible to achieve: one should try to have not too different r_c's for different angular momenta

To make a long story short: the difficult question is "how much should I push r_c outwards in order to have reasonable results for unreasonable hard atoms". There is no well-defined answer. A typical wavefunction maximum for hard atoms is 0.7-0.8 a.u (in general, the outermost peak, but hard atoms are those with $2p$, $3d$, $4f$ valence states, with no orthogonalization to core states of the same l and no nodes) (this is why they are hard, anyway). $r_c = 0.8$ a.u. will yield unacceptably hard PP's. With a little bit of experience one can say that for second-row ($2p$) elements $r_c = 1.1 - 1.2$ will yield reasonably good results for 50-70 Ry PW kinetic energy cutoff; for $3d$ transition metals, the same r_c will require > 80 Ry cutoff (highest l have slower convergence for the same r_c).

NOTE 1: in the TM approach the wavefunction matching beyond r_c is exact; in the VBC approach, it is approximate (through the "penalty function"). As a consequence, in order to have similar convergence and transferability, the r_c in the VBC approach should be chosen smaller than in the TM approach.

NOTE 2: it is the hardest atom that determines the PW cutoff in a solid or molecule. Do not waste time optimizing soft PP if you have harder atoms around.

3.6 Choosing whether you want the 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 separate a "bare" PP from the screening part, one removes the screening 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 halides 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 may be useless.

The smoothing works by replacing the true core charge with a fake, smoother, core charge for $r < r_{cc}$, with r_{cc} defined by $\rho_v(r_{cc}) = a\rho_c(r_{cc})$ and $a \sim 0.5$ typically. In fact the effect of nonlinearity is important only in regions where $\rho_c(r) \sim \rho_v(r)$. Note that

for smaller a one gets smaller r_{cc}, more accurate core correction, but also a harder core charge, and vice versa,

3.7   Generating the pseudopotential

The generation step is done in program `trou` for the TM case, in program `fitcar` for the VBC approach. In the TM case one has to supply interactively:

1. a wavefunction file produced by `ld1`

2. Number of valence electrons

3. if the core charge is desired. If yes, the smoothing factor for the core charge, see above.

4. Number of valence states to be used in the generation: one (ONLY one) for each angular momentum

5. For each angular momentum ($s, p, d, f, ...$ states, IN THIS ORDER): wavefunction label. A default core radius (the outermost maximum) is proposed. Type carriage return to accept or a different core radius. If there is more than one state of same l in the valence, the PP must be built on the lowest, the other must be empty.

6. A filename of a previously generated PP (if any). When different configurations are used for different l (typically one for s and p and one for d) first one produces a file for one configuration (for s and p), then one produces the PP for the second configuration (for s, p, d) and reads the previous file (so discarding the s and p produced in the second run).

7. A filename for the newly generated PP

Information on how the the PP was generated is written in human-readable form in the PP file.

In the VBC case one has to give the starting values of the parameters and instruction on the minimization between items 5 and 6.

3.8   Checking for transferability

There is no well-established criterion, to the best of my knowledge, for checking PP transferability. An obvious way to check correctness and to get a feeling for transferability, 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. It is very important to check that the pseudowavefunction match the atomic wavefunction as accurately as possible beyond r_c.

Very primitive plotting programs, producing postscript output, are available for plotting and comparing pseudo and atomic wavefunctions (`wfcgraf`) and for plotting the pseudopotential in real space (`psrgraf`).

NOTE: in order to use the plotting programs, the labeling of states of the PP calculation should be the same as for the AE calculations without core states. The appropriate principal quantum numbers for valence electrons (i.e. the lowest s state must have $n = 1$, the lowest p and second s states must have $n = 2$ and so on) are automagically generated from wavefunctions labels (please check!). Labels are used to identify the AE-PP wavefunction correspondence.

Another way to check for transferability is to compare AE and PS logarithmic derivatives, calculated by program ld1. 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 latter is optional). The energy range should cover the typical valence one-electron energy range expected in the targeted application of the PP.

– an output file name (optional) where results are written.

The output file name is appended .ae for an AE calculation, .pp and .kb (see later) for a PS calculation. The format can be easily read and plotted using for instance the plotting program xmgr. Sizable discrepancies between AE and PS logarithmic derivatives are a sign of trouble (unless your energy range is too low or too high, of course).

3.9 Checking for required cutoff

One can have a feeling of the hardness of the potential from the following indicators:

– the value of r_c (see above)

– the behavior of the Fourier transform $V_l(q)$. A very primitive plotting program psqgraf can be used to plot $V_l(q)$ on a postscript file. Note that q^2 is the energy in Ry.

– an atomic calculation with a basis set of spherical Bessel functions $j_l(qr)$ (equivalent to projecting PW's on states with given l).

The latter is performed by program ldb. It requires on input a PP file, the energy cutoff in Ry, and the electronic configuration. The first set of numbers, under "Semilocal orbital energies:", shows the one-electron levels. At convergence they should be, within numerical accuracy, equal to those of program ld1 with the same PP and electronic configuration. The behavior of the energy levels as a function of the cutoff will be very close to what one will get in a PW calculation.

NOTE: the above does not replace the usual cutoff checks in solid-state or molecular calculations. Direct checks on the quantity of interest in your system are still needed.

3.10 Checking for separable form

The separable (KB) form of PP's is exceedingly convenient in electronic structure calculations, unless you are doing exceedingly simple systems (crystalline silicon...). In the KB formalism, one rewrites the BHS PP's as projectors. An arbitrary function can be added to the local (l-independent) part of the PP and subtracted to all l components. Generally one exploits this arbitrariness to remove one l component using it as local part.

Unfortunately the KB projection can lead to loss of transferability (often negligible, sometimes not) or even to the appearance of "ghost states" – states with the wrong number of nodes that are absent in the all-electron atom – that make the PP completely useless. For practical use, one has to check carefully whether KB PP are useful or not. In particular one must exploit the freedom in choosing the "local part" in order avoid 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. This is the default choice done by program `trou`. According to Murphy's law, this is also the choice that more often gives raise to problems. In the PP file `lloc` is the 7th field (between the two logicals) of the second line. This MUST BE CHANGED if it is not appropriate.

There are three methods to check for ghosts:

– use known criteria for the appearance of "ghosts"[2] (not implemented).

– compare the all-electron (or PP) logarithmic derivative with the KB logarithmic derivative (see above). The former are written by program `ld1` to a file with suffix `.pp`. The latter (calculated using the KB form for PP but keeping the same SCF potential of semilocal PP) are written to a file with suffix `.kb`. Plot the two, for instance using `xmgr`. Any sizable discrepancy between the two is a sign of trouble. Note that for `l=lloc` you must obtain exactly the same values. Use the value of `lloc` that seems to yield better results: edit the PP file to change it (see above).

– use program `ldb`. Atomic codes using the integration of radial wavefunction (like `ld1`) are unable to find spurious states, because their algorithm discards states with wrong number of nodes. However if a basis set is used the spurious states will show up. The self-consistent calculation is performed first on the PP in the normal form. Then a KB Hamiltonian is constructed, using the atomic pseudowavefunction and the screening potential just calculated, and diagonalized on the spherical Bessel basis set (without redoing the self-consistency, it is not needed for our purposes). The results are written in the right-hand part of the output, after "KB orbital energies:", and should match as exactly as possible the results of the left-hand part. This must be tried at least on the reference configuration, for a few occupied and unoccupied (pseudo)states, for all possible local parts.

A very unfortunate case may sometimes occur: you generate a marvelous PP but, no matter what you choose as local reference state, ghosts appear. You have to retry with a different reference configuration or matching radii. Good luck.

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 equation:

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

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

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

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

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

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

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

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

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

The charge density is given by

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

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):

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

Laplacian in spherical coordinates:

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

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:

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

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

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

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

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: $\hbar = 1, m = 1/2, e^2 = 2$):

$$-\frac{d^2 R_{nl}(r)}{dr^2} + \left(\frac{l(l + 1)}{r^2} + M(r)\left(V(r) - \epsilon\right)\right) R_{nl}(r)$$

$$-\frac{\alpha^2}{4M(r)} \frac{dV(r)}{dr}\left(\frac{dR_{nl}(r)}{dr} + \langle\kappa\rangle\frac{R_{nl}(r)}{r}\right) = 0, \tag{12}$$

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

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

The charge density is defined as in the nonrelativistic case:

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

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

$$\int_0^\infty f(r)dr = \int_0^\infty f(x)r(x)dx \tag{15}$$

and

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

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:

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

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)

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

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 $r \sim 10 r_t$ down to r_t, using as a starting point the asymptotic behavior of $R_{nl}(r)$ at large r:

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

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:

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

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

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

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

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

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

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

B   Equations for the Troullier-Martins method

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

$$R^{ps}(r) = r^{l+1} e^{p(r)} \quad r \leq r_c \tag{24}$$
$$R^{ps}(r) = R(r) \quad r \geq r_c \tag{25}$$

where

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

On this pseudowavefunction we impose the norm conservation condition:

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

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

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

• Continuity of the wavefunction:

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

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

• Continuity of the first derivative of the wavefunction:

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

that is

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

• 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 $p''(r)$:

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

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

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

$$
\begin{aligned}
V(r) &= \frac{\hbar^2}{2m}\left(\frac{1}{R^{ps}(r)}\frac{d^2 R^{ps}(r)}{dr^2} - \frac{l(l+1)}{r^2}\right) + \epsilon \tag{38} \\
&= \frac{\hbar^2}{2m}\left(2\frac{l+1}{r}p'(r) + [p(r)]^2 + p''(r)\right) + \epsilon \tag{39}
\end{aligned}
$$

Keeping only lower-order terms in r:

$$
\begin{aligned}
V(r) &\simeq \frac{\hbar^2}{2m}\left(2\frac{l+1}{r}(2c_2 r + 4c_4 r^3) + 4c_2^2 r^2 + 2c_2 + 12c_4 r^2\right) + \epsilon \tag{40} \\
&= \frac{\hbar^2}{2m}\left(2c_2(2l+3) + \left((2l+5)c_4 + c_2^2\right)r^2\right) + \epsilon. \tag{41}
\end{aligned}
$$

The additional constraint is:

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

References

[1] G.B. Bachelet, D.R. Hamann and M. Schlüter, Phys. Rev. B **26**, 4199 (1982).

[2] X. Gonze, R. Stumpf, and M. Scheffler, Phys. Rev. B **44**, 8503 (1991).

[3] S. Goedecker, M. Teter, and J. Hutter, Phys. Rev. B **54**, 1703 (1996).

[4] S.G. Louie, S. Froyen, and M.L. Cohen, Phys. Rev. B **26**, 1738 (1982).

[5] M.Fuchs and M. Scheffler, Comput. Phys. Commun. **119**, 67 (1999). Web site: http://www.fhi-berlin.mpg.de/th/fhi98md/fhi98PP.

[6] N. Troullier and J.L. Martins, Phys. Rev. B **43**, 1993 (1991).

[7] L. Kleinman and D.M. Bylander, Phys. Rev. Lett. 48, 1425 (1982).

[8] D.R. Hamann, Phys. Rev. B **40**, 2980 (1989).

[9] S.H. Wei and A. Zunger, Phys. Rev. B **37**, 8958 (1987).