Lecture Notes per il Corso di *Struttura della Materia* (Dottorato di Fisica, Università di Pisa, v.2005): DENSITY FUNCTIONAL THEORY FOR ELECTRONIC STRUCTURE CALCULATIONS

Paolo Giannozzi
Scuola Normale Superiore, Piazza dei Cavalieri 7
I-56126 Pisa, Italy

Contents

1 Density Functional Theory

Density Functional Theory (DFT) is a *ground-state* theory in which the emphasis is on the *charge density* as the relevant physical quantity. DFT has proved to be highly successful in describing structural and electronic properties in a vast class of materials, ranging from atoms and molecules to simple crystals to complex extended systems (including glasses and liquids). Furthermore DFT is computationally very simple. For these reasons DFT has become a common tool in *first-principles* calculations aimed at describing – or even predicting – properties of molecular and condensed matter systems.

1.1 The Hohenberg-Kohn Theorem

Let us consider a system of N interacting (spinless) electrons under an external potential $V(\mathbf{r})$ (usually the Coulomb potential of the nuclei). If the system has a nondegenerate ground state, it is obvious that there is only one ground-state charge density $n(\mathbf{r})$ that corresponds to a given $V(\mathbf{r})$. In 1964 Hohenberg and Kohn demonstrated the opposite, far less obvious result: there is only one external potential $V(\mathbf{r})$ that yields a given ground-state charge density $n(\mathbf{r})$. The demonstration is very simple and uses a *reductio ad absurdum* argument.

Let us consider a many-electron Hamiltonian $H = T + U + V$, with ground state wavefunction Ψ. T is the kinetic energy, U the electron-electron interaction, V the external potential. The charge density $n(\mathbf{r})$ is defined as

$$n(\mathbf{r}) = N \int |\Psi(\mathbf{r}, \mathbf{r}_2, \mathbf{r}_3, ..., \mathbf{r}_N)|^2 d\mathbf{r}_2 ... d\mathbf{r}_N. \tag{1}$$

Let us consider now a *different* Hamiltonian $H' = T + U + V'$ (V and V' do not differ simply by a constant: $V - V' \neq$const.), with ground state wavefunction Ψ'. Let us assume that the ground state charge densities are the same: $n[V] = n'[V']$. The following inequality holds:

$$E' = \langle \Psi'|H'|\Psi'\rangle < \langle \Psi|H'|\Psi\rangle = \langle \Psi|H + V' - V|\Psi\rangle \tag{2}$$

that is,

$$E' < E + \int (V(\mathbf{r}) - V'(\mathbf{r}))n(\mathbf{r})d\mathbf{r}. \tag{3}$$

The inequality is strict because Ψ and Ψ' are different, being eigenstates of different Hamiltonians. By reversing the primed and unprimed quantities, one obtains an absurd result. This demonstrates that no two different potentials can have the same charge density.

A subtle point about the existence of the potential corresponding to a given ground state charge density (the *v-representability* problem), and the various extensions of the Hohenberg and Kohn theorem, are discussed in the specialized literature.

A straightforward consequence of the first Hohenberg and Kohn theorem is that the ground state energy E is also uniquely determined by the ground-state charge density. In mathematical terms E is a *functional $E[n(\mathbf{r})]$* of $n(\mathbf{r})$. We can write

$$E[n(\mathbf{r})] = \langle \Psi|T + U + V|\Psi\rangle = \langle \Psi|T + U|\Psi\rangle + \langle \Psi|V|\Psi\rangle = F[n(\mathbf{r})] + \int n(\mathbf{r})V(\mathbf{r})d\mathbf{r} \tag{4}$$

where $F[n(\mathbf{r})]$ is a *universal* functional of the charge density $n(\mathbf{r})$ (and *not* of $V(\mathbf{r})$). For this functional a variational principle holds: the ground-state energy is *minimized* by the ground-state charge density. In this way, DFT exactly reduces the N-body problem to the determination of a 3-dimensional function $n(\mathbf{r})$ which minimizes a functional $E[n(\mathbf{r})]$. Unfortunately this is of little use as $F[n(\mathbf{r})]$ is not known.

1.2 The Kohn-Sham equations

One year later, Kohn and Sham (KS) reformulated the problem in a more familiar form and opened the way to practical applications of DFT. The system of interacting electrons is mapped on to an auxiliary system of non-interacting electrons having the same ground state charge density $n(\mathbf{r})$.

For a system of non-interacting electrons the ground-state charge density is representable as a sum over one-electron orbitals (the *KS orbitals*) $\psi_i(\mathbf{r})$:

$$n(\mathbf{r}) = 2\sum_i |\psi_i(\mathbf{r})|^2, \tag{5}$$

where i runs from 1 to $N/2$ if we assume double occupancy of all states, and the KS orbitals are the solutions of the Schrödinger equation

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V_{KS}(\mathbf{r})\right)\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\mathbf{r}) \tag{6}$$

(m is the electron mass) obeying orthonormality constraints:

$$\int \psi_i^*(\mathbf{r})\psi_j(\mathbf{r})d\mathbf{r} = \delta_{ij}. \tag{7}$$

The existence of a unique potential $V_{KS}(\mathbf{r})$ having $n(\mathbf{r})$ as its ground state charge density is a consequence of the Hohenberg and Kohn theorem, which holds irrespective of the form of the electron-electron interaction U.

1.3   Kohn-Sham equations and the variational principle

The problem is now to determine $V_{KS}(\mathbf{r})$ for a given $n(\mathbf{r})$. This problem is solved by considering the variational property of the energy. For an arbitrary variation of the $\psi_i(\mathbf{r})$, under the orthonormality constraints of Eq. (7), the variation of E must vanish. This translates into the condition that the functional derivative (see appendix) with respect to the ψ_i of the constrained functional

$$E' = E - \sum_{ij} \lambda_{ij}\left(\int \psi_i^*(\mathbf{r})\psi_j(\mathbf{r})d\mathbf{r} - \delta_{ij}\right), \tag{8}$$

where λ_{ij} are Lagrange multipliers, must vanish:

$$\frac{\delta E'}{\delta\psi_i^*(\mathbf{r})} = \frac{\delta E'}{\delta\psi_i(\mathbf{r})} = 0. \tag{9}$$

It is convenient to rewrite the energy functional as follows:

$$E = T_s[n(\mathbf{r})] + E_H[n(\mathbf{r})] + E_{xc}[n(\mathbf{r})] + \int n(\mathbf{r})V(\mathbf{r})d\mathbf{r}. \tag{10}$$

The first term is the kinetic energy of *non-interacting* electrons:

$$T_s[n(\mathbf{r})] = -\frac{\hbar^2}{2m}2\sum_i \int \psi_i^*(\mathbf{r})\nabla^2\psi_i(\mathbf{r})d\mathbf{r}. \tag{11}$$

The second term (called the Hartree energy) contains the electrostatic interactions between clouds of charge:

$$E_H[n(\mathbf{r})] = \frac{e^2}{2}\int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}d\mathbf{r}d\mathbf{r}'. \tag{12}$$

The third term, called the *exchange-correlation energy*, contains all the remaining terms: our ignorance is hidden there. The logic behind such procedure is to subtract out easily computable terms which account for a large fraction of the total energy.

Using

$$\frac{\delta n(\mathbf{r})}{\delta\psi_i^*(\mathbf{r}')} = \psi_i(\mathbf{r})\delta(\mathbf{r} - \mathbf{r}') \tag{13}$$

and the formulae given in the appendix, one finds

$$\frac{\delta T_s}{\delta\psi_i^*(\mathbf{r})} = -\frac{\hbar^2}{2m}2\sum_i \nabla^2\psi_i(\mathbf{r}), \tag{14}$$

4

$$\frac{\delta E_H}{\delta \psi_i^*(\mathbf{r})} = e^2 \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \psi_i(\mathbf{r}) \tag{15}$$

and finally

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V_H(\mathbf{r}) + V_{xc}[n(\mathbf{r})] + V(\mathbf{r}) \right) \psi_i(\mathbf{r}) = \sum_j \lambda_{ij} \psi_j(\mathbf{r}) \tag{16}$$

where we have introduced a *Hartree potential*

$$V_H(\mathbf{r}) = e^2 \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \tag{17}$$

and an *exchange-correlation potential*

$$V_{xc}[n(\mathbf{r})] = \frac{\delta E_{xc}}{\delta n(\mathbf{r})}. \tag{18}$$

The Lagrange multiplier λ_{ij} are obtained by multiplying both sides of Eq.16 by $\psi_k^*(\mathbf{r})$ and integrating:

$$\lambda_{ik} = \int \psi_k^*(\mathbf{r}) \left(-\frac{\hbar^2}{2m}\nabla^2 + V_H(\mathbf{r}) + V_{xc}[n(\mathbf{r})] + V(\mathbf{r}) \right) \psi_i(\mathbf{r}) d\mathbf{r}. \tag{19}$$

For an insulator, whose states are either fully occupied or completely empty, it is always possible to make a subspace rotation in the space of ψ's (leaving the charge density invariant). We finally get the KS equations:

$$\left(H_{KS} - \epsilon_i \right) \psi_i(\mathbf{r}) = 0, \tag{20}$$

where $\lambda_{ij} = \delta_{ij}\epsilon_j$ and the operator H_{KS}, called KS Hamiltonian, is defined as

$$H_{KS} = -\frac{\hbar^2}{2m}\nabla^2 + V_H(\mathbf{r}) + V_{xc}(\mathbf{r}) + V(\mathbf{r}) \equiv -\frac{\hbar^2}{2m}\nabla^2 + V_{KS}(\mathbf{r}) \tag{21}$$

and is related to the functional derivative of the energy:

$$\frac{\delta E}{\delta \psi_i^*(\mathbf{r})} = H_{KS}\psi_i(\mathbf{r}). \tag{22}$$

1.4 DFT, Hartree-Fock, and Slater's exchange

The KS equations are somewhat reminiscent of the Hartree-Fock (HF) equations. Both are derived from a variational principle: the minimization of the energy functional for the latter, of the energy for a single Slater determinant wavefunction for the former. Both are self-consistent equations for one-electron wavefunctions. In the HF equations the *exchange* term appears in the place of the exchange-correlation potential of KS equations:

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V_H(\mathbf{r}) + V(\mathbf{r}) \right) \psi_i(\mathbf{r}) + e^2 \sum_{j,\|} \int \frac{\psi_j(\mathbf{r})\psi_j^*(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \psi_i(\mathbf{r}') d\mathbf{r}' = \epsilon_i \psi_i(\mathbf{r}) \tag{23}$$

where the sum over j extends only to states with parallel spins. Traditionally, one defines the *correlation energy* as the difference between the HF and the real energy. The name "exchange-correlation" in DFT reflects such tradition, although the exchange-correlation energy of DFT is not exactly the same as HF exchange plus correlation energy: in fact the former contains a contribution coming from the difference between the true many-body kinetic energy $\langle \Psi | T | \Psi \rangle$ and the kinetic energy $T_s[n(\mathbf{r})]$ of non-interacting electrons.

The exchange term in the HF equations is a *nonlocal* operator – one acting on a function ϕ as $(V\phi)(\mathbf{r}) = \int V(\mathbf{r}, \mathbf{r}')\phi(\mathbf{r}')d\mathbf{r}'$, and is quite difficult to compute. In earlier calculations, done with primitive computer machinery (or even *without* any computer machinery), an approximated form was often used. In the homogeneous electron gas, the average exchange energy and exchange potential for an electron are

$$\langle \epsilon_x \rangle = -\frac{3}{4}\frac{e^2 k_F}{\pi}, \qquad \langle v_x \rangle = -\frac{3}{2}\frac{e^2 k_F}{\pi} \tag{24}$$

5

where k_F is the Fermi wavevector: $k_F = (3\pi^2 n)^{1/3}$. In 1951 Slater proposed to replace the nonlocal exchange potential with the above form valid for the homogeneous electron gas, with k_F evaluated at the local density. This procedure yields a *local* (multiplicative) exchange potential

$$V_x(\mathbf{r}) = -\frac{3e^2}{2\pi} \left[3\pi^2 n(\mathbf{r}) \right]^{1/3}, \tag{25}$$

sometimes multiplied by coefficient α varying between 2/3 and 1 as an adjustable parameter. This approximation was rather popular in early solid-state physics but was never regarded as an especially good one (and it wasn't, actually).

1.5 Local Density Approximation for the exchange-correlation energy

We still don't have a reasonable estimate for the exchange-correlation energy $E_{xc}[n(\mathbf{r})]$. Kohn and Sham introduced, as early as 1965, the Local Density Approximation (LDA): they approximated the functional with a *function* of the local density $n(\mathbf{r})$:

$$E_{xc}[n(\mathbf{r})] = \int \epsilon_{xc}(n(\mathbf{r}))n(\mathbf{r})d\mathbf{r}, \qquad \frac{\delta E_{xc}}{\delta n(\mathbf{r})} \equiv \mu_{xc}(n(\mathbf{r})) = \left(\epsilon_{xc}(n) + n\frac{d\epsilon_{xc}(n)}{dn} \right)_{n=n(\mathbf{r})} \tag{26}$$

and for $\epsilon_{xc}(n(\mathbf{r}))$ used the same dependence on the density as for the homogeneous electron gas (also known as *jellium*) for which $n(\mathbf{r})$ is constant.

Even in such simple case the exact form of $\epsilon(n)$ is unknown (except at the HF level, see above). However, approximate forms have been known for a long time, going back to Wigner (1931). Numerical results from Monte-Carlo calculations (in principle exact) by Ceperley and Alder have been parameterized by Perdew and Zunger with a simple analytical form:

$$\begin{aligned} \epsilon_{xc}(n) = & \quad -0.4582/r_s - 0.1423/(1 + 1.0529\sqrt{r_s} + 0.3334 r_s) & , & \quad r_s \geq 1 \\ = & \quad -0.4582/r_s - 0.0480 + 0.0311\ln r_s - 0.0116 r_s + 0.0020 r_s \ln r_s & , & \quad r_s \leq 1 \end{aligned} \tag{27}$$

where r_s is the usual parameter appearing in the theory of metals: $r_s = (3/4\pi n)^{1/3}$, and atomic units are used ($e^2 = \hbar = m = 1$: lengths in Bohr radii, energies in Hartree=27.2 eV). Following HF tradition, the first term is called "exchange" (it has the same form as Slater's local approximation to exchange), the remaining terms "correlation". We note however that such distinction is to some extent arbitrary. Actually it has been shown that LDA contains a fair amount of error compensation between "exchange" and "correlation".

The Perdew-Zunger form for ϵ_{xc} is often used. Several other expressions have appeared in the literature. All forms yield very similar results in condensed-matter calculations, which is not surprising, since all parameterizations are very similar in the range of r_s applicable for solid-state phenomena.

1.6 Successes and failures of DFT

DFT, even in the simplest LDA approximation, turns out to be much more successful than expected. LDA is computationally much simpler than HF, yet it yields results of similar quality in atoms and molecules – highly inhomogeneous systems for which an approximation based on the homogeneous electron gas would hardly look appropriate. Structural and vibrational properties of solids are in general accurately described: the correct crystal structure is usually found to have the lowest energy; bond lengths, bulk moduli, phonon frequencies are accurate within a few percent.

LDA also has some well-known problems. Some can be avoided by using better functionals, some others have a deeper and more fundamental nature.

1.6.1 Accuracy of LDA, gradient corrections

The accuracy of LDA is often considered satisfactory in condensed-matter systems, but it is much less so in atomic and molecular physics, for which highly accurate experimental data are available. Also, LDA badly overestimates ($\sim 20\%$ and more) cohesive energies and bond strengths in molecules and solids, and as a consequence bond lengths are often underestimated.

Such problems are mostly corrected by the introduction of *gradient corrections*. The exchange-correlation functional is written as a function of the local density *and* of the local gradient of the density:

$$E_{xc}[n(\mathbf{r})] = \int \epsilon_{xc}(n(\mathbf{r}), |\nabla n(\mathbf{r})|)n(\mathbf{r})d\mathbf{r}. \tag{28}$$

Gradient-corrected functionals are the simplest extension – although not problem-free – of LDA to inhomogeneous systems one can think of. Their adoption has been quite slow, but finally they have found widespread acceptance. Gradient-corrected functionals yield much better atomic energies and binding energies than LDA, at a modest additional computational cost. In particular, they yield a good description of the Hydrogen bond, thus opening the way to calculations for systems, such as water, in which Hydrogen bonds play a crucial role (LDA is unusable for those systems: the Hydrogen bond in LDA is way too strong).

The weak van der Waals forces between closed-shells systems, that are responsible for the physisorption, are still beyond the reach of DFT. The van der Waals (or dispersive) interactions have a non-local character: they are due to charge fluctuations on one system, inducing a dipole on the other. This phenomenon is absent by construction from LDA, as well as from any functional based on the local density and on its local derivatives: these can reproduce only phenomena due to charge overlap. Since however LDA overestimates the attractive potential coming from the overlap of the tails of the charge density, closed-shell systems in LDA are actually bound with binding energies and binding distances in apparent agreement with experimental results. This is a fictitious result (and the dependence on the separation distance is wrong) that disappear if better-behaved gradient-corrected functionals are used.

1.6.2 Self-interaction

Self-interaction, i.e. the interaction of an electron with the field it generates, should cancel exactly, as it does in HF, by construction. The cancellation is actually quite good: in the H atom, the exchange-correlation and Hartree energy typically cancel within 5%, but they do not cancel exactly with current DFT approximations. Recipes for "Self-Interaction Correction" (SIC) exist, but their usefulness is limited or unclear.

The self-interaction affects finite systems, or systems containing localized electrons, while its effect is vanishing for delocalized electronic states in extended systems (solids). In finite systems the presence of self-interaction is reflected in an incorrect long-range behavior of the potential felt by an electron. For an atom, we should have $V_{xc}(r) \to -1/r$ for $r \to \infty$, but current functionals yield instead a potential that decays exponentially.

A remarkable example of the effects of self-interaction is the Al defect in SiO_2. In SiO_2, four O atoms are bound to a Si atom, forming a tetrahedron; each O is shared between two such tetrahedra. The strong bonding involves the sp^3 hybrid Si orbital and the $2p$ O orbitals. The presence of Al introduces a "hole" in the electronic structure. In DFT, one finds that the "hole" is evenly distributed on the Al-O bonds and that the local structure is still tetrahedral. Experiments show however that the hole is localized on one of the Al-O bonds, that is longer and weaker than the three others, thus yielding a local distorted structure around Al. The reason for this failure – affecting not only LDA but gradient corrections as well – has been tracked to self-interaction.

1.6.3 Strongly correlated electrons

Current DFT approximations often fail to correctly describe strongly correlated materials. These systems contain localized, atomic-like electronic states, typically originating from d or f atomic states, together with delocalized, band-like states originating from s and p states. The one-electron way of filling KS orbitals may break down: adding an electron to a localized state pushes the level up in energy, due to strong electron correlation (i.e. Coulomb repulsion between electrons). A famous example of failure is NiO: an insulator with a simple structure, that current DFT approximations consistently predict to be a metal. A similar behavior is displayed by K_4C_{60}, even if no d or f states are present: in this case the electron correlation on the C_{60} molecule is responsible.

1.6.4 The "band gap problem"

One would like very much to be able to calculate one-electron energies having the meaning of removal (or addition) energies, as for a non interacting system (in the language of many-body theory, *quasiparticle* energies). If one electron in the state v is removed from the system, $E_N - E_{N-1} = \epsilon_v$, where E_N is the energy of the system with N electrons. If one electron is added to the system in the state c, $E_{N+1} - E_N = \epsilon_c$. The difference between the largest addition energy and the smallest removal energy defines the energy band gap: $E_g = \epsilon_c - \epsilon_v = E_{N+1} + E_{N-1} - 2E_N$. In solids this is the onset of the continuum of optical transitions, if the gap is direct (if the lowest empty state and the highest filled state have the same $\mathbf{k}$ vector). From atomic and molecular physics, the highest occupied and lowest unoccupied states are respectively called HOMO (Highest Occupied Molecular Orbital) and LUMO (Lowest Unoccupied MO), while addition and removal energy are respectively called *electron affinity*, A, and *ionization potential*, I.

In HF the one-electron energies have the meaning of removal (or addition) energies for extended systems (Koopman's theorem). If the world were described by single Slater determinants, the difference between the LUMO and HOMO one-electron HF energies would yield the real energy gaps in solids (neglecting polarization effects, i.e. the change in the one-electron states upon addition or removal of an electron). Since the world is not well described by single Slater determinants, the band gap is usually quite overestimated in HF (with the true exchange potential, not Slater's local approximation).

In DFT, the one-electron energies have acquired a rather bad reputation, mostly due to the failure of KS band gaps (that is: calculated as the difference between LUMO and HOMO KS energies) to reproduce with an acceptable accuracy the true band gap in solids: gaps in DFT are strongly underestimated (up to 50%). One may say that after all DFT is a ground state theory, and that KS eigenvalues and eigenvectors are purely mathematical quantities without any physical meaning, but it wouldn't be a very satisfactory answer. It can be demonstrated that in *exact* DFT, $I = -\epsilon_{HOMO}$ holds. In finite systems ionization potentials and electron affinities can be calculated as energy differences between the ground state and a state with one electron added or removed. In extended systems this is of course not possible.

The reason for the infamous "band gap problem" lies in the dependence of the exact energy functional upon the number of electrons and in the inability of approximate functionals to reproduce it. In the next section we consider the extension of DFT to a fractionary number of electrons.

1.6.5 The discontinuity of exchange-correlation potential

The basic variational property of the density functional can be expressed by the stationary condition

$$\frac{\delta}{\delta n(\mathbf{r})} \left(E - \mu \left(\int n(\mathbf{r}) d\mathbf{r} - N \right) \right) = 0 \tag{29}$$

where μ is a Lagrange multiplier and N an integer number. The formulation of DFT can be extended to noninteger number of particles $N + \omega$ ($\omega > 0$) via the following definition:

$$E[n(\mathbf{r})] = F_{frac}[n(\mathbf{r})] + \int V(\mathbf{r}) n(\mathbf{r}) d\mathbf{r} \tag{30}$$

and

$$F_{frac}[n(\mathbf{r})] = \min \mathrm{tr}\{D(T+U)\}, \qquad D = (1-\omega)|\Psi_N\rangle\langle\Psi_N| + \omega|\Psi_{N+1}\rangle\langle\Psi_{N+1}| \tag{31}$$

where the minimum must be searched on all density matrices D that yield the prescribed density $n(\mathbf{r})$. It is easily verified that integration of $n(\mathbf{r})$ over all space yields $N + \omega$ electrons. With this definition the variational principle, Eq. 29, is defined for any number of electrons and yields the Euler equations

$$\frac{\delta E}{\delta n(\mathbf{r})} = \mu \tag{32}$$

and that μ is really the *chemical potential*: if we call E_N the energy at the ground state for N electrons, one has

$$\mu(N) = \frac{\partial E_N}{\partial N}. \tag{33}$$

There is an obvious problem if we consider $\mu(N)$ a continuous function of N for all values of N. Consider two neutral isolated atoms: in general, they will have two different values for μ. As a consequence the total energy of the two atoms will be lowered by a charge transfer from the atom at a higher chemical potential to the one at lower chemical potential.

In reality there is no paradox, because the E_N curve is not continuous. If we write down explicitly $E_{N+\omega}$, we find that both energy and minimizing charge density at fractionary number of electrons are simply a linear interpolation between the respective values at the end points with N and $N+1$ electrons:

$$E_{N+\omega} = (1-\omega)E_N + \omega E_{N+1}, \qquad n_{N+\omega}(\mathbf{r}) = (1-\omega)n_N(\mathbf{r}) + \omega n_{N+1}(\mathbf{r}) \qquad (34)$$

with obvious notations. The interesting and far-reaching consequence is that there is a discontinuity of the chemical potential $\mu(N)$ and of the functional derivative $\delta E/\delta n(\mathbf{r})$ at integer N. This is an important and essential characteristic of the exact energy functional that simply reflects the discontinuity of the energy spectrum.

Coming back to our paradox: for an atom with nuclear charge Z, ionization potential $I(Z)$ and electron affinity $A(Z)$ in the ground state,

$$\mu(N) = -I(Z) \qquad Z-1 < N < Z \qquad (35)$$
$$= -A(Z) \qquad Z < N < Z+1. \qquad (36)$$

For a system of two neutral atoms with nuclear charges X and Y, in which ω electrons are transferred from the first to the second atom:

$$\mu(\omega) = \mu(0) + I(Y) - A(X) \qquad -1 < \omega < 0 \qquad (37)$$
$$= \mu(0) + I(X) - A(Y) \qquad 0 < \omega < 1. \qquad (38)$$

Since the largest A (3.62 eV, for Cl) is still smaller than the smallest I (3.89 eV, for Cs), the neutral ground state is stable.

1.6.6 Band gaps and discontinuity of exchange-correlation potential

A consequence of the results of the previous section is that the true band gap of a solid, $E_g = I - A$, can be written as

$$E_g = -\mu(N-\delta) + \mu(N+\delta) = \left.\frac{\delta E}{\delta n(\mathbf{r})}\right|_{N+\delta} - \left.\frac{\delta E}{\delta n(\mathbf{r})}\right|_{N-\delta} \qquad (39)$$

with $\delta \to 0$.

Let us substitute to $E[n(\mathbf{r})]$ the explicit KS form, Eq.10. The Hartree and external potential terms of the functional will yield no discontinuity and no contribution to E_g. Only the kinetic and exchange-correlation terms may have a discontinuity and contribute to E_g.

For a non interacting system, only the kinetic term contributes, and the gap is exactly given by the KS gap:

$$E_g^{KS} = \left.\frac{\delta T_s}{\delta n(\mathbf{r})}\right|_{N+\delta} - \left.\frac{\delta T_s}{\delta n(\mathbf{r})}\right|_{N-\delta} = \epsilon_{LUMO} - \epsilon_{HOMO}. \qquad (40)$$

We remark that even the kinetic energy of non interacting electrons, considered as a functional of the density, must have a discontinuous derivative when crossing an integer number of electrons. This is one reason why it is so difficult to produce explicit functionals of the charge density for T_s that are able to yield good results: no simple functional form will yield the discontinuity, but this is needed in order to get the correct energy spectrum.

For the interacting system:

$$E_g = \left.\frac{\delta T_s}{\delta n(\mathbf{r})}\right|_{N+\delta} - \left.\frac{\delta T_s}{\delta n(\mathbf{r})}\right|_{N-\delta} + \left.\frac{\delta E_{xc}}{\delta n(\mathbf{r})}\right|_{N+\delta} - \left.\frac{\delta E_{xc}}{\delta n(\mathbf{r})}\right|_{N-\delta} = E_g^{KS} + E_g^{xc}. \qquad (41)$$

Note that the kinetic term is evaluated at the same charge density as for the non interacting system, so it coincides with the KS gap.

In conclusion: the KS gaps are not, by construction, equal to the true gap, because they are missing a term (E_g^{xc}) coming from the discontinuity of derivatives of the exchange-correlation functional. This is absent by construction from any current approximated functional (be it LDA or gradient-corrected or more complex). There is some evidence that this missing term is responsible for a large part of the band gap problem, at least in common semiconductors.

1.7 Adiabatic continuation formula and the exchange-correlation hole

The exchange-correlation energy can be recast into a form that sheds some light on the unexpected success of LDA and gives a possible path for the production of better functionals. One considers a system in which the Coulomb interaction between electrons is adiabatically switched on:

$$U_\lambda = \lambda \frac{e^2}{2} \sum_{i,j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} = \lambda U \tag{42}$$

where λ is a parameter that goes from $\lambda = 0$, for the noninteracting system, to $\lambda = 1$, for the true interacting system. The charge density is forced to remain equal to the charge density of the interacting system:

$$n_\lambda(\mathbf{r}) = n(\mathbf{r}), \tag{43}$$

while the potential V_λ will depend on λ. At $\lambda = 0$ the potential is nothing but the KS potential: and the energy functional at $\lambda = 0$ has the simple form:

$$E_0 = T_s[n(\mathbf{r})] + \int n(\mathbf{r}) V_{KS}(\mathbf{r}) d\mathbf{r}. \tag{44}$$

The following step is to write the energy functional for the true interacting system as an integral of the derivative with respect to λ:

$$E_1 = E_0 + \int_0^1 \frac{dE_\lambda}{d\lambda} d\lambda. \tag{45}$$

The derivative can be simply expressed using the Hellmann-Feynman theorem:

$$\frac{dE_\lambda}{d\lambda} = \langle \Psi_\lambda | \frac{\partial H}{\partial \lambda} | \Psi_\lambda \rangle \tag{46}$$

(see section on Hellmann-Feynman forces for the demonstration). Explicitly:

$$\frac{dE_\lambda}{d\lambda} = \langle \Psi_\lambda | U | \Psi_\lambda \rangle + \langle \Psi_\lambda | \frac{\partial V_\lambda}{\partial \lambda} | \Psi_\lambda \rangle. \tag{47}$$

By performing the integration, one finally finds

$$E_{xc} = \frac{1}{2} \int \frac{f_{xc}(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} n(\mathbf{r}) d\mathbf{r} d\mathbf{r}' \tag{48}$$

where $f_{xc}(\mathbf{r}, \mathbf{r}')$ is the *exchange-correlation hole*: the charge missing around a point $\mathbf{r}$ due to exchange effects (Pauli antisymmetry) and to Coulomb repulsion. The exchange-correlation hole obeys the sum rule

$$\int f_{xc}(\mathbf{r}, \mathbf{r}') d\mathbf{r}' = -1. \tag{49}$$

The exchange-correlation hole is related to the pair correlation function $g(\mathbf{r}, \mathbf{r}')$, giving the probability to find an electron in $\mathbf{r}'$ if there is already one in $\mathbf{r}$. Its exact definition is:

$$f_{xc}(\mathbf{r}, \mathbf{r}') = n(\mathbf{r}') \int_0^1 (g_\lambda(\mathbf{r}, \mathbf{r}') - 1) \, d\lambda \tag{50}$$

where $g_\lambda(\mathbf{r}, \mathbf{r}')$ is the pair correlation function the system having the electron-electron interaction multiplied by λ, Eq.(42). In homogeneous systems $f_{xc}(\mathbf{r}, \mathbf{r}')$ and $g(\mathbf{r}, \mathbf{r}')$ are well known and studied functions. It has been shown that in inhomogeneous systems LDA does not give a good

approximation for $f_{xc}(\mathbf{r}, \mathbf{r}')$. However LDA yields a very good approximation for its spherical part $f_{xc}(\mathbf{r}, s)$:

$$\widetilde{f}_{xc}(\mathbf{r}, s) = \int f_{xc}(\mathbf{r}, \mathbf{r} + s\hat{r}) \frac{d\hat{r}}{4\pi}. \tag{51}$$

It is easily shown the Eq.48 depends only on the spherical part of the exchange-correlation hole:

$$E_{xc} = \frac{1}{2} \int \frac{\widetilde{f}_{xc}(\mathbf{r}, s)}{s} n(\mathbf{r}) d\mathbf{r} ds. \tag{52}$$

This explains at least partially the good performances of LDA. The above procedure is a good starting point in the search for better functional, via better modeling of the exchange-correlation hole.

1.8   The exact exchange-correlation potential from many-body theory

Many-body perturbation theory yields the following exact solution for the many-body problem:

$$\left(-\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{r}) + V_H(\mathbf{r}) - \epsilon_i \right) \psi_i(\mathbf{r}) + \int \Sigma(\mathbf{r}, \mathbf{r}', \epsilon_i) \psi_i(\mathbf{r}') d\mathbf{r}' = 0 \tag{53}$$

where the *self-energy* $\Sigma(\mathbf{r}, \mathbf{r}', \epsilon)$ is a complex, nonlocal, energy-dependent operator, the $\psi_i(\mathbf{r})$ and ϵ_i have the physical meaning of quasiparticle states and energies. The energies ϵ_i are also complex and their imaginary part is related to the lifetime of the state.

Both DFT and many-body perturbation theory are exact on the ground state (and the latter also on excited states). This implies

$$n(\mathbf{r}) = \int \mathrm{Im} G_{\mathrm{DFT}}(\mathbf{r}, \mathbf{r}, \epsilon) d\epsilon = \int \mathrm{Im} G(\mathbf{r}, \mathbf{r}, \epsilon) d\epsilon \tag{54}$$

where $G(\mathbf{r}, \mathbf{r}', \epsilon)$ is the Green's function of the system, $G_{\mathrm{DFT}}(\mathbf{r}, \mathbf{r}', \epsilon)$ is the same in DFT, and the integration extends to the energies of occupied states. The Dyson equation must also apply between G and G_{DFT}:

$$G = G_{\mathrm{DFT}} + G_{\mathrm{DFT}} \left(\Sigma - V_{xc} \right) G. \tag{55}$$

By combining the above equations, one finally gets the following result:

$$\mathrm{Im} \int \left[G_{\mathrm{DFT}} \left(\Sigma - V_{xc} \right) G \right]_{\mathbf{r}=\mathbf{r}'} = 0. \tag{56}$$

This equation can be used to deduce the exact exchange-correlation potential. Practical many-body perturbation theory calculations are very difficult but not impossible. Some test calculations on simple systems have shown that the LDA V_{xc} is a good approximation to the true V_{xc}.

2 Practical DFT calculations

2.1 Atoms

Atomic DFT calculations are usually performed assuming a spherically averaged charge density. For closed-shell atoms, such procedure does not introduce any approximation, while for open-shell atoms, it introduces an error that turns out to be quite small (it can be accounted for using perturbation theory if a higher accuracy is desired). Under such assumption, an atom can be described as in elementary Quantum Mechanics by an *electronic configuration* $1s^2 2s^2 2p^6$...: the KS equation has spherical symmetry and is separable into a radial equation and an angular part (whose solutions are the spherical harmonics). The solution of the KS equations for an atom proceeds as follows. For a given electronic configuration, and starting from some initial guess of the KS potential,

- the radial KS equations are solved for those radial orbitals that correspond to occupied states;

- the (spherically averaged) charge density is recalculated;

- a new KS potential is calculated from the charge density, and the procedure is iterated until self-consistency is reached.

The minimum energy is obtained for the ground state electronic configuration, that is well known for all atoms.

The solution of the radial KS equation (step 1 above) is typically done by numerical integration on a grid, using any of the many well-known techniques that have been developed for one-dimensional differential equations.

The iteration to self-consistency (step 3) is done using the methods explained in Sec. "Iteration to self-consistency".

One may wonder why we fix the electronic configuration instead of filling the one-electron state starting from the lowest energies and up. For many atoms there is no difference between the two approaches. Atoms with incomplete d and f states however present a problem. The incomplete d and f shells may have KS energies that are lower than those of outer s and p states; if however we try to move one more electron from s and p states into the d or f shell, the KS level is "pushed up" by strong Coulomb repulsion between highly localized electrons. This is a manifestation of *strong correlation* that is responsible for a wealth of interesting phenomena (such as magnetism). Currently available functionals are unable to reproduce this behavior and may produce an incorrect occupancy of state if this is assigned in "the one-electron way". Fixing the electronic configuration solves the problem (unfortunately only in atoms) by imposing the correct occupancy of the highly localized (correlated) d and f states.

2.2 Molecules

In molecules, KS equations are usually solved by expanding KS orbitals into some suitable basis set. Methods of solutions based on the discretization of the problem on a 3-d grid have also been proposed, though. Localized basis sets (atomic-like wavefunctions centered on atoms) are often used, especially in Quantum Chemistry. The most common basis sets are Linear Combinations of Atomic Orbitals (LCAO), Gaussian-type Orbitals (GTO), Slater-type Orbitals (STO). These atomic-like functions are tailored for fast convergence, so that only a few (some tens at most) functions per atom are needed. An impressive body of technique has been developed during the years on the use of localized basis sets.

Localized orbitals are quite delicate to use. One problem is the difficult to check systematically for convergence. Another problem is the difficulty of calculating the Hellmann-Feynman forces acting on atoms, due to the presence of *Pulay forces* (see later). In the following we will concentrate on the opposite approach, that is, choosing extended, atomic-independent Plane Waves (PW) as basis set.

2.3 Extended systems: unit cells and supercells

The atomic arrangement in perfect crystals is described by a periodically repeated *unit cell*. For many interesting physical systems, however, perfect periodicity is absent, but the system is either

approximately periodic or periodic in one or two directions or periodic except for a small part. Examples of such systems include surfaces, point defects in crystals, substitutional alloys, heterostructures ("superlattices" and quantum wells). In all such cases it is convenient to simulate the system with a periodically repeated fictitious *supercell*. The form and the size of the supercell depend on the physical system being studied. The study of point defects requires that a defect does not interact with its periodic replica in order to accurately simulate a truly isolated defect. For disordered solids, the supercell must be large enough to guarantee a significant sampling of the configuration space. For surfaces, one uses a crystal slab alternated with a slab of empty space, both large enough to ensure that the bulk behavior is recovered inside the crystal slab and that the surface behavior is unaffected by the presence of the periodic replica of the crystal slab. In the examples mentioned above, the supercell approach is usually more convenient than the "cluster approach", that is, simulating an extended system by taking a finite piece of material (the more traditional approach in Quantum Chemistry). The reason is the absence of an abrupt termination in the supercell approach.

Even finite systems (molecules, clusters) can be studied using supercells. Enough empty space between the periodic replicas of the finite system must be left so that the interactions between them are weak. The use of supercells for the simulation of molecular or completely aperiodic systems (liquids, amorphous systems) has become quite common in recent years, in connection with first-principles simulations (especially molecular dynamics simulations) using a PW basis set. In fact there are important computational advantages in the use of PW's that may offset the disadvantage of inventing a periodicity where there is none.

The size of the unit cell – the number of atoms and the volume – is very important. Together with the type of atoms it determines the difficulty of the calculation: large unit cells mean large calculations. Unfortunately many interesting physical systems are described – exactly or approximately – by large unit cells.

2.4 Plane wave basis set

In the following we will assume that our system is a crystal with lattice vectors $\mathbf{R}$ and reciprocal lattice vectors $\mathbf{G}$. It is not relevant whether the cell is a real unit cell of a real periodic crystal or if it is a supercell describing an aperiodic system. The KS wavefunctions are classified by a band index and a Bloch vector $\mathbf{k}$ in the Brillouin Zone (BZ).

A PW basis set is defined as

$$\langle \mathbf{r} | \mathbf{k} + \mathbf{G} \rangle = \frac{1}{V} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}, \quad \frac{\hbar^2}{2m} |\mathbf{k}+\mathbf{G}|^2 \leq E_{cut}, \tag{57}$$

where V is the crystal volume, E_{cut} is a cutoff on the kinetic energy of PW's (from now on, simply "the cutoff"). PW's have many attractive features: they are simple to use (matrix elements of the Hamiltonian have a very simple form), orthonormal by construction, unbiased (there is no freedom in choosing PW's: the basis is fixed by the crystal structure and by the cutoff) and it is very simple to check for convergence (by increasing the cutoff).

Unfortunately the extended character of PW's makes it very difficult to accurately reproduce localized functions such as the charge density around a nucleus or even worse, the orthogonalization wiggles of inner (core) states. In order to describe features which vary on a length scale δ, one needs Fourier components up to $q \sim 2\pi/\delta$. In a solid, this means $\sim 4\pi(2\pi/\delta)^3/3\Omega$ PW's (where Ω is the dimension of the BZ). A simple estimate for diamond is instructive. The $1s$ wavefunction of the carbon atom has its maximum around 0.3 a.u., so $\delta \simeq 0.1$ a.u. is a reasonable value. Diamond has an fcc lattice ($\Omega = (2\pi)^3/(a_0^3/4)$) with lattice parameter $a_0 = 6.74$ a.u., thus yielding $\sim 250,000$ PW's. This is clearly too much for practical use.

2.5 Pseudopotentials

Core states prevent the use of PW's. However they do not contribute in a significant manner to chemical bonding and to solid-state properties. Only outer (valence) electrons do, while core electron are "frozen" in their atomic state. This suggests that one can safely ignore changes in core states (*frozen core approximation*). However the soundness of this approach was challenged by a 1976 paper by Janak, showing that large variations in the energy of core states can be induced by

changes in the chemical environment. The controversy was solved in 1980 by Von Barth and Gelatt. Their argument is briefly sketched here. Let us introduce the notations n_c and n_v for the true selfconsistent core and valence charge; n_c^0 and n_v^* for the frozen-core charge and the corresponding valence charge. A "frozen-core functional" $E[n_c, n_v]$ is introduced. The frozen-core error is

$$\delta = E[n_c^0, n_v^*] - E[n_c, n_v]. \tag{58}$$

By expanding around n_c and n_v one finds

$$\delta \simeq \int \frac{\delta E}{\delta n_c}(n_c^0 - n_c)d\mathbf{r} + \int \frac{\delta E}{\delta n_v}(n_v^* - n_c)d\mathbf{r} + \text{2nd order terms.} \tag{59}$$

The important point is that the following stationary conditions hold:

$$\frac{\delta E}{\delta n_c} = \mu_c, \qquad \frac{\delta E}{\delta n_v} = \mu_v \tag{60}$$

where μ_c and μ_v are constants, so that the first-order terms in the error vanish.

The idea of replacing the full atom with a much simpler pseudoatom with valence electrons only arises naturally (apparently in a 1934 paper by Fermi for the first time). *Pseudopotentials* (PP's) have been widely used in solid state physics starting from the 1960's. In earlier approaches PP's were devised to reproduce some known experimental solid-state or atomic properties such as energy gaps or ionization potentials. Other types of PP's were obtained from band structure calculations with the OPW (orthogonalized PW) basis set, by separating the smooth (PW) part from the orthogonalization part in the wavefunctions.

Modern PP's are called *norm-conserving*. These are *atomic* potentials which are devised so as to mimic the scattering properties of the true atom. For a given reference atomic configuration, a norm-conserving PP must fulfill the following condition:
1) all-electron and pseudo-wavefunctions must have the same energy, and
2) they must be the same beyond a given "core radius" r_c, which is usually located around the outermost maximum of the atomic wavefunction;
3) the pseudo-charge and the true charge contained in the region $r < r_c$ must be the same.

This last condition explains the name norm-conserving. There is an historical reason for this: some earlier PP's violated condition 3 (this was known as the "orthogonality hole" problem). Note that the definition "all-electron", here and in the following, refers to a KS calculation that includes core electrons, not to a many-electron wavefunctions.

Norm-conserving PP are relatively smooth functions, whose long-range tail goes like $-Z_v e^2/r$ where Z_v is the number of valence electrons. They are *nonlocal* because it is usually impossible to mimic the effect of orthogonalization to core states on different angular momenta l with a single function. There is a PP for every l:

$$\widehat{V}^{ps} = V_{loc}(r) + \sum_l V_l(r)\widehat{P}_l = V_{loc}(r) + \sum_{lm} Y_{lm}(\mathbf{r})V_l(r)\delta(r - r')Y_{lm}^*(\mathbf{r}'), \tag{61}$$

where $V_{loc}(r) \simeq -Z_v e^2/r$ for large r and $\widehat{P}_l = |l\rangle\langle l|$ is the projection operator on states of angular momentum l. They are however seldom used in this form. For computational reasons, they are recast into a *separable* form (see appendix). The nonlocality of PP's introduces some additional but limited complications in the calculation. In particular, one has to do the following generalization:

$$\int V(\mathbf{r})n(\mathbf{r})d\mathbf{r} \longrightarrow \sum_i \langle \psi_i|\widehat{V}|\psi_i \rangle = \sum_i \int \psi_i^*(\mathbf{r})V(\mathbf{r}, \mathbf{r}')\psi_i(\mathbf{r}')d\mathbf{r}d\mathbf{r}'. \tag{62}$$

Experience has shown that PP's are practically equivalent to the frozen core approximation: PP and all-electron calculations on the same systems yield almost indistinguishable results (except for those cases in which core states are not sufficiently frozen). It should be remarked that the use of PP's is not limited to PW basis sets: PP's can be used in conjunction with localized basis sets as well.

2.6   Another way of looking at pseudopotentials

Norm-conserving PP's are still "hard" – that is, they contain a significant amount of Fourier components with large q – for a number of atoms, such as N, O, F, and the first row of transition metals. For these atoms little is gained in the pseudization, because there are no orthonormality wiggles that can be removed in the $2p$ and $3d$ states, respectively. More complex *Ultrasoft* PP's have been devised that are much softer than ordinary norm-conserving PP's, at the price of a considerable additional complexity.

The heavy formalism of ultrasoft PP's tends to hide the underlying logic (and physics). An alternative approach, called Projector Augmented Waves (PAW), is much more transparent. Moreover PAW includes as special cases a number of other methods and provides a simple and consistent way to reconstruct all-electron wavefunctions from pseudo-wavefunctions. These are needed for reliable calculation of a number of observables, such as NMR chemical shifts and hyperfine coupling coefficients.

The idea of PAW is to find a mapping between the complete wavefunction and the pseudo-wavefunction via a suitable linear operator. The pseudo-wavefunction must be a smooth object that can be expanded into PW's.

Let us consider for simplicity the case of a single atom in the system. In a region R centered around the atom, the mapping is defined as

$$|\tilde{\phi}_l\rangle = (1 + \mathcal{T})|\phi_l\rangle \tag{63}$$

where the functions $\tilde{\phi}_l$ are solutions, regular at the origin but not necessarily bound, of the all-electron atomic KS equation; the functions ϕ_l are corresponding pseudo-functions, that are much smoother in the region R and join smoothly to the $\tilde{\phi}_l$ at the border of region R. Outside the region R, we set $\mathcal{T} = 0$.

In the region R, we assume that we may write a pseudo-wavefunction ψ for our molecular or solid-state system as a sum over the atomic pseudo-waves ϕ_l:

$$|\psi\rangle = \sum_l c_l |\phi_l\rangle \tag{64}$$

By applying the operator $(1 + \mathcal{T})$ to both sides of the above expansion we find

$$|\tilde{\psi}\rangle = \sum_l c_l |\tilde{\phi}_l\rangle \tag{65}$$

where $\tilde{\psi}$ is the all-electron wavefunction. The above result can be recast into the form

$$|\tilde{\psi}\rangle = |\psi\rangle + \sum c_l \left(|\tilde{\phi}_l\rangle - |\phi_l\rangle \right). \tag{66}$$

It remains to define the c_l coefficients. Let us introduce the projectors β_l with the following properties:

$$\langle \beta_l | \phi_m \rangle = \delta_{lm}, \qquad \sum_l |\phi_l\rangle\langle\beta_l| = I. \tag{67}$$

It is easy to verify that $c_l = \langle \beta_l | \psi \rangle$ and that we can write

$$|\tilde{\psi}\rangle = |\psi\rangle + \sum_l \langle\beta_l|\psi\rangle \left(|\tilde{\phi}_l\rangle - |\phi_l\rangle \right) \tag{68}$$

$$= \left[I + \sum_l \left(|\tilde{\phi}_l\rangle - |\phi_l\rangle \right) \langle\beta_l| \right] |\psi\rangle. \tag{69}$$

The quantity between square brackets is our $1 + \mathcal{T}$ operator. This replaces the pseudo-states ϕ from the pseudo-wavefunctions around the atoms and replaces them with the all-electron states $\tilde{\phi}$. The $1 + \mathcal{T}$ operator is a purely atomic quantity that is obtained from a judicious choice of the $\tilde{\phi}_l$ all-electron atomic states, the corresponding pseudo-states ϕ_l, and the projectors β_l.

The equations to solve in the PAW method are then obtained by inserting the above form for $\tilde{\psi}$ in the energy functional and by finding its minimum with respect to the variation of the smooth part

only, ψ. Rather cumbersome expressions results. An important feature of the resulting equations is that the charge density is no longer given simply by the square of the orbitals, but it contains in general an additional (*augmentation*) term:

$$n(\mathbf{r}) = \sum_i |\psi_i(\mathbf{r})|^2 + \sum_i \sum_{lm} \langle \psi_i | \beta_l \rangle q_{lm}(\mathbf{r}) \langle \beta_m | \psi_i \rangle \qquad (70)$$

where

$$q_{lm}(\mathbf{r}) = \tilde{\phi}_l(\mathbf{r})\tilde{\phi}_m(\mathbf{r}) - \phi_l(\mathbf{r})\phi_m(\mathbf{r}) \qquad (71)$$

(using the completeness relation, Eq.(67)). Conversely the pseudo-wavefunctions are no longer orthonormal, but obey instead a generalized orthonormality relation:

$$\langle \psi_i | S | \psi_j \rangle = \delta_{ij}, \qquad S = I + \sum_{lm} |\beta_l\rangle Q_{lm} \langle \beta_m| \qquad Q_{lm} = \int_R q_{lm}(\mathbf{r})d\mathbf{r}. \qquad (72)$$

Ultrasoft PP's can be derived from PAW assuming a pseudized form for $q_{lm}(\mathbf{r})$. Norm-conserving PP's in the separable form can be derived from PAW if the atomic states $\tilde{\phi}_l$ and ϕ_l obey the norm-conservation rule (thus $S = 1$). The LAPW method can also be recast under this form. The careful reader will also remark some similarity between the PAW approach and the venerable PP's based on the OPW method (those with the infamous "orthogonality hole": PAW plugs the hole by defining the charge density in the correct way).

2.7 Brillouin-Zone sampling

In order to calculate the charge density $n(\mathbf{r})$ in a periodic system one has to sum over an infinite number of $\mathbf{k}$-points:

$$n(\mathbf{r}) = \sum_{\mathbf{k}} \sum_i |\psi_{\mathbf{k},i}(\mathbf{r})|^2 \qquad (73)$$

where the index i runs over occupied bands. Assuming periodic (Born-Von Kàrmàn) boundary conditions

$$\psi(\mathbf{r} + L_1 \mathbf{R}_1) = \psi(\mathbf{r} + L_2 \mathbf{R}_2) = \psi(\mathbf{r} + L_3 \mathbf{R}_3) = \psi(\mathbf{r}), \qquad (74)$$

a crystal has $L = L_1 L_2 L_3$ allowed $\mathbf{k}$-points (L is also the number of unit cells). In the "thermodynamic" limit of an infinite crystal, $L \rightarrow \infty$, the discrete sum over $\mathbf{k}$ becomes an integral over the BZ.

Experience shows that this integral can be approximated by a discrete sum over an affordable number of $\mathbf{k}$-points, at least in insulators and semiconductors. When present, symmetry can be used to further reduce the number of calculations to be performed. Only one $\mathbf{k}$-point is left to represent each *star* – the set of $\mathbf{k}$-points that are equivalent by symmetry – with a weight w_i that is proportional to the number of $\mathbf{k}$-points in the star. The infinite sum over the BZ is replaced by a discrete sum over a set of points $\{\mathbf{k}_i\}$ and weights w_i:

$$\frac{1}{L} \sum_{\mathbf{k}} f_{\mathbf{k}}(\mathbf{r}) \longrightarrow \sum_i w_i f_{\mathbf{k}_i}(\mathbf{r}). \qquad (75)$$

The resulting sum is then symmetrized to get the charge density.

Suitable sets for BZ sampling in insulators and semiconductors are called "special points". This name is somewhat misleading: in most cases those sets just form uniform grids in the BZ.

In metals things are more difficult because one needs an accurate sampling of the Fermi surface. A suitable extension of DFT to fractionary occupation numbers is needed. The *Gaussian broadening* and the *tetrahedron* techniques, or variations of the above, are generally used.

In supercells, the $\mathbf{k}$-point grid is often limited to the Γ point ($\mathbf{k} = 0$). A better sampling may be needed only if it is important to accurately describe the band structure of a subjacent crystal structure. This is the case of point defects in solids and of surfaces. If, on the contrary, supercells are used to simulate completely aperiodic or finite systems, the Γ point is the good choice: a better $\mathbf{k}$-point grid would better account for the periodicity of the system, but this is fictitious anyway.

3 Finding the electronic ground state

There are two possible ways to find the electronic ground state, for fixed atomic positions. The first is to solve self-consistently the KS equations, by diagonalizing the Hamiltonian matrix and iterating on the charge density (or the potential) until self-consistency is achieved. The second is to directly minimize the energy functional as a function of the coefficients of KS orbitals in the PW (or other) basis set, under the constraint of orthonormality for KS orbitals. The basic ingredients are in both cases the same.

3.1 Iteration to self-consistency

In the following I will consider the charge density as the quantity to be determined self-consistently, but similar considerations apply to the self-consistent potential V_{KS} as well.

We supply an input charge density $n_{in}(\mathbf{r})$ to the KS equations and we get an output charge density $n_{out}(\mathbf{r})$. This defines a functional A:

$$n_{out}(\mathbf{r}) = A[n_{in}(\mathbf{r})]. \tag{76}$$

At self-consistency,

$$n(\mathbf{r}) = A[n(\mathbf{r})]. \tag{77}$$

The first algorithm that comes to the mind is to simply use $n^{out}(\mathbf{r})$ as the new input charge density:

$$n_{in}^{(i+1)} = n_{out}^{(i)}, \tag{78}$$

where the superscripts indicate the iteration number. Unfortunately there is no guarantee that this will work, and experience shows that it usually does not. The reason is that the algorithm will work only if the error on output is smaller than the error on input. If you have an error $\delta n^{in}(\mathbf{r})$ on input, the error on output, close to self-consistency, will be

$$\delta n_{out}(\mathbf{r}) \simeq \int \frac{\delta A}{\delta n(\mathbf{r})} \delta n_{in}(\mathbf{r}) d\mathbf{r} \equiv J \delta n_{in} \tag{79}$$

which may or may not be smaller than the input error: it depends on the size of the largest eigenvalue, e_J, of the operator J, which is related to the dielectric response of the system. Usually, $e_J > 1$ and the iteration does not converge.

A simple algorithm that generally works, although sometimes slowly, is the "simple mixing". A new input charge density is generated by mixing the input and output charges:

$$n_{in}^{(i+1)} = (1 - \alpha)n_{in}^{(i)} + \alpha n_{out}^{(i)} \tag{80}$$

The value of α must be chosen empirically in order to get fast convergence. The error with respect to self-consistency becomes

$$\delta n_{out} = [(1 - \alpha) + \alpha J] \delta n_{in} \tag{81}$$

and it is easily seen that the iteration converges if $\alpha < |1/e_J|$. In general, the convergence is easier for small cells and symmetric systems, more difficult for larger cells, low symmetry, cells elongated along one directions, surfaces. Relatively big values ($\alpha = 0.3 - 0.5$) can be chosen in "easy" systems, smaller values are appropriate for cases of difficult convergence.

Better results are obtained with more sophisticated algorithms (to name a few: Anderson, Broyden, Direct Iteration in Inverse Space, DIIS) that use informations collected from several preceding iterations. Let us sketch the logic of such algorithms. We have a sequence of $n_{in}^{(i)}$ producing $n_{out}^{(i)}$ from preceding iterations. We look for the linear combination of input n_{in}^{new}:

$$n_{in}^{new} = \sum_l c_l n_{in}^{(l)}, \qquad \sum_l c_l = 1 \tag{82}$$

that minimizes an appropriate norm $||n_{in}^{new} - n_{out}^{new}||$. Close to self-consistency,

$$||n_{in}^{new} - n_{out}^{new}|| \simeq ||\sum_l c_l(n_{in}^{(l)} - n_{out}^{(l)})|| \tag{83}$$

and the coefficients c_l are determined by imposing that such norm is minimum. Then we mix n_{in}^{new} with $n_{out}^{new} = \sum_l c_l n_{out}^{(l)}$ (using simple mixing or whatever algorithm is appropriate).

3.2 Diagonalization of the Hamiltonian

When the wavefunctions are expanded on a finite basis set the KS equations take the form of a secular equation:

$$\sum_{\mathbf{G}'} H(\mathbf{k}+\mathbf{G}, \mathbf{k}+\mathbf{G}')\psi_{\mathbf{k},i}(\mathbf{G}') = \epsilon_{\mathbf{k},i}\psi_{\mathbf{k},i}(\mathbf{G}),\tag{84}$$

where the matrix elements of the Hamiltonian have the form

$$H(\mathbf{k}+\mathbf{G}, \mathbf{k}+\mathbf{G}') = \frac{\hbar^2}{2m}(\mathbf{k}+\mathbf{G})^2\delta_{\mathbf{G},\mathbf{G}'} + V_{scf}(\mathbf{G}-\mathbf{G}') + V_{loc}(\mathbf{G}-\mathbf{G}') + V_{NL}(\mathbf{k}+\mathbf{G}, \mathbf{k}+\mathbf{G}').\tag{85}$$

The term $V_{scf}(\mathbf{G}-\mathbf{G}')$ is the Fourier transform of the the screening potential:

$$V_{scf}(\mathbf{G}-\mathbf{G}') = \frac{1}{V}\int V_{scf}(\mathbf{r})e^{i(\mathbf{G}-\mathbf{G}')\mathbf{r}}d\mathbf{r}.\tag{86}$$

(V is the volume of the crystal: the integration extends over the entire crystal) and the same applies to V_{loc} that comes from the local term in the PP's. The nonlocal contribution V_{NL} comes from the nonlocal part of the PP's:

$$V_{NL}(\mathbf{k}+\mathbf{G}, \mathbf{k}+\mathbf{G}') = \frac{1}{V}\int V_{NL}(\mathbf{r},\mathbf{r}')e^{-i(\mathbf{k}+\mathbf{G})\mathbf{r}'}e^{i(\mathbf{k}+\mathbf{G}')\mathbf{r}'}d\mathbf{r}d\mathbf{r}'.\tag{87}$$

The problem is reduced in this way to the well-known problem of finding the lowest eigenvalues and eigenvectors (only the valence states for insulators, a few more for metals) of an $N_{pw} \times N_{pw}$ Hermitian matrix (where N_{pw} is the number of PW's). This task can be performed with well-known bisection-tridiagonalization algorithms, for which very good public-domain computer packages (for instance, LAPACK) exist. Unfortunately this straightforward procedure has serious limitations. In fact:

i) the computer time required to diagonalize a $N_{pw} \times N_{pw}$ matrix grows as N_{pw}^3;

ii) the matrix must be stored in memory, requiring $\mathcal{O}(N_{pw}^2)$ memory.

As a consequence a calculation requiring more than a few hundred PW's becomes exceedingly time- and memory-consuming. As the number of PW's increases with the size of the unit cell it is very hard to study systems containing more than a few (say 5-10) atoms. Both limitations can be pushed much further using *iterative techniques* (see Appendix).

3.3 Direct minimization

It is not necessary to go through KS equations and self-consistency to find the electronic ground state. The energy functional can be written as a function of the coefficients in the basis set of the KS orbitals and directly minimized, under the usual orthonormality constraints. One has to find the minimum of

$$E'(\psi_{\mathbf{k},i}(\mathbf{G})) = E(\psi_{\mathbf{k},i}(\mathbf{G})) - \sum_{ij}\lambda_{ij}\left(\sum_{\mathbf{G}}\psi^*_{\mathbf{k},i}(\mathbf{G})\psi_{\mathbf{k},j}(\mathbf{G}) - \delta_{ij}\right),\tag{88}$$

with respect to the variables $\psi_{\mathbf{k},i}(\mathbf{G})$ and the Lagrange multipliers λ_{ij}. The problem is made much simpler by the knowledge of the gradients of the function to be minimized. In fact, remembering Eq.22, one easily finds

$$\frac{\partial E'}{\partial\psi_{\mathbf{k},i}(\mathbf{G})} = H(\mathbf{G},\mathbf{G}')\psi^*_{\mathbf{k},i}(\mathbf{G}') - \sum_{ij}\lambda_{ij}\psi^*_{\mathbf{k},j}(\mathbf{G}).\tag{89}$$

Note that, as in iterative diagonalization, the basic ingredients are $H\psi$ products. Note also that the Hamiltonian depends on the variables $\psi_{\mathbf{k},i}(\mathbf{G})$ through V_{scf} and the charge density.

The problem of minimizing a function of many variables whose gradients are known, with the additional complication due to the presence of constraints, can be solved using appropriate extensions to textbook algorithms, or specialized algorithms, such as steepest descent (bad) or conjugate gradient (better) or DIIS (even better).

4 Moving atoms - complex materials

Until now we have assumed that the atomic positions were known and fixed. This is the case for simple crystals (silicon for instance), but in more complex crystals (for instance, SiO_2) the equilibrium positions are not fixed by symmetry. In even more complex materials we simply don't know the equilibrium atomic positions and would like to calculate them.

In the following we assume that ions are classical objects. At zero temperature the equilibrium atomic positions $\mathbf{R}_i$, $i = 1, ..., N$ (N = number of atoms in the unit cell) are determined by the minimum of the *total energy* E_{tot} of the system, that is, the sum of the electronic (DFT) energy E and of the ion-ion interaction (electrostatic) energy E_{II}. If we consider the electrons in their ground state for any given configuration of $\mathbf{R}_i$ (collectively indicated by $\{\mathbf{R}\}$), the total energy will be a function of the atomic positions:

$$E_{tot}(\{\mathbf{R}\}) = E(\{\mathbf{R}\}) + E_{II}(\{\mathbf{R}\}). \tag{90}$$

The procedure to find the atomic configuration yielding the minimum energy is usually called *structural optimization* or *relaxation.*

For an infinite system we must distinguish between atomic displacements that change the form and volume of the unit cell (related to *elastic* modes) and atomic displacements internal to the unit cell (related to *phonon* modes). Such distinction does not exist for a finite system. The optimization of the lattice and that of atomic positions have to be done separately, or in any case, using different procedures (Unless we use *variable-cell molecular dynamics*, a very powerful but very complex technique).

4.1 Optimization of lattice parameters

The determination of the equilibrium lattice parameters and of the relative stability of different structures for simple semiconductors was one of the first remarkable applications of the LDA PW-PP approach (around 1980). The total energy is calculated as a function of the volume V of the unit cell for various different candidate structures. The lowest-energy structure will be the equilibrium structure at zero temperature and at zero pressure.

The $E(V)$ curve can in principle be directly calculated. However it is much more convenient to fit an equation of state to a few calculated points. Empirical equations of state depending on a few parameters and covering a wide range of volumes around the equilibrium are well known and widely used in geology and geophysics. The most famous is possibly the Murnaghan equation of state:

$$P(V) = \frac{B}{B'} \left[\left(\frac{V_0}{V} \right)^{B'} - 1 \right] \tag{91}$$

where the fit parameters are the equilibrium volume V_0, the bulk modulus B:

$$B = -V \frac{\partial P}{\partial V} = V \frac{\partial^2 E}{\partial V^2} \tag{92}$$

and its derivative with respect to the pressure, $B' = dB/dP$, an adimensional quantity ranging from 3 to 10 for almost all solids. The Murnaghan $E(V)$ is obtained by integrating the former expression. All these quantities are directly comparable to experimental results (at zero temperature).

The reason for this fit procedure is that the straightforward calculation of $E(V)$ suffers from important errors. In particular, when using PW's with a given energy cutoff, the number of PW's depends on V. As most calculations are done far from convergence, this will cause large oscillations in the calculated $E(V)$ (this is reminiscent of the "Pulay force" problem). Experience show that the fit to an equation of state effectively smoothes the oscillations and yields very good results even if the cutoff of PW's is low.

The statement "most calculations are done far from convergence" is not as alarming as it may seem: in fact the slow convergence is due to the region of charge close to the atomic cores. This is an essentially atomic-like charge that changes little from one structure to another. If we are interested in comparing different structures of the same materials, the relative energy differences will converge with the cutoff well before the absolute energy values. Of course, one has to check carefully the relative convergence with respect to the BZ sampling as well.

It is also possible to find the pressure at which the crystal makes a transition from one structure to the other. This is achieved by connecting with a common tangent two $E(V)$ curves for two different structures. It is easy to show that this construction determines the pressure at which the enthalpies of the two phases are equal: $E_1 + PV_1 = E_2 + PV_2$. The minimum enthalpy state is the thermodynamic condition of stability at zero temperature and at constant pressure. The crossing of the enthalpies of the two phases at equal P signals the possibility of a first-order structural transition.

Of course this approach relies on some knowledge or intuition of reasonable candidates crystal structures. Generally the results are in good to very good agreement with experiments.

In more complex crystals: noncubic or with atomic positions in the unit cell that are not fixed by symmetry, the equilibrium is determined not only by the volume of the unit cell but also by other lattice parameters (for instance, c/a for tetragonal crystals) and by atomic positions in the unit cell. The approach sketched above is still valid, provided one determines the equilibrium atomic positions (see next section) and the equilibrium lattice parameters for a given volume. For the latter the calculation of *stresses* may be useful:

$$\sigma_{\alpha\beta} = \frac{1}{V}\frac{\partial E}{\epsilon_{\alpha\beta}} \tag{93}$$

where $\epsilon_{\alpha\beta}$ is the *strain*: a homogeneous deformation of all coordinates, sending $\mathbf{r}$ into $\mathbf{r}' = (1+\epsilon)\mathbf{r}$ (where ϵ is a matrix). The stresses can be calculated in DFT. At equilibrium and at zero pressure, the stresses are zero. The pressure is related to the stress by $P = -\mathrm{Tr}\sigma/3$.

4.2 Optimization of atomic positions

The problem of finding minimum of the total energy as a function of atomic positions, having fixed the unit cell, is much easier if one can calculate the gradients of the energy with respect to the variables (the atomic positions in the unit cell). This can be done quite easily, at least for a PW basis set, as shown in the next section. We are left with the textbook problem of finding the minimum of a $3N-$dimensional problem. Several well-known and well-studied algorithms exist: conjugate gradient, quasi-Newton methods, DIIS. In the appendix the conjugate gradient algorithm is examined

The two following points however must be remarked. The first is that, if we start from a system having a given symmetry, the forces will *not* break such symmetry. This may be both an advantage and a disadvantage. The second is that algorithms based on forces will very likely bring the system to the closer local minimum (a zero gradient point), rather than to the absolute minimum (the lowest-energy minimum). In situations in which there are many local minima separated by energy barriers this kind of approach can easily fail to find the global minimum. Unfortunately this is a typical situation: for instance, clusters of atoms are known to have a large number of local minima.

4.3 Hellmann-Feynman forces

Hellmann-Feynman forces are the derivative of the total energy with respect to atomic positions $\mathbf{R}_i$. For many-body Hamiltonians and wavefunctions, only terms containing *explicit* derivatives in the Hamiltonian contribute (Hellmann-Feynman theorem). The terms containing *implicit* derivatives through the wavefunctions, that we indicate with $\widetilde{\mathbf{F}}_i$, vanish:

$$\mathbf{F}_i = -\frac{d}{d\mathbf{R}_i}\langle\Psi|H|\Psi\rangle = -\langle\Psi|\frac{\partial H}{\partial\mathbf{R}_i}|\Psi\rangle - \widetilde{\mathbf{F}}_i \tag{94}$$

with

$$\widetilde{\mathbf{F}}_i = \langle\frac{d\Psi}{d\mathbf{R}_i}|H|\Psi\rangle + \langle\Psi|H|\frac{d\Psi}{d\mathbf{R}_i}\rangle = E\langle\frac{d\Psi}{d\mathbf{R}_i}|\Psi\rangle + E\langle\Psi|\frac{d\Psi}{d\mathbf{R}_i}\rangle = E\frac{d}{d\mathbf{R}_i}\langle\Psi|\Psi\rangle. \tag{95}$$

The last term vanish because it is the derivative of a constant quantity. Note that partial derivative are used to indicate explicit derivation, otherwise the total derivative is used.

In DFT the same applies, thanks to the variational character of the energy. Let us write the force as

$$\mathbf{F}_i = -\frac{dE}{d\mathbf{R}_i} = -\int n(\mathbf{r})\frac{\partial V(\mathbf{r})}{\partial\mathbf{R}_i}d\mathbf{r} - \frac{\partial E_{II}}{\partial\mathbf{R}_i} - \widetilde{\mathbf{F}}_{\mathbf{i}} \tag{96}$$

where the first term comes from explicit derivation of the energy functional, E_{II} is the ion-ion (classical) interaction energy, and the $\widetilde{\mathbf{F}}_\mathbf{i}$ contains the implicit derivation through KS orbitals:

$$\widetilde{\mathbf{F}}_\mathbf{i} = \sum_k \int \left(\frac{d\psi_k^*(\mathbf{r})}{d\mathbf{R}_i} \frac{\delta E}{\delta \psi_k^*(\mathbf{r})} + \frac{d\psi_k(\mathbf{r})}{d\mathbf{R}_i} \frac{\delta E}{\delta \psi_k(\mathbf{r})} \right) d\mathbf{r}. \tag{97}$$

Using the expression for the functional derivative of the energy functional, Eq.22, and the identity

$$0 = \frac{d}{d\mathbf{R}_i} \int n(\mathbf{r}) d\mathbf{r} = \sum_k \left(\int \frac{d\psi_k^*(\mathbf{r})}{d\mathbf{R}_i} \psi_k(\mathbf{r}) d\mathbf{r} + \int \psi_k^*(\mathbf{r}) \frac{d\psi_k(\mathbf{r})}{d\mathbf{R}_i} d\mathbf{r} \right), \tag{98}$$

the term $\widetilde{\mathbf{F}}_\mathbf{i}$ can be recast as

$$\widetilde{\mathbf{F}}_\mathbf{i} = \sum_k \int \left(\frac{d\psi_k^*(\mathbf{r})}{d\mathbf{R}_i} (H - \epsilon_k)\psi_k(\mathbf{r}) + \frac{d\psi_k(\mathbf{r})}{d\mathbf{R}_i} (H - \epsilon_k)\psi_k^*(\mathbf{r}) \right) d\mathbf{r}. \tag{99}$$

This term vanishes on the ground state. Finally, one finds that, in perfect analogy to the many-body case, the forces acting on atoms are the matrix element on the ground state of the gradient of the external potential plus an ion-ion term:

$$\mathbf{F}_i = -\int n(\mathbf{r}) \frac{\partial V(\mathbf{r})}{\partial \mathbf{R}_i} d\mathbf{r} - \frac{\partial E_{II}}{\partial \mathbf{R}_i}. \tag{100}$$

4.4 Pulay forces

Unfortunately the term $\widetilde{\mathbf{F}}_\mathbf{i}$ in Eq.99 vanishes only if we have ground state charge density and wavefunctions at perfect convergence. In the real world, this is never the case. In particular, the wavefunctions are expanded on a finite basis set that is never complete. This may produce a nonzero value of $\widetilde{\mathbf{F}}_\mathbf{i}$, called *Pulay force*.

Let us write the expansion of wavefunctions into a basis set, taken to be orthonormal for simplicity:

$$\psi_k(\mathbf{r}) = \sum_n c_n^{(k)} \phi_n(\mathbf{r}). \tag{101}$$

This will yield a secular equation

$$\sum_m (H_{nm} - \epsilon_i) c_m^{(k)} = 0, \qquad H_{nm} = \int \phi_n^*(\mathbf{r}) H \phi_m(\mathbf{r}) d\mathbf{r}. \tag{102}$$

By inserting the expansion of the KS orbitals into Eq.99 one finds

$$\widetilde{\mathbf{F}}_\mathbf{i} = \sum_k \sum_{mn} \frac{\partial c_n^{(k)}}{\partial \mathbf{R}_i} (H_{nm} - \epsilon_k) c_m^{(k)} + \sum_{mn} c_n^{(k)} c_m^{(k)} \int \frac{\partial \phi_n^*(\mathbf{r})}{\partial \mathbf{R}_i} (H - \epsilon_k)\phi_m(\mathbf{r}) d\mathbf{r} + \text{c.c.} \tag{103}$$

The first term vanish exactly even if the basis set is not complete (see Eq.102). The second term instead vanishes only if i) if the basis set is complete, or ii) if $\partial \phi_n^*(\mathbf{r})/\partial \mathbf{R}_i$ has no component outside the subspace defined by the $\phi_n(\mathbf{r})$, or iii) if the basis set does not depend explicitly on the atomic positions. The latter is the case of PW's. Pulay forces do not arise because the basis set is incomplete, but because it is "incomplete in a different way" when atoms are moved. Using PW one has also an incomplete basis set, but it is "equally incomplete" for all atomic positions in the unit cell.

In practical calculations with localized basis sets, Pulay forces must be taken into account, otherwise the error on the forces is quite large. If one wants to minimize the energy, or to do molecular dynamics simulations, it is crucial that the forces are the derivative of the energy within numerical accuracy. Although much progress has been done in the last years towards reliable calculation of forces with localized basis sets, PW's are still much more used than localized basis sets for all applications in which forces are important.

It should be kept in mind that the above results holds at perfect electronic self-consistency (or at the perfect minimum of the energy functional, in the case of direct minimization). Practical calculations of forces will always contain a small error. We will come back to this point later.

5 DFT and Molecular Dynamics

We can safely assume that ions behave as classical particles (a very good approximation, except in some cases for Hydrogen). Also, we can assume that the electrons are always on the *Born-Oppenheimer (BO) surface*, that is, in the ground state corresponding to their instantaneous positions. Under these assumptions the dynamical behavior of ions can be described by a classical Lagrangian

$$L = \frac{1}{2} \sum_i M_i \dot{\mathbf{R}}_i^2 - E_{tot}(\{\mathbf{R}\}) \tag{104}$$

where M_i are the mass of ions. The corresponding equations of motion:

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{\mathbf{R}}_i} - \frac{\partial L}{\partial \mathbf{R}_i} = 0, \qquad \mathbf{P}_i = \frac{\partial L}{\partial \dot{\mathbf{R}}_i} \tag{105}$$

are nothing but Newton's equations.

It is tempting to use Eq.104 as the basis for a *molecular dynamics* (MD) study. In classical MD, the forces are generated by an interatomic potential (often a sum of two-body terms like Lennard-Jones potentials) and the Newton equations are discretized and numerically solved. The discrete interval of time is called *time step*. A sequence of atomic coordinates and velocities is generated starting from a suitable initial set of coordinates and velocities. The sequence can be used to calculate thermodynamical averages. Straightforward MD will sample the microcanonical ensemble: constant energy at fixed volume, but it is possible to build a dynamics at constant temperature (canonical ensemble) using a *Nosé thermostat* that simulates a thermal bath, or at constant pressure, by adding a fictitious dynamics on the volume, and even more complex cases.

MD can also be used to find the global minima using the *simulated annealing* technique. The configuration space is sampled at equilibrium, then the kinetic energy is gradually removed from the systems that has the possibility (but is not guaranteed to do so) to reach the global minimum. Such procedure is sometimes the only practical way to find the global minimum for especially hard problems. In mathematical terms, "easy" problems are exactly solvable by computer algorithms in polynomial time, that is, in a number of steps that is a polynomial function of the dimension of the problem; "hard" problems are solved in exponential time. A problem is NP (nondeterministic polynomial) if its solution (if one exists) can be guessed and verified in polynomial time. This is the kind of problems for which the simulated annealing has been devised. The determination of the structure in clusters is believed to be a NP-hard problem.

5.1 Classical Molecular Dynamics

Let us consider the most basic MD : a purely mechanical system of N atoms, enclosed in a volume V (usually with periodical boundary conditions, PBC, for a condensed-matter system), having mechanical energy $E = T + E_p$, where $T = \frac{1}{2} \sum_i M_i \dot{\mathbf{R}}_i^2$ is the kinetic energy of ions, $E_p = E_p(\{\mathbf{R}\})$ is the interatomic potential energy. This is known as the NVE, or microcanonical, ensemble.

5.1.1 Discretization of the equation of motion

The numerical solution (integration) of the equations of motions is generally performed using the *Verlet algorithm*. This is obtained from the following basic and very simple equations :

$$\mathbf{R}_i(t + \delta t) = \mathbf{R}_i(t) + \delta t \mathbf{V}_i(t) + \frac{\delta t^2}{2M_i} \mathbf{f}_i(t) + \frac{\delta t^3}{6} \mathbf{b}_i(t) + \mathcal{O}(\delta t^4) \tag{106}$$

$$\mathbf{R}_i(t - \delta t) = \mathbf{R}_i(t) - \delta t \mathbf{V}_i(t) + \frac{\delta t^2}{2M_i} \mathbf{f}_i(t) - \frac{\delta t^3}{6} \mathbf{b}_i(t) + \mathcal{O}(\delta t^4) \tag{107}$$

where $\mathbf{V}_i = \dot{\mathbf{R}}_i$ are velocities, $\mathbf{f}_i$ forces acting on ion i. By summing and subtracting Eqs. (106) and (107) we get the Verlet algorithm:

$$\mathbf{R}_i(t + \delta t) = 2\mathbf{R}_i(t) - \mathbf{R}_i(t - \delta t) + \frac{\delta t^2}{M_i} \mathbf{f}_i(t) + \mathcal{O}(\delta t^4) \tag{108}$$

$$\mathbf{V}_i(t) = \frac{1}{2\delta t} \left[\mathbf{R}_i(t + \delta t) - \mathbf{R}_i(t - \delta t) \right] + \mathcal{O}(\delta t^3). \tag{109}$$

The velocities are one step behind the positions, but they are not used to update the positions. It is possible to recast the Verlet algorithm into an equivalent form (one giving exactly the same trajectories) in which both velocities and positions are updated in the same step. By combining Eq.(106) with Eq.(107) displaced in time by $+\delta t$, one finds

$$\mathbf{V}_i(t + \delta t) = \mathbf{V}_i(t) + \frac{\delta t}{2 M_i} \left[\mathbf{f}_i(t) + \mathbf{f}_i(t + \delta t) \right] \tag{110}$$

$$\mathbf{R}_i(t + \delta t) = \mathbf{R}_i(t) + \delta t \mathbf{V}_i(t) + \frac{\delta t^2}{2 M_i} \mathbf{f}_i(t). \tag{111}$$

Note that the update of velocities requires the forces for the new positions. This algorithm is known as *Velocity Verlet*. Its equivalence to the Verlet algorithm may not seem evident, but it can be proved quite simply.

In spite of his simplicity, the Verlet algorithm, in any incarnation, is efficient and numerically stable. In particular, it yields trajectories that conserve to a very good degree of accuracy the energy E. A small loss of energy conservation, due both to numerical errors and to the discretization, is unavoidable, but a systematic drift of the energy is not acceptable. In this respect Verlet is superior to apparently better (i.e. higher-order) schemes. In one of the following sections we will see one reason why this happen.

5.1.2 Thermodynamical averages

In the following we will use the *phase space* canonical variables, collectively indicated as $\mathbf{R}, \mathbf{P}$, instead of coordinates and velocities. From a practical point of view, the calculation of thermodynamical averages in classical MD is an average over many time steps:

$$A_T = \frac{1}{T} \int_0^T A(\mathbf{R}(t), \mathbf{P}(t)) dt \simeq \frac{1}{M} \sum_{n=1}^{M} A(t_n), \quad t_n = n\delta t, \quad t_M = M\delta t = T. \tag{112}$$

For an ergodic system (that is, one whose trajectories in a sufficiently long time pass arbitrary close to any point in the phase space), it is believed that:

$$\lim_{T \to \infty} A_T \to \langle A \rangle \tag{113}$$

where $\langle \rangle$ is the average over the corresponding ensemble:

$$\langle A \rangle = \int \rho(\mathbf{R}, \mathbf{P}) A(\mathbf{R}, \mathbf{P}) d\mathbf{R} d\mathbf{P} \tag{114}$$

where ρ is the probability of a microscopic state. In NVE MD the microcanonical ensemble is sampled:

$$\rho_{NVE}(\mathbf{R}, \mathbf{P}) = \frac{g(N)}{\Omega} \delta(H - E) \tag{115}$$

where H is the Hamiltonian corresponding to the Lagrangian of Eq.(104), E is the mechanical energy (including kinetic energy of ions) $g(N) = (h^{3N} N!)^{-1}$ for N indistinguishable atoms, and Ω, related to the entropy S by the Boltzmann relation $S = k_B \log \Omega$, is the total number of microscopic states:

$$\Omega = g(N) \int d\mathbf{R} d\mathbf{P} \delta(H - E). \tag{116}$$

The time step must be as big as possible in order to sample as much phase space as possible, but at the same time it must be small enough to allow to follow the motion the ions with little loss of accuracy (which usually appears as a drift in the energy). Typically $\delta t \sim 0.01 - 0.1 \delta t_{max}$, where δt_{max} is the period of the fastest phonon mode: $\delta t_{max} = 1/\omega_{max}$.

5.1.3 Verlet algorithm as unitary discretization of the Liouvillian

Let us consider an observable $A = A(\mathbf{R}, \mathbf{P}, t)$. Its time evolution can be written as

$$\frac{dA}{dt} = \sum_i \left(\dot{\mathbf{R}}_i \frac{\partial A}{\partial \mathbf{R}_i} + \dot{\mathbf{P}}_i \frac{\partial A}{\partial \mathbf{P}_i} + \frac{\partial A}{\partial t} \right) = i\mathcal{L}A + \frac{\partial A}{\partial t} \tag{117}$$

where the operator $\mathcal{L}$ is called the *Liouvillian*. Assuming that $A = A(\mathbf{R}, \mathbf{P})$ does not depend explicitly on the time, the Liouvillian determines entirely the time evolution of A: formally,

$$A(t) = e^{i\mathcal{L}t}A(t=0) = U(t)A(t=0). \tag{118}$$

It can be shown that $\mathcal{L}$ is an Hermitian operator and thus U is a unitary operator (as it should be: time-reversal symmetry must hold). We can write $\mathcal{L}$ as

$$i\mathcal{L} = \sum_i \left(\dot{\mathbf{R}}_i \frac{\partial}{\partial \mathbf{R}_i} + \dot{\mathbf{P}}_i \frac{\partial}{\partial \mathbf{P}_i} \right) = \sum_i \left(\dot{\mathbf{R}}_i \frac{\partial}{\partial \mathbf{R}_i} + \mathbf{f}_i \frac{\partial}{\partial \mathbf{P}_i} \right) \tag{119}$$

and finally as a sum of two terms, one acting on coordinates and one on momenta: $i\mathcal{L} = i\mathcal{L}_p + i\mathcal{L}_r$, where

$$i\mathcal{L}_p = \sum_i \mathbf{f}_i \frac{\partial}{\partial \mathbf{P}_i}, \qquad i\mathcal{L}_r = \sum_i \dot{\mathbf{R}}_i \frac{\partial}{\partial \mathbf{R}_i}. \tag{120}$$

Until now, we have just recast the classical equation of motion into an elegant but not especially useful formalism. Let us discretize now the time evolution operator, by dividing t into N small intervals $\delta t = t/N$, and apply the *Trotter* approximation:

$$e^{i(\mathcal{L}_p + \mathcal{L}_r)t} = \left[e^{i(\mathcal{L}_p + \mathcal{L}_r)\delta t} \right]^N = \left[e^{i\mathcal{L}_p \delta t/2} e^{i\mathcal{L}_r \delta t} e^{i\mathcal{L}_p \delta t/2} + \mathcal{O}(\delta t^3) \right]^N. \tag{121}$$

Remember that $\mathcal{L}_p$ and $\mathcal{L}_r$ are operators: the Trotter approximation is not trivial. Let us apply the operator between square brackets to a point $(\mathbf{R}_i(t), \mathbf{P}_i(t))$ in phase space at time t. We will use the known result

$$e^{a\partial/\partial x} f(x) = f(x+a) \tag{122}$$

if a does not depend on x. Since $\mathcal{L}_p$ and $\mathcal{L}_r$ are sums of terms acting on each particle separately, we can consider their action on each particle independently.

$$e^{i\mathcal{L}_p \delta t/2} (\mathbf{R}_i, \mathbf{P}_i) = \left(\mathbf{R}_i, \mathbf{P}_i + \frac{\delta t}{2} \mathbf{f}_i(\mathbf{R}) \right) \equiv (\mathbf{R}_i', \mathbf{P}_i') \tag{123}$$

$$e^{i\mathcal{L}_r \delta t} (\mathbf{R}_i', \mathbf{P}_i') = \left(\mathbf{R}_i' + \frac{\delta t}{M_i} \mathbf{P}_i', \mathbf{P}_i' \right) \tag{124}$$

$$= \left(\mathbf{R}_i + \frac{\delta t}{M_i} \mathbf{P}_i + \frac{\delta t^2}{2m} \mathbf{f}_i(\mathbf{R}_i), \mathbf{P}_i + \frac{\delta t}{2} \mathbf{f}_i(\mathbf{R}) \right) \equiv (\mathbf{R}_i'', \mathbf{P}_i'') \tag{125}$$

$$e^{i\mathcal{L}_p \delta t/2} (\mathbf{R}_i'', \mathbf{P}_i'') = \left(\mathbf{R}_i'', \mathbf{P}_i'' + \frac{\delta t}{2} \mathbf{f}_i(\mathbf{R}'') \right) \tag{126}$$

$$= \left(\mathbf{R}_i + \frac{\delta t}{M_i} \mathbf{P}_i + \frac{\delta t^2}{2M_i} \mathbf{f}_i(\mathbf{R}), \mathbf{P}_i + \frac{\delta t}{2} [\mathbf{f}_i(\mathbf{R}) + \mathbf{f}_i(\mathbf{R}'')] \right) \tag{127}$$

Noting that $\mathbf{f}_i(\mathbf{R}) = \mathbf{f}_i(t)$, $\mathbf{f}_i(\mathbf{R}'') = \mathbf{f}_i(t + \delta t)$, the last expression is nothing but the velocity Verlet algorithm for $(\mathbf{R}_i(t + \delta t), \mathbf{P}_i(t + \delta t))$.

In conclusion: the Verlet algorithm may be derived by a discretization of the time evolution operator that conserves unitarity. Such property is crucial for any well-behaved algorithm one can think of.

5.1.4 Canonical ensemble in MD

We are often interested in systems in thermal equilibrium with a thermal bath at temperature T: the NVT or *canonical* ensemble, for which

$$\rho_{NVT}(\mathbf{R}, \mathbf{P}) = \frac{g(N)}{Z} e^{-H(\mathbf{R}, \mathbf{P})/k_B T} \tag{128}$$

where Z is the partition function:

$$Z = g(N) \int d\mathbf{R} d\mathbf{P} e^{-H(\mathbf{R}, \mathbf{P})/k_B T}. \tag{129}$$

Integration over $\mathbf{P}$ gives for the partition function of N identical atoms:

$$Z = Z_r/(N!\lambda^{3N}) \tag{130}$$

where λ is the thermal wavelength:

$$\lambda = \frac{h}{\sqrt{2\pi M k_b T}} \tag{131}$$

and Z_r is the configurational partition function:

$$Z_r = \int d\mathbf{R}_1 ... \int d\mathbf{R}_n e^{-E_p(\mathbf{R})/k_B T}. \tag{132}$$

In the canonical ensemble, the temperature is related to the expectation value of the kinetic energy:

$$\langle \sum_{i=1}^{N} \frac{\mathbf{P}_i^2}{2M_i} \rangle_{NVT} = \frac{3}{2} N k_B T. \tag{133}$$

The canonical ensemble can be simulated using what is called *Nosé-Hoover thermostat*: an additional fictitious degree of freedom produces a dynamical friction force having the effect of heating ions when the kinetic energy is lower than the desired value, cooling them in the opposite case. Specifically, the equations of motion become

$$\ddot{\mathbf{R}}_i = \frac{\mathbf{f}_i}{M_i} - \dot{\zeta}\dot{\mathbf{R}}_i \tag{134}$$

$$\ddot{\zeta} = \frac{1}{Q} \left[\sum_{i=1}^{N} M_i \dot{\mathbf{R}}_i^2 - 3Nk_b T \right] \tag{135}$$

where Q plays the role of "thermal mass". The constant of motion for this system is

$$\widetilde{H} = H + \frac{Q}{2}\dot{\zeta}^2 + 3Nk_b T\zeta \tag{136}$$

but $\widetilde{H}$ does not generate the dynamics (the dynamics is *non-canonical*). It can be shown that such dynamics samples the canonical ensemble.

Although all thermodynamical properties could in principle be determined from the free energy F, it is not possible to calculate directly F from a MD simulation. The free energy (like the partition function and the entropy) cannot be simply expressed as a thermodynamical average (like the energy). Specialized algorithms are needed for free energy calculation.

5.1.5 Constant-pressure MD

Very often we are interested in simulating systems kept at a given pressure P rather that occupying a fixed volume V. *Constant-pressure* MD can be obtained by adding the volume V or, in a more general case, the cell parameters, to the dynamical variables. In the simple case of a liquid, one defines a Lagrangian:

$$\widetilde{L} = \frac{1}{2}\sum_{i=1}^{N} M_i \left(V^{1/3}\dot{\sigma}_i \right)^2 - E_p(\{V^{1/3}\sigma\}) + \frac{1}{2}W\dot{V}^2 - PV \tag{137}$$

where $\sigma_i = \mathbf{R}_i/V^{1/3}$ are scaled variables, P is the desired external pressure, and W is a (fictitious) mass for V.

For a solid, we may be interested in knowing the equilibrium unit cell volume and form under a given stress state (typically a constant external hydrostatic pressure) rather than working at fixed cell and calculating the corresponding stress. In this case one introduces a matrix $\mathbf{h}$, formed by the unit cell vectors $\mathbf{a}_i$: $\mathbf{h} = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$, and defines scaled variables $\mathbf{S}_i$ as $\mathbf{S}_i = \mathbf{h}^{-1}\mathbf{R}_i$. The extended Lagrangian becomes

$$\widetilde{L} = \frac{1}{2}\sum_{i=1}^{N} M_i \dot{\mathbf{S}}_i \mathcal{G} \dot{\mathbf{S}}_i - E_p(\{\mathbf{h}\mathbf{S}\}) + \frac{1}{2}W\mathrm{Tr}\,{}^t\dot{\mathbf{h}}\dot{\mathbf{h}} - PV \tag{138}$$

where $\mathcal{G} = \mathbf{h}^t\mathbf{h}$ is the metric tensor. The interest of variable-cell dynamics for solid-state systems reside in the possibility to simulate structural phase transitions (under applied pressure but also as a function of temperature).

5.2 Car-Parrinello Molecular Dynamics

Implementations of MD using first-principle interatomic potential calculated from DFT, as in Eq.(104), are widely used. All the MD machinery developed for classical interatomic potentials can be used. However these implementation suffer from a serious drawback. MD is quite sensitive to the quality of forces. If the forces are not the derivatives of the energy with high accuracy, the MD simulation will have problems, appearing as a drift of quantities that should be conserved (like e.g. the energy) from their values. The error on DFT forces is *linear* in the selfconsistency error of the charge density (while for the DFT energy it is *quadratic*). As a consequence, a very good and expensive convergence to self-consistency is required at every time step.

In 1985 Car and Parrinello (CP) proposed a different approach. They introduced a Lagrangian for both electronic and ionic degrees of freedom:

$$ L = \frac{\mu}{2} \sum_k \int d\mathbf{r} |\dot{\psi}_k(\mathbf{r})|^2 + \frac{1}{2} \sum_i M_i \dot{\mathbf{R}}_i^2 - E_{tot}(\{\mathbf{R}\}, \{\psi\}) + \sum_{k,l} \Lambda_{kl} \left(\int \psi_k^*(\mathbf{r}) \psi_l(\mathbf{r}) d\mathbf{r} - \delta_{kl} \right) \quad (139) $$

which generates the following set of equations of motion:

$$ \mu \ddot{\psi}_k = H\psi_k - \sum_l \Lambda_{kl}\psi_l, \quad M_i \ddot{\mathbf{R}}_i = -\frac{\partial E_{tot}}{\partial \mathbf{R}_i} \quad (140) $$

where μ is a fictitious electronic mass, and the Lagrange multipliers Λ_{kl} enforce orthonormality constraints.

The electronic degrees of freedom are, in the typical implementation, expansion coefficients of KS orbitals into PW. The forces acting on them at each time step are determined by the KS Hamiltonian calculated from the current values of ψ_k and of $\mathbf{R}_i$. The sum over orbitals for an insulating system of n electrons includes $n/2$ states, assuming that spin polarization is neglected (every orbital is occupied by two electrons). Most CP calculations are done for aperiodic systems or for systems having a large unit cell (or supercell), so that typically only the Γ point ($\mathbf{q} = 0$) is used to sample the Brillouin Zone. Note that the entire Hamiltonian operator is not required: only products $H\psi_i$ are.

The forces acting on ions have the Hellmann-Feynman form:

$$ \frac{\partial E_{tot}}{\partial \mathbf{R}_i} = \sum_k \langle \psi_k | \frac{\partial V}{\partial \mathbf{R}_i} | \psi_k \rangle \quad (141) $$

where V is the electron-ion interaction (pseudo-)potential. Note however that Hellmann-Feynman theorem holds only on the exact ground state. The relation of Car-Parrinello forces to Hellmann-Feynman forces is explained in the next section.

Orthonormality constraints are exactly imposed to the ψ at each time step, using an iterative procedure that exploits the fact that the loss of orthonormality at each time step is small.

The simulation starts by bringing the electrons to the BO surface (that is, to the ground state) at fixed ions and proceeds, using classical MD technology, on both electronic and ionic degrees of freedom. With appropriate values of μ and δt, the electrons always remain close to the BO surface, while the ions follow a trajectory that is close to the trajectory they would follow in the BO approximation.

The Car-Parrinello dynamics has turned out to be very successful especially in the study of low-symmetry situations: surfaces, clusters, liquids, disordered materials, and for the study of chemical reactions.

5.2.1 Why Car-Parrinello works

The reasons why the Car-Parrinello dynamics works so effectively are quite subtle. The dynamics for the electrons is purely classical (and fictitious: it has nothing to do with real electron dynamics). As a consequence the energy would tend to equipartition between electronic and ionic degrees of freedom, causing an energy transfer from ionic to electronic degrees of freedom. This does not happen (and must not happen, otherwise the electrons will leave the BO surface) even on long simulation times. If we analyze the dynamics in terms of oscillators, we find that the typical

frequencies associated to the fictitious electron dynamics are given by $\omega^{el} \sim \sqrt{(\epsilon_i - \epsilon_j)/\mu}$, if there is a gap in the electronic spectrum. For ions, the oscillator frequencies are the typical phonon frequencies. It turns out that, for reasonable values of the gap and of the fictitious electron mass μ, the maximum phonon frequency is much smaller than the minimum electron frequency: $\omega^{ph}_{max} << \omega^{el}_{min}$. The energy transfer from ionic to electronic degrees of freedom is as a consequence very small even on long times.

This situation generates a fast electron dynamics that keeps the electrons close to the BO surface and averages out the error on the forces, so that the much slower ionic dynamics turns out to be correct (that is, very close to the BO dynamics one would obtain from highly converged selfconsistency). A detailed explanation is contained in a 1991 paper by Pastore, Smargiassi, and Buda.

If there is no gap in the electronic spectrum, or if the gap is too small, the above picture breaks down. It may be needed to add separate thermostats to ionic and electronic degrees of freedom in order to prevent the flow of energy from the former to the latter.

5.2.2 Choice of the parameters

The choice of the electronic mass μ must strike a compromise between conservation of adiabaticity (favored by small values of μ, see above) and maximum admissible time step (that is limited by the maximum electronic frequency, so that the heaviest μ, the smaller ω^{el}_{max}, the larger δt_{max}. Typically $\mu \sim 200$ amu (1 amu=1 electron mass). For large gap systems, such as SiO_2 or H_2O, in which adiabaticity problems are minor, μ may be increased up to ~ 500-700 amu and even more. Such values of μ correspond to a typical timestep of $\sim 0.1 - 0.2 fs$.

In order to increase the time step, it is customary to introduce the so-called *mass preconditioning*. In a PW basis set, the time step is limited by high-frequency components with the largest $\mathbf{G}$ vector. These components are dominated by the kinetic energy $\hbar^2 \mathbf{G}^2/2\mu$. Since electronic masses are fictitious, it is advantageous to introduce a mass that for high-frequency components goes like $\mu(\mathbf{G}) \simeq \mu(1 + \mathbf{G}^2)$. The corresponding equations of motions are only slightly more complex.

It should be noticed, however, that too heavy electron masses adversely affect the quality of simulation via an "electron drag" effect. The electron motion follow the ionic motion with some delay, thus introducing a drag force that appears as if the ions were heavier than their masses. This "mass renormalization" must be taken into account when extracting vibrational frequencies from MD runs. In some cases, this effect can introduce a nonnegligible deviation from the true ionic dynamics.

6 Appendix

6.1 Functionals and functional derivatives

The concept of *functional* is the generalization of the concept of function: function associates a value with another value, while a functional associates a value with a given function. The functional dependence is indicated by square brackets, like in $E[n(\mathbf{r})]$.

Functional derivatives $\delta F[f(x)]/\delta f(y)$ are defined implicitly through the expression

$$\delta F = \int \left(\frac{\delta F[f(x)]}{\delta f(y)} \right) \delta f(y) dy \tag{142}$$

where δF is the first-order variation of $F[f(x)]$ produced by an arbitrary variation $\delta f(y)$ of $f(y)$. Functional derivatives obeys some simple rules similar to those for normal derivatives. If $f(x)$ is a *function*,

$$\frac{\delta f(x)}{\delta f(y)} = \delta(x - y). \tag{143}$$

If a functional is the product of two functionals $F[f(x)]$ and $G[f(x)]$,

$$\frac{\delta F[f(x)]G[f(x)]}{\delta f(y)} = \frac{\delta F[f(x)]}{\delta f(y)}G[f(x)] + F[f(x)]\frac{\delta G[f(x)]}{\delta f(y)}. \tag{144}$$

The following "chain relation" applies:

$$\frac{\delta F[f(g(x))]}{\delta g(z)} = \int \frac{\delta F}{\delta f(y)}\frac{\delta f(y)}{\delta g(z)}dy. \tag{145}$$

Note that the functional dependence is sometimes removed in functional derivatives in order to simplify the notations.

6.2 Iterative diagonalization

Iterative diagonalization can be used whenever
    i) the number of states to be calculated is much smaller than the dimension of the basis set, and
    ii) a reasonable and economical estimate of the inverse operator H^{-1} is available.
    Both conditions are satisfied in practical calculation in a PW basis set: the number of PW's is usually much larger than the number of bands, and the Hamiltonian matrix is dominated by the kinetic energy at large $\mathbf{G}$ (the Hamiltonian is *diagonally dominant*).
    Iterative methods are based on a repeated refinement of a trial solution, which is stopped when satisfactory convergence is achieved. The number of iterative steps cannot be predicted in advance. It depends heavily on the structure of the matrix, on the type of refinement used, and on the starting point. A well-known and widely used algorithm is due to Davidson. In this method, a set of correction vectors $|\delta\psi_i\rangle$ to the M trial eigenvectors $|\psi_i\rangle$ are generated as follows:

$$|\delta\psi_i\rangle = \frac{1}{D - \epsilon_i}(H - \epsilon_i)|\psi_i\rangle \tag{146}$$

where the $\epsilon_i = \langle\psi_i|H|\psi_i\rangle$ are the trial eigenvalues. The $|\delta\psi_i\rangle$'s are orthogonalized and the Hamiltonian is diagonalized (with conventional techniques) in the subspace spanned by the trial and correction vectors. A new set of trial eigenvectors is obtained and the procedure is iterated until convergence is achieved. A good set of starting trial vectors is supplied by the eigenvectors at the preceding iteration of the potential.
    An important point is the following. The Hamiltonian matrix is never explicitly required excepted for its diagonal part. Only $H\psi_i$ products are required, which can be calculated in a very convenient way by applying the *dual-space technique*. In fact the kinetic energy is diagonal in $\mathbf{G}$-space, whereas the local potential term is diagonal in real space. Using FFT's (see below) one can go quickly back and forth from real to reciprocal space and perform the products where it is more convenient. There is still a nonlocal term which appears to require the storage of the matrix. The trick is to write V_{NL} in a *separable* form:

$$V_{NL}(\mathbf{k} + \mathbf{G}, \mathbf{k} + \mathbf{G}') = \sum_{\mu=1}^{N_{at}}\sum_{j=1}^{n} f_j^\mu(\mathbf{k} + \mathbf{G})g_j^\mu(\mathbf{k} + \mathbf{G}'), \tag{147}$$

where n is a small number and N_{at} is the number of atoms in the unit cell. This allows us to perform the products by storing only the f and g vectors.

6.3 Fast-Fourier Transform

An important computational advantage of PW's is the existence of very fast algorithms (known as the Fast Fourier-Transform, FFT) to perform the discrete Fourier transforms. This allows simple and fast transformation from reciprocal to real space and vice versa. The basic one-dimensional FFT executes the following transformation:

$$f_i = \sum_{j=0}^{N-1} g_j e^{2\pi i j/N}, \quad i = 0, ..., N-1 , \tag{148}$$

and its inverse

$$g_i = \frac{1}{N}\sum_{j=0}^{N-1} f_j e^{-2\pi i j/N}. \tag{149}$$

The transformation is usually performed "in place", that is the result is overwritten on the input vector. This takes $\mathcal{O}(N \log N)$ operations instead of $\mathcal{O}(N^2)$ of a straightforward summation. In three dimensions the discrete Fourier transform maps a function $\widetilde{f}(\mathbf{g}_i)$ in reciprocal space into a function $f(\mathbf{r}_i)$ in the unit cell (and vice versa):

$$\mathbf{g}_i = i_1 \mathbf{G}_1 + i_2 \mathbf{G}_2 + i_3 \mathbf{G}_3, \quad \mathbf{r}_i = \frac{j_1}{N_1} \mathbf{R}_1 + \frac{j_2}{N_2} \mathbf{R}_2 + \frac{j_3}{N_3} \mathbf{R}_3 \qquad (150)$$

where $\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3$ $(\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3)$ are the three fundamental translations that generate the real-space (reciprocal) lattice, $i_1 = -N_1/2, ..., N_1/2$, and so on. N_1, N_2, N_3 must be sufficiently large to include all available Fourier components; the more Fourier components, the larger the grid in $\mathbf{G}$-space and the finer the grid in $\mathbf{R}$-space. It is easily verified that this 3-d FT can be done in a very fast way by performing 3 inter-nested 1-d FFT.

6.4 Conjugate Gradient

In the following let us consider a function $f(\mathbf{x})$ of the variables $\mathbf{x} \equiv (x_i, ..., x_N)$ and its gradients $\mathbf{g}(\mathbf{x}) = -\nabla_{x_i} f(\mathbf{x})$.

The first obvious minimization algorithm that comes to mind is *steepest descent* (SD). This consists in minimizing $f(\mathbf{x})$ along the direction $\mathbf{g}(\mathbf{x})$. Once the minimum along such direction is reached, the gradient is recalculated, a new minimum is sought along the new direction of the gradient, and so on.

SD is a prototypical *direction set* method: the gradient is eliminated one component at the time along a set of directions. In SD every direction is orthogonal to the previous one (by construction). SD is not bad far from convergence, but it becomes very bad very quickly. A reason for bad convergence is that the set of directions in SD is not optimal. Let us consider such aspect in more detail in the following.

The function in the region not far from the minimum is approximately quadratic:

$$f(\mathbf{x}) \simeq \frac{1}{2}{}^t\mathbf{x} \cdot \mathbf{A} \cdot \mathbf{x} - \mathbf{b} \cdot \mathbf{x} + f_0, \qquad \mathbf{g}(\mathbf{x}) = -\mathbf{A} \cdot \mathbf{x} + \mathbf{b} \qquad (151)$$

where A is a matrix, $\mathbf{b}$ is a vector (not necessarily known).

An optimal set of directions should ensure that when we search for a minimum along the new direction, we do not lose what we have gained in the preceding step. Let us assume that at step n we reached the minimum along line $\mathbf{h}_n$. This implies: $\mathbf{g}(\mathbf{x}_n) \cdot \mathbf{h}_n = 0$. We move from $\mathbf{x}_n$ along direction $\mathbf{h}_{n+1}$. The gradient change $\delta \mathbf{g}$ is proportional to $\mathbf{A} \cdot \mathbf{h}_{n+1}$. If we impose that this change has no component along all previous minimization directions $\mathbf{h}_n$, we get the condition

$$\mathbf{h}_n \cdot \mathbf{A} \cdot \mathbf{h}_m = 0 \qquad (152)$$

that defines *conjugate* directions. The simpler *conjugate gradient* (CG) algorithm is as follows:

1. start by minimizing along $\mathbf{h}_0 = \mathbf{g}_0 = -\nabla f(\mathbf{x}_0)$. If the function is quadratic, the minimum can be found analytically: $\mathbf{x}_1 = \mathbf{x}_0 + \lambda_0 \mathbf{h}_0$, where $\lambda_0 = -\mathbf{h}_0 \cdot \mathbf{g}_0 / \mathbf{h}_0 \cdot \mathbf{A} \cdot \mathbf{h}_0$.

2. find the next direction $\mathbf{h}_1 = \mathbf{g}_1 + \gamma_1 \mathbf{h}_0$ and impose that it is conjugate to the preceding one, Eq.152. One finds $\gamma_1 = \mathbf{g}_1 \cdot \mathbf{g}_1 / \mathbf{g}_0 \cdot \mathbf{g}_0$.

3. iterate the procedure until the desired convergence. The sequence of gradients $\mathbf{g}_n$ and of conjugate gradients $\mathbf{h}_n$ is found to obey $\mathbf{g}_n \cdot \mathbf{g}_m = 0$, $\mathbf{g}_n \cdot \mathbf{h}_m = 0$, and Eq.152, for all n, m.

If the problem is not quadratic, so that $\mathbf{A}$ is not a priori known, the algorithm remain the same, but the analytical determination of the line minimum in step 1) is not performed. A numerical minimization along the $\mathbf{h}$ directions is performed and the gradient $\mathbf{g}$ is calculated at the line minimum.

CG converges much better than SD, with negligible additional effort. If the problem is purely quadratic, *exact* convergence is guaranteed to be reached in N steps. This would take $\mathcal{O}[N^3]$ operations, not better than the inversion of $\mathbf{A}$. *Approximate* convergence, however, can be reached in a much smaller number of steps. Moreover CG can be applied in presence of large matrices $\mathbf{A}$

for which inversion is impractical or impossible. Only the results of operator $\mathbf{A}$ on trial vectors $\mathbf{A} \cdot \mathbf{x}$ are required, not the entire operator.

In general, the rate of convergence is determined by the ratio between the largest and smallest eigenvalues of the matrix $\mathbf{A}$: the closer to 1, the better. Since in real-life example such ratio may considerably differ from 1, a technique known as *preconditioning* is often used to produce an equivalent problem for which such ratio is closer to 1, thus yielding better convergence properties.

The CG method, in many variants, is much used not only for structural optimization but also as an alternative method to self-consistency for finding the minimum of the energy functional at fixed ions ("electronic" minimization). In this case the variables $\mathbf{x}$ are the expansion coefficients of KS orbitals into the PW or any other basis set. The algorithm becomes slightly more complicated because orthonormality constraints between orbitals must be taken into account.

Other minimization methods   The CG method does not use explicitly the second derivative matrix $\mathbf{A}$ or its inverse $\mathbf{A}^{-1}$. This is an advantage if the number of variables in the problem is large (as i.e. in the electronic minimization problem mentioned above): the storage of an approximate $\mathbf{A}$ or $\mathbf{A}^{-1}$ would be prohibitely large. For structural optimization however the number of variables never exceeds a few hundreds. Moreover it is conceivable to find with little effort reasonable approximations to $\mathbf{A}$, which is related to the force constant matrix. *Quasi-Newton* methods make use of some guess for $\mathbf{A}$ and produce an iterative refinement of $\mathbf{A}$ (or of $\mathbf{A}^{-1}$) using the available information.

6.5   Essential Bibliography

The following is not meant to be a complete list of references. Only a few recent and detailed references are given.

Density Functional Theory of Atoms and Molecules, R.G. Parr and W. Yang (Oxford University Press, New York, 1989).
General DFT

Density Functional Theory, R.M. Dreizler and E.K.U. Gross, Springer-Verlag, Berlin (1990).
General DFT, more formal.

R.O. Jones and O. Gunnarson, Rev. Mod. Phys. **61**, 689 (1989).
Performances of DFT and LDA in atoms.

W.E. Pickett, Computer Phys. Reports **9**,115 (1989).
Very good review paper of the plane-wave-pseudopotential LDA method. Unfortunately almost impossible to find.

U. von Barth and C.G. Gelatt, Phys. Rev. B **21**, 2222 (1980).
The paper that rejects the conclusions of Janak's paper (J. F. Janak, Solid State Commun. **20**, 151 (1976)) and clarifies the validity of the frozen-core approximation

A. Baldereschi, Phys. Rev. B **7**, 5212 (1973).
The really "special" special point

A Molecular Dynamics Primer, by Furio Ercolessi, `http://www.fisica.uniud.it/~ercolessi/md`
An introduction to Classical Molecular Dynamics, with examples and sample codes.

G. Galli and A. Pasquarello, in *Computer Simulation in Chemical Physics*, edited by M.P. Allen and D.J. Tildesley (Kluwer, Amsterdam, 1993), p. 261.
An introduction to the Car-Parrinello molecular dynamics

D. Marx and J. Hutter, in *Modern Methods of Quantum Chemistry*, (John von Neumann Insitute for Computing, FZ Jülich, 2000) p. 301-449.
All you wanted to know about Car-Parrinello molecular dynamics

G. Pastore, E. Smargiassi, F. Buda, Phys. Rev. A **44**, 6334 (1991).
The paper that explains why the Car-Parrinello dynamics work.

S. Baroni, S. de Gironcoli, A. Dal Corso, and P. Giannozzi, Rev. Mod. Phys. **73**, 515-562 (2001).
Review paper on Density-Functional Perturbation Theory and its applications

7 Electronic screening: basics

We want to study the effect of applying a perturbation $\delta V_0(\mathbf{r})$ to a system of electrons under an external potential $V_0(\mathbf{r})$. At linear order in the perturbation, the perturbation produces a variation of the charge density $\delta n(\mathbf{r})$ proportional to $\delta V_0(\mathbf{r}')$:

$$\delta n(\mathbf{r}) = \int \chi(\mathbf{r}, \mathbf{r}') \delta V_0(\mathbf{r}') d\mathbf{r}' \tag{153}$$

where $\chi(\mathbf{r}, \mathbf{r}')$ is defined as the *density response of the system*. We remark that $\delta n(\mathbf{r})$ is a functional of $\delta V_0(\mathbf{r})$, and that $\chi(\mathbf{r}, \mathbf{r}')$ is a functional derivative:

$$\chi(\mathbf{r}, \mathbf{r}') = \frac{\delta n(\mathbf{r})}{\delta V_0(\mathbf{r}')}. \tag{154}$$

The polarization $\delta n(\mathbf{r})$ produces an electrostatic field that *screens* the perturbing potential $\delta V_0(\mathbf{r})$. The screened potential $\delta V(\mathbf{r})$ becomes, from Poisson's equation:

$$\delta V(\mathbf{r}) = \delta V_0(\mathbf{r}) + e^2 \int \frac{\delta n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \tag{155}$$

that is:

$$\delta V(\mathbf{r}) = \delta V_0(\mathbf{r}) + e^2 \int \frac{\chi(\mathbf{r}', \mathbf{r}'')}{|\mathbf{r} - \mathbf{r}'|} \delta V_0(\mathbf{r}'') d\mathbf{r}' d\mathbf{r}'' \tag{156}$$

$$= \int \left(\delta(\mathbf{r} - \mathbf{r}') + e^2 \int \frac{\chi(\mathbf{r}'', \mathbf{r}')}{|\mathbf{r} - \mathbf{r}''|} d\mathbf{r}'' \right) \delta V_0(\mathbf{r}') d\mathbf{r}' \tag{157}$$

$$= \int \epsilon^{-1}(\mathbf{r}, \mathbf{r}') \delta V_0(\mathbf{r}') d\mathbf{r}'. \tag{158}$$

For an applied electrostatic potential, the response function we have called $\epsilon^{-1}(\mathbf{r}, \mathbf{r}')$ is nothing but the dielectric response function as usually defined in electrostatics.

In summary, $\chi(\mathbf{r}, \mathbf{r}')$ yields the charge response from a bare (external) perturbing potential via

$$\delta n(\mathbf{r}) = \int \chi(\mathbf{r}, \mathbf{r}') \delta V_0(\mathbf{r}') d\mathbf{r}' \tag{159}$$

while $\epsilon^{-1}(\mathbf{r}, \mathbf{r}')$, defined as

$$\epsilon^{-1}(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}') + e^2 \int \frac{\chi(\mathbf{r}'', \mathbf{r}')}{|\mathbf{r} - \mathbf{r}''|} d\mathbf{r}'' \tag{160}$$

yields the screened potential from the bare one via

$$\delta V(\mathbf{r}) = \epsilon^{-1}(\mathbf{r}, \mathbf{r}') \delta V_0(\mathbf{r}') d\mathbf{r}'. \tag{161}$$

Their determination is a difficult many-body problem.

Several approximate models for dielectric screening are known. Let us consider a homogeneous and isotropic medium. We will expect that $\chi(\mathbf{r}, r') = \chi(\mathbf{r} - \mathbf{r}')$ and the same for ϵ. In reciprocal space we will have the so-called *diagonal screening*:

$$\delta n(\mathbf{k}) = \chi(\mathbf{k}) \delta V_0(\mathbf{k}). \tag{162}$$

The corresponding dielectric response is

$$\epsilon^{-1}(\mathbf{k}) = 1 + \frac{4\pi e^2}{k^2} \chi(k) \equiv \frac{1}{\epsilon(\mathbf{k})} \tag{163}$$

and the screened potential

$$\delta V(\mathbf{k}) = \frac{\delta V_0(\mathbf{k})}{\epsilon(\mathbf{k})}. \tag{164}$$

For metals, an approximate form for $\epsilon(k)$ is given by Thomas-Fermi theory:

$$\epsilon(\mathbf{k}) = 1 + \frac{k_{TF}^2}{k^2} \tag{165}$$

where k_{TF} is a parameter depending on the homogeneous charge density.

7.1 Electronic screening: DFT linear-response theory

Let us assume a system of electrons obeying KS equations:

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V_{KS}(\mathbf{r})\right)\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\mathbf{r}) \tag{166}$$

where $V_{KS}(\mathbf{r}) = V(\mathbf{r}) + V_H(\mathbf{r}) + V_{xc}[n(\mathbf{r})]$ and the charge density is generalized to the case of noninteger occupancy f_i:

$$n(\mathbf{r}) = \sum_i f_i|\psi_i(\mathbf{r})|^2. \tag{167}$$

An external potential $\delta V(\mathbf{r})$ modifies V_{ks} through $\delta V_{ks} = \delta V_0(\mathbf{r}) + \delta V_H(\mathbf{r}) + \delta V_{xc}[n(\mathbf{r})]$. Let us consider first-order changes (linear response). We introduce am auxiliary quantity $\chi_0(\mathbf{r},\mathbf{r}')$, the *independent-particle polarizability*, defined via

$$\delta n(\mathbf{r}) = \int \chi_0(\mathbf{r},\mathbf{r}')\delta V_{KS}(\mathbf{r}')d\mathbf{r}'. \tag{168}$$

This is the functional derivative of the charge density with respect to the screened KS potential

$$\chi_0(\mathbf{r},\mathbf{r}') = \frac{\delta n(\mathbf{r})}{\delta V_{KS}(\mathbf{r}')} \tag{169}$$

(compare with the definition of $\chi(\mathbf{r},\mathbf{r}')$ above). The quantity $\chi_0(\mathbf{r},\mathbf{r}')$ can be easily derived from perturbation theory. The first-order variation of KS orbitals is given by

$$\delta\psi_i(\mathbf{r}) = \sum_{j\neq i}\psi_j(\mathbf{r})\frac{\langle\psi_j|\delta V_{ks}|\psi_i\rangle}{\epsilon_i - \epsilon_j} \tag{170}$$

(note that at linear order the new KS orbitals $\psi_i + \delta\psi_i$ are orthonormal) and the first-order variation of the charge density is:

$$\delta n(\mathbf{r}) = \sum_i f_i\delta\psi_i^*(\mathbf{r})\psi_i(\mathbf{r}) + c.c. = \sum_{i,j,i\neq j} f_i\left(\psi_i^*(\mathbf{r})\psi_j(\mathbf{r})\frac{\langle\phi_j|\delta V_{ks}|\psi_i\rangle}{\epsilon_i - \epsilon_j} + c.c\right). \tag{171}$$

The last equation can be rewritten in more symmetric form:

$$\delta n(\mathbf{r}) = \sum_{i,j,i\neq j}\frac{f_i - f_j}{\epsilon_i - \epsilon_j}\psi_i^*(\mathbf{r})\psi_j(\mathbf{r})\langle\psi_j|\delta V_{ks}|\psi_i\rangle \tag{172}$$

explicitly showing that the contribution from i,j states vanishes if both are fully occupied. For a closed-shell (insulating) system:

$$\delta n(\mathbf{r}) = 4\text{Re}\sum_{v,c}\psi_v^*(\mathbf{r})\psi_c(\mathbf{r})\frac{\langle\psi_c|\delta V_{ks}|\psi_v\rangle}{\epsilon_v - \epsilon_c} \tag{173}$$

where v denotes filled (valence) states, c empty (conduction) states. We can write the independent-particle polarizability $\chi_0(\mathbf{r},\mathbf{r}')$ as

$$\chi_0(\mathbf{r},\mathbf{r}') = 4\text{Re}\sum_{v,c}\frac{\psi_v^*(\mathbf{r})\psi_c(\mathbf{r})\psi_c^*(\mathbf{r}')\psi_v(\mathbf{r}')}{\epsilon_v - \epsilon_c}. \tag{174}$$

Such expression can be recast into the form

$$\chi_0(\mathbf{r},\mathbf{r}') = 4\text{Re}\sum_v \psi_v^*(\mathbf{r})P_c\frac{1}{\epsilon_v - H_{KS}}P_c\psi_v(\mathbf{r}') \tag{175}$$

where P_c is the projector operator over conduction states Note that:

- this expression is valid only if $V_K S$ is described by a local potential: $V_{KS} = V_{KS}(\mathbf{r})$

- $\chi_0(\mathbf{r},\mathbf{r}')$ *is* a ground-state property, even if it apparently depends on excited states.

7.2 Physical Response operator

The above expression provides to us a way to calculate a "fake" response operator $\chi_0(\mathbf{r}, \mathbf{r}')$, that gives the response to a given *screened* potential. How does it lead to the "real" response operator $\chi(\mathbf{r}, \mathbf{r}')$?

The screening from the Hartree potential, at first order, is:

$$\delta V_H(\mathbf{r}) = e^2 \int \frac{\delta n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' = e^2 \int v_c(\mathbf{r}, \mathbf{r}') \delta n(\mathbf{r}') d\mathbf{r}' \tag{176}$$

where $v_c(\mathbf{r}, \mathbf{r}') = 1/|\mathbf{r} - \mathbf{r}'|$. From exchange-correlation:

$$\delta V_{xc}(\mathbf{r}) = \int f_{xc}(\mathbf{r}, \mathbf{r}') \delta n(\mathbf{r}') d\mathbf{r}' \tag{177}$$

where

$$f_{xc}(\mathbf{r}, \mathbf{r}') = \frac{\delta V_{xc}(\mathbf{r})}{\delta n(\mathbf{r}')} = \frac{\delta^2 E_{xc}}{\delta n(\mathbf{r}) \delta n(\mathbf{r}')}. \tag{178}$$

In operator notations:

$$\delta n = \hat{\chi}_0 \delta V = \hat{\chi}_0 \left(\delta V_0 + \hat{v}_c \delta n + \hat{f}_{xc} \delta n \right) \tag{179}$$

$$\left(1 - \hat{\chi}_0 \hat{v}_c - \hat{\chi}_0 \hat{f}_{xc} \right) \delta n = \hat{\chi}_0 \delta V_0 \tag{180}$$

$$\delta n = \left(1 - \hat{\chi}_0 \hat{v}_c - \hat{\chi}_0 \hat{f}_{xc} \right)^{-1} \hat{\chi}_0 \delta V_0 \tag{181}$$

and finally

$$\hat{\chi} = \left(1 - \hat{\chi}_0 \hat{v}_c - \hat{\chi}_0 \hat{f}_{xc} \right)^{-1} \hat{\chi}_0 = \left(\hat{\chi}_0^{-1} - \hat{v}_c - \hat{f}_{xc} \right)^{-1} \tag{182}$$

(remember that these are operators!!!).

Early calculations used the RPA (Random Phase Approximation), consisting in neglecting the f_{xc} term. The addition of exchange-correlation at the LDA level is however straightforward: the exchange-correlation function f_{xc} is a local operator

$$f_{xc}(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}') \left. \frac{dV_{xc}(n)}{dn} \right|_{n=n(\mathbf{r})}. \tag{183}$$

The major problem of such straightforward approach is how to invert the operators. In solids, the response function $\chi_0(\mathbf{r}, \mathbf{r}')$ can be expressed in reciprocal space as a matrix (the *dielectric matrix*): $\chi_0(\mathbf{k} + \mathbf{q} + \mathbf{G}, \mathbf{k} + \mathbf{q} + \mathbf{G}')$ for the response to an external perturbation of wavevector $\mathbf{q}$. The operations between operators becomes matrix operations over infinite matrices. By truncating the matrices at an appropriate $\mathbf{G}$ one has a practical scheme for calculating response operators.

7.3 Density-Functional Perturbation Theory

The method described in the preceding section is computationally very heavy. In fact what we calculate is the response to *all* possible perturbations. We are often interested to the response to a specific perturbation instead. In this sense the calculation of the dielectric matrix is an overkill. Moreover the method is not suitable for nonlocal perturbing potential. Since most DFT calculations rely on nonlocal PP's, this is a severe limitation.

An alternative method is the following. We consider the basic equations

$$\delta V_{ks} = \delta V_0 + \hat{v}_c \delta n + \hat{f}_{xc} \delta n \tag{184}$$

and

$$\delta n(\mathbf{r}) = 4 \text{Re} \sum_{v,c} \psi_v^*(\mathbf{r}) \psi_c(\mathbf{r}) \frac{\langle \psi_c | \delta V_{ks} | \psi_v \rangle}{\epsilon_v - \epsilon_c} \tag{185}$$

solve self-consistently. The sum over unoccupied states is better expressed via an implicit summation:

$$\sum_c \psi_c(\mathbf{r}) \frac{\langle \psi_c | \delta V_{ks} | \psi_v \rangle}{\epsilon_v - \epsilon_c} = P_c \frac{1}{\epsilon_v - H_{KS}} P_c \delta V_{ks} \psi_v. \tag{186}$$

The projector over empty states can be written as

$$P_c = 1 - P_v = 1 - \sum_v |\psi_v\rangle\langle\psi_v| \tag{187}$$

so that conduction bands are never explicitly required. The variation of the charge density can be recast into the form

$$\delta n(\mathbf{r}) = 4\mathrm{Re}\sum_v \psi_v^*(\mathbf{r})\Phi_v(\mathbf{r}) \tag{188}$$

where the right-hand term $\Phi_v(\mathbf{r})$:

$$\Phi_v = P_c \frac{1}{\epsilon_v - H_{KS}} P_c \delta V_{ks}\psi_v \tag{189}$$

is obtained from the solution of a linear equation:

$$(\epsilon_v - H_{KS}) P_c \Phi_v = P_c \delta V_{ks}\psi_v. \tag{190}$$

The above equations define a self-consistent procedure that can be solved by iteration, much in the same way as in the solution of KS equations.

A typical application of such procedure is the calculation of phonon dispersion spectra. The perturbing potential is in this case the displacement of an ion. Once $\delta n(\mathbf{r})$ is calculated, the second-order expansion term of the energy, $E^{(2)}$, can be easily calculated. In the next section we show that

$$E^{(2)} = \frac{1}{2}\int \delta n(\mathbf{r})\delta V_0(\mathbf{r})d\mathbf{r} \tag{191}$$

holds.

7.4   Total energy expansion

Let us consider the many-body Hamiltonian (not the KS Hamiltonian) and assume that the external potential depend on some external parameter λ: $V = V_\lambda$. The energy will be

$$E(\lambda) = \langle\Psi_\lambda|T + U + V_\lambda|\Psi_\lambda\rangle \tag{192}$$

where $H_\lambda = T + U + V_\lambda$, T is the many-body kinetic energy, U the Coulomb interaction, and

$$H_\lambda|\Psi_\lambda\rangle = E_\lambda|\Psi_\lambda\rangle. \tag{193}$$

The first-order derivative of the energy with respect to λ is

$$\frac{dE(\lambda)}{d\lambda} = \langle\frac{d\Psi_\lambda}{d\lambda}|T + U + V_\lambda|\Psi_\lambda\rangle + \langle\Psi_\lambda|\frac{dV_\lambda}{d\lambda}|\Psi_\lambda\rangle + \langle\Psi_\lambda|T + U + V_\lambda|\frac{d\Psi_\lambda}{d\lambda}\rangle \tag{194}$$

$$= E_\lambda\left(\langle\frac{d\Psi_\lambda}{d\lambda}|\Psi_\lambda\rangle + \langle\Psi_\lambda|\frac{d\Psi_\lambda}{d\lambda}\rangle\right) + \langle\Psi_\lambda|\frac{dV_\lambda}{d\lambda}|\Psi_\lambda\rangle. \tag{195}$$

The term between brackets vanishes because it is equal to the derivative of a constant: $d/d\lambda\langle\Psi_\lambda|\Psi_\lambda\rangle$. We find the well-known Hellmann-Feynman result:

$$\frac{dE(\lambda)}{d\lambda} = \langle\Psi_\lambda|\frac{dV_\lambda}{d\lambda}|\Psi_\lambda\rangle \tag{196}$$

stating that there is no contribution to the first-order energy derivative from $|d\Psi_\lambda/d\lambda\rangle$.

Let us consider 2nd-order derivatives:

$$\frac{d^2E(\lambda)}{d\lambda^2} = \frac{d}{d\lambda}\langle\Psi_\lambda|\frac{dV_\lambda}{d\lambda}|\Psi_\lambda\rangle \tag{197}$$

$$= \langle\frac{d\Psi_\lambda}{d\lambda}|\frac{dV_\lambda}{d\lambda}|\Psi_\lambda\rangle + \langle\Psi_\lambda|\frac{dV_\lambda}{d\lambda}|\frac{d\Psi_\lambda}{d\lambda}\rangle + \langle\Psi_\lambda|\frac{d^2V_\lambda}{d\lambda^2}|\Psi_\lambda\rangle. \tag{198}$$

Let us write the energy, potential and wavefunctions as a power series expansion with respect to the external parameter λ:

$$E(\lambda) \simeq E^{(0)} + \lambda E^{(1)} + \frac{1}{2}\lambda^2 E^{(2)} + ... \tag{199}$$

$$V(\lambda) \simeq V^{(0)} + \lambda V^{(1)} + \frac{1}{2}\lambda^2 V^{(2)} + ... \tag{200}$$

$$\Psi(\lambda) \simeq \Psi^{(0)} + \lambda \Psi^{(1)} + \frac{1}{2}\lambda^2 \Psi^{(2)} + ... \tag{201}$$

$$\tag{202}$$

The above results can be resumed into the formula

$$E^{(1)} = \langle \Psi^{(0)} | V^{(1)} | \Psi^{(0)} \rangle \tag{203}$$

and

$$E^{(2)} = \langle \Psi^{(0)} | V^{(2)} | \Psi_0 \rangle + \frac{1}{2} \left(\langle \Psi^{(1)} | V^{(1)} | \Psi^{(0)} \rangle + \langle \Psi^{(0)} | V^{(1)} | \Psi^{(1)} \rangle \right). \tag{204}$$

Note that there is no contribution to the second-order energy term from $\Psi_{(2)}\rangle$. This is a special case of a general result known as *2n+1 theorem*: the knowledge of wavefunctions up to order n in the expansion determines the energy terms up to order $2n + 1$. For a local potential we can recast the above result into a simpler form:

$$E^{(1)} = \int n^{(0)}(\mathbf{r}) V^{(1)}(\mathbf{r}) d\mathbf{r}, \qquad E^{(2)} = \int n^{(0)}(\mathbf{r}) V^{(2)}(\mathbf{r}) d\mathbf{r} + \frac{1}{2} \int n^{(1)}(\mathbf{r}) V^{(1)}(\mathbf{r}). \tag{205}$$

The form of higher-order terms becomes increasingly complex with increasing order. The $2n+1$ theorem can however be demonstrated in one line (courtesy of Stefano de Gironcoli, SISSA, Trieste). Let us consider the energy at each λ as a functional of the charge density and of the external potential:

$$E_\lambda = E[V_\lambda, n_\lambda] = E[V_\lambda, n^{(0)} + \lambda n^{(1)} + ... + \frac{1}{i!}\lambda^i n^{(i)}] + \mathcal{O}(\lambda^{2i+2}). \tag{206}$$

The last step is a consequence of the variational character of the energy: the first term we neglect in the charge density is $\mathcal{O}(\lambda^{i+1})$, so the error on the energy will be of the order of the square of the error on the charge density.