

Chapter 2

Methods of electronic structure

Introduction

A set of approximations and techniques are introduced that allow an affordable and efficient implementation of the concepts discussed in Chapter 1. First and foremost, the Local Density Approximation [4][47] to the exchange-correlation energy is described. This approximation, or one of its recent refinements [48], brings the conceptualization operated by Density Functional Theory down to an explicit formulation, establishing a limit to its ultimate quantum-mechanical accuracy while making at the same time quantum mechanics practically applicable to an immense variety of systems. The pseudopotential scheme [6][7] is then presented, with its goal of decoupling the small and computationally expensive length scales typical of the core electrons, confined around each nucleus, from those of the interacting gas of the valence electrons, which are responsible for the majority of structural and chemical properties of interest.

These two approximations result in an explicit and manageable form for the total energy of the valence electron gas in the field of the nuclei and their frozen cores; the problem then turns computational, and so a brief description of the most successful methods for the self-consistent minimization of the LDA total-energy functional [49][50] follows. Finally the different problems that arise in dealing with systems that are either metallic or have a vanishing gap are discussed, together with the present state of research in this area.

2.1 The Local Density Approximation

The Kohn-Sham scheme provides an appealing decomposition of the total-energy functional, via its rigorous treatment of the non-interacting kinetic energy term T_s . A local density approximation can then be introduced [4], at variance with the Thomas-Fermi choice of the kinetic energy itself, solely for the remaining unknown *exchange-correlation* functional $E_{xc}[n(\mathbf{r})]$. Kohn and Sham proposed the approximation (universally known as LDA)

$$E_{xc}^{LDA}[n(\mathbf{r})] = \int \epsilon_{xc}(n(\mathbf{r}))n(\mathbf{r}) d\mathbf{r}, \quad (2.1)$$

where $\epsilon_{xc}(n) = \epsilon_{xc}(n(\mathbf{r}))$ is the exchange and correlation energy of the homogeneous electron gas of constant density $n = n(\mathbf{r})$. This defines automatically the exchange-correlation potential appearing in the Kohn-Sham equations (1.32) or in the total-energy functional (1.34) as:

$$v_{xc}^{LDA}(\mathbf{r}) = \frac{\delta E_{xc}^{LDA}}{\delta n(\mathbf{r})} = \epsilon_{xc}(n(\mathbf{r})) + \frac{\partial \epsilon_{xc}(n)}{\partial n}n(\mathbf{r}). \quad (2.2)$$

The function $\epsilon_{xc}(n)$ can be decomposed in an exchange part, for which the Dirac [51] result gives

$$\epsilon_x(n) = -\frac{3}{4} \left(\frac{3}{\pi} \right)^{1/3} n^{1/3},$$

and a remaining correlation term $\epsilon_c(n)$, a function solely of n , that can be determined once for all by interpolating between quantum MonteCarlo calculations [47] and the appropriate asymptotic high-density limit [52][39]. The rationale for this approximation is that in the limit of slowly varying density an exchange-correlation contribution $\epsilon_x(n)dn(\mathbf{r})$ can be attributed to each infinitesimal element in \mathbf{r} , and the total exchange-correlation obtained by quadrature [2][4]. The domain of applicability of LDA has been found, unexpectedly, to go much beyond the nearly free-electron gas, and accurate results can be obtained for very inhomogeneous systems (see Ref. [53]). An explanation for this success [54] comes from the analysis of the properties of the *exchange-correlation hole* $n_{xc}(\mathbf{r}, \mathbf{r}')$: the exchange-correlation energy of the interacting system can be reformulated exactly as

$$E_{xc} = \frac{1}{2} \int d\mathbf{r}n(\mathbf{r}) \int d\mathbf{r}' \frac{1}{|\mathbf{r} - \mathbf{r}'|} n_{xc}(\mathbf{r}, \mathbf{r}' - \mathbf{r}), \quad (2.3)$$

where $n_{xc}(\mathbf{r}, \mathbf{r}')$ is defined in terms of the pair-correlation function $g(\mathbf{r}, \mathbf{r}', \lambda)$ for the system described by a density $n(\mathbf{r})$ but interacting via a *reduced* Coulomb interaction $\lambda \hat{U}_{e-e}$

$$n_{xc}(\mathbf{r}, \mathbf{r}' - \mathbf{r}) \equiv n(\mathbf{r}') \int_0^1 [g(\mathbf{r}, \mathbf{r}', \lambda) - 1] d\lambda. \quad (2.4)$$

The exchange-correlation hole describes intuitively the concept of Pauli repulsion, in that an electron at a position \mathbf{r} reduces the probability of finding one in a position \mathbf{r}' , creating a charge depletion around \mathbf{r} . The exchange-

correlation energy can be regarded as the energy resulting from the interaction of an electron and its surrounding exchange-correlation hole. Since $g(\mathbf{r}, \mathbf{r}', \lambda)$ tends to 1 for an increasing separation $|\mathbf{r} - \mathbf{r}'|$, the exchange-correlation hole is essentially associated with the short-range quantum effects of the Coulomb interaction. Given that the latter is isotropic, it can be demonstrated [54] that the exchange correlation energy depends only on the *spherical average* of $n_{xc}(\mathbf{r}, \mathbf{r}')$ for a given \mathbf{r} ; it has been shown in fact [55] that even if the exact and LDA exchange-correlation holes differ significantly, their spherical averages are similar even for very inhomogeneous systems, and this provides the key to understanding the reasons for the success of LDA in a wide and general variety of environments for the electron gas. Finally, the LDA satisfies exactly the sum rule

$$\int n_{xc}^{LDA}(\mathbf{r}, \mathbf{r}' - \mathbf{r}) = \int n_{xc}(\mathbf{r}, \mathbf{r}' - \mathbf{r}) d\mathbf{r}' = -1, \quad (2.5)$$

which explains some of its accuracy; the necessity of enforcing this sum rule is at the origin of some of the difficulties encountered in improving the LDA by means of perturbative expansions (the most recent and successful development being the generalized gradient approximation (GGA) of Perdew and Wang (PW91, Ref. [48])). The LDA will be used in all the applications presented here, which deal with systems where the charge density varies smoothly and which are within the domain of accuracy of the LDA.

2.2 The pseudopotential approximation

The pseudopotential approximation¹ provides a rational way to isolate the fundamental role played by the valence electrons in the electronic-structure problem. In a mean-field picture, the valence electrons can be thought of as the loosely bound orbitals which are most strongly modified on formation of the chemical bonds in a molecule or the bands in a solid. The length-scales of these orbitals (and energies, since the Schrödinger equation mixes the two via the Laplacian) are those characteristic of inter-atomic separations and chemical bond energies. In contrast, inner electrons are tightly bound around each atomic nucleus and are largely unperturbed by the environment surrounding their atom. In a Hartree-Fock or Kohn-Sham mean-field picture, the higher orbitals must have features on the inner core length-scales via the orthogonality constraint to the lower orbitals. Nevertheless, it can be argued that the atomic problem can be projected for the valence electrons into an effective energy-dependent Hamiltonian, where the nuclear attraction is largely screened by a repulsive term that mimics the effects of the orthogonality constraint; this is the rationale behind the *cancellation theorem* [6]. The resulting potential is commonly referred to as a *pseudopotential* and it is much weaker and smoother than the original Coulombic potential. This weak potential brings in a picture for the electron-ion interaction in solids of an inhomogeneous electron gas weakly perturbed by an array of effective pseudopotentials [7].

¹See Refs. [56] and [57] for some recent and comprehensive reviews.

2.2.1 Basis representation

The pseudopotential approach allows for a representation of the quantities of interest—the charge density and its decomposition into the Kohn-Sham orbitals—that is as general and as economical as possible, since the electronic structure problem can then be solved on a discrete grid in real space where the accuracy on the spacing needs to be only that necessary to follow the oscillations of the pseudo-orbitals and the motion of the pseudo-ions. Methods that operate exclusively in real-space are being introduced nowadays, based either on regular grids [58], or on the variable metrics provided by adaptive coordinates [59], or based on multi-grid techniques [60].

In what follows we will adopt what is presently the method of choice for the description of periodic solids, introducing periodic-boundary conditions [61] that eliminate or reduce the finite size errors in the description of perfect bulk crystals or systems with some degree of periodicity². The induced periodicity in the external potential makes the Hamiltonian operator commute with the set of translation operators identified by the periodic boundary conditions [62]; the set of common eigenstates for these operators is (*Floquet's*, or *Bloch's theorem*):

$$\psi_{\hat{H}}(\mathbf{r}) = \psi_{n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u_{n\mathbf{k}}(\mathbf{r}), \quad (2.6)$$

where \mathbf{k} is the quasi-momentum and $u_{n\mathbf{k}}(\mathbf{r})$ has the periodicity of the unit cell. The infinite wavefunctions of an extended system become represented, with these periodic boundary conditions, by a finite number of wavefunctions for each quasi-momentum \mathbf{k} , spanning the infinite set of points \mathbf{k} inside the

²E.g. surfaces, that can be described by periodically repeated thick slabs in the direction perpendicular to the surface, while keeping the perfect periodicity parallel to it.

Brillouin Zone. The periodicity of $u_{n\mathbf{k}}(\mathbf{r})$ is exploited when representing the wavefunction with the *discrete* basis of plane waves (that are an orthogonal and complete set):

$$\langle \mathbf{r} | \mathbf{k} + \mathbf{G} \rangle = \frac{1}{\sqrt{V}} e^{i(\mathbf{k} + \mathbf{G}) \cdot \mathbf{r}},$$

where V is the volume of the unit cell; additionally, a metric for the completeness can be naturally introduced by selecting the *finite set* of plane waves for which $|\mathbf{k} + \mathbf{G}|^2 \leq E_{cutoff}$. The expectation value for the kinetic energy of a wavefunction represented on a basis of plane waves is calculated efficiently in reciprocal space:

$$\psi_{n\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} c_{n\mathbf{k},\mathbf{G}} e^{i(\mathbf{k} + \mathbf{G}) \cdot \mathbf{r}} \Rightarrow \langle \psi_{n\mathbf{k}}(\mathbf{r}) | -\nabla^2 | \psi_{n\mathbf{k}}(\mathbf{r}) \rangle = \sum_{\mathbf{G}} c_{n\mathbf{k},\mathbf{G}}^2 |\mathbf{k} + \mathbf{G}|^2,$$

while the action of the (local) pseudopotential is expressed in real space; in other words the kinetic-energy operator is diagonal in reciprocal space, and the external potential is diagonal in real space. It is computationally more convenient to calculate these expectation values in the representation for which they are diagonal, switching back and forth from one representation (the Fourier coefficients $c_{n\mathbf{k},\mathbf{G}}$) to the other (the wavefunction $\psi_{n\mathbf{k}}(\mathbf{r})$ on a grid in real space), since efficient algorithms are available (the so-called Fast Fourier Transforms, see Chapter 12 of Ref. [63]) whose cost is roughly proportional to the number N of elements in the mesh to be transformed (more precisely it is $\mathcal{O}(N \ln N)$).

2.2.2 Smoothness and transferability

Simple and fairly accurate models of pseudopotentials were developed initially by choosing some simple analytical form (either in real space, like in

the “empty-core” potential [64], or for its Fourier components, as in the “Empirical Pseudopotential Method” [65]), and then adjusting the parameters in order to fit the atomic energy levels or the band structures. Whilst providing considerable insight for the theory of the electronic structure of solids, these pseudopotentials do not provide the fundamental requirement of *transferability*, in that their ad-hoc fitting procedure does not guarantee a faithful representation outside the configuration space in which the fitting has been performed, and so little success is expected when using these pseudopotential in widely different chemical environments. This generally precludes the ability to perform truly first-principles calculations, that require to determine material properties without the need of any experimental input, and in particular without the need of adopting procedures that are tuned in advance to the specific problem addressed. To improve this accuracy, and given the angular symmetry of the atomic orbitals and the independent orthogonality constraints that they generate, it is natural to develop a scattering approach [7] in which the pseudopotential acts differently on each angular-momentum component of the scattering wavefunction; the pseudopotentials is thus decomposed onto a sum of projections over different angular momenta, and becomes a *non-local* (or semi-local, more accurately in this case) operator

$$V_{ps}(\mathbf{r}, \mathbf{r}') = V^{loc}(r) \delta(\mathbf{r} - \mathbf{r}') + \sum_l V_l^{non-loc}(r) \delta(r - r') P_l(\hat{\mathbf{r}}, \hat{\mathbf{r}}').$$

In addition to this, the modern theory of pseudopotentials builds around the key issue of transferability, constructing a pseudopotential that scatters the incoming wavefunction as closely as possible to the original potential over a wide range of energies. A systematic and successful procedure for the development of accurate and transferable pseudopotentials has been developed [66]

[67], where the added constraint of *norm-conservation* is introduced in order to preserve transferability. In summary, the requirements in constructing the pseudopotential are:

- the lowest eigenvalues for $T + V_{ps}$ must equal the valence all-electron eigenvalues.
- the pseudo-wavefunctions ϕ_l should be nodeless, and they and their first derivatives must be differentiable.
- the pseudo-wavefunctions ϕ_l must be identical with the all-electron wavefunctions beyond a chosen core radius.

or equivalently (norm-conservation)

- the total integrated pseudo-charge density from a given ϕ_l and the corresponding all-electron charge density are identical inside the core radius.

The norm-conservation requirement ensures that the logarithmic derivative of ϕ_l (related to the phase shifts in the scattering) has the same behaviour—up to the first order in changes in energy—as in the all-electron case. Thus the pseudopotential is constructed from the beginning with an intrinsic degree of transferability. An example of a non-local norm-conserving pseudopotential, together with its pseudo-wavefunctions, is given in Fig. 2.1.

The original choice [67] for the pseudopotentials' core radii along the periodic table was conservatively small, thus aiding transferability at the cost of smoothness and computational effort. Several groups have proposed modified schemes for generating pseudopotentials aimed at reducing the number of

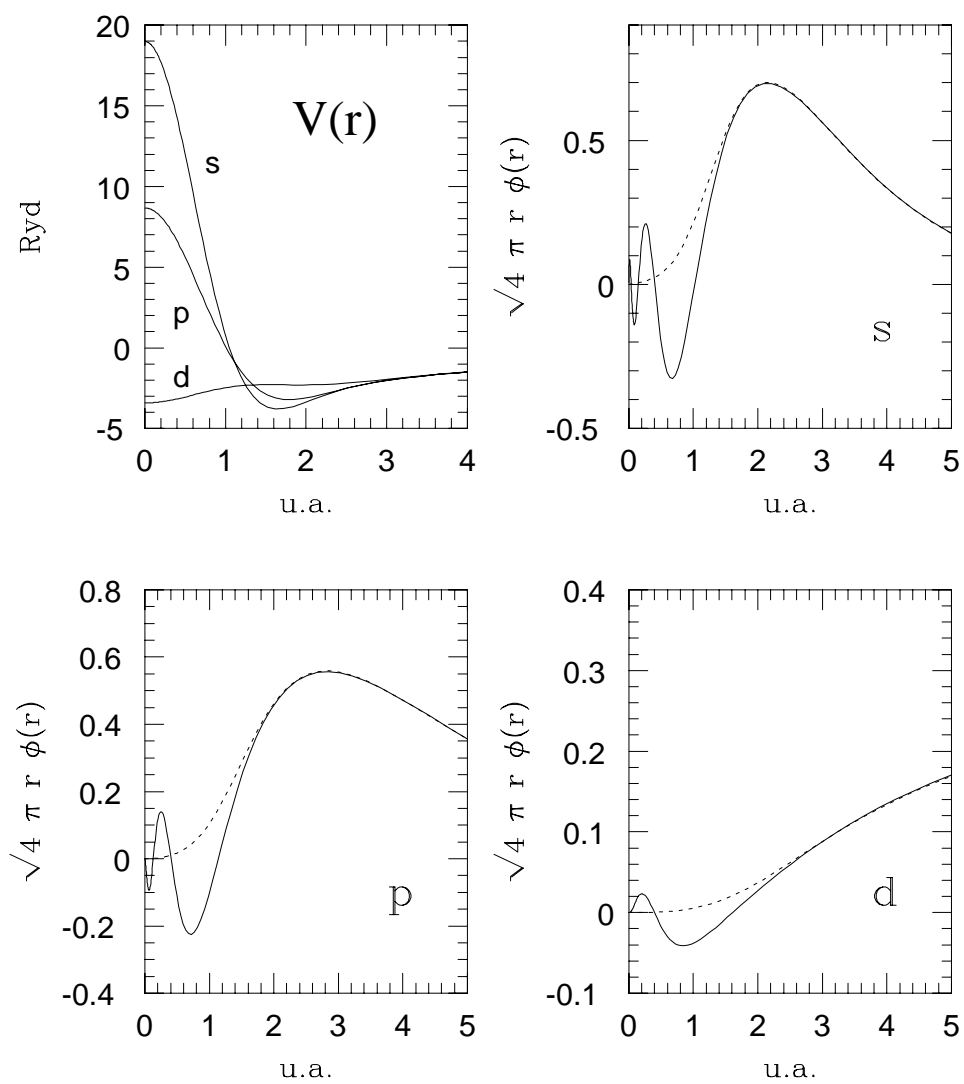


Figure 2.1: s -, p - and d - components of a non-local norm-conserving pseudopotential (In), with the all-electron and pseudo-wavefunctions $r\phi(r)$ (solid and dashed, clockwise from top left).

plane waves required to describe properly the electron-ion interaction, either focusing on producing improved smoother wave-functions in the pseudizing process [68] [69] [70] [71] or smoother potentials [72]. In particular, a formulation has been proposed (the *ultrasoft pseudopotential formulation* [73]) that partially releases the constraint of norm-conservation while it imposes at the same time the matching of the scattering properties on a broader range of energies. This allows for improved transferability *and* much increased smoothness, at the price of introducing a mechanism of *charge augmentations* to restore norm-conservation and hence the proper balance of valence charge density in the core region. In all the applications presented here an optimized pseudopotential for Al (courtesy of M.H. Lee, Ref. [71]) has been used, which was generated along the lines of Refs. [69] [70] [71], with the aim of reducing the residual kinetic-energy of the pseudo-wavefunction beyond a given cutoff in reciprocal space in order to improve smoothness:

$$\Delta E_{kin} = \sum_{i, \mathbf{G}_i > \mathbf{q}_c} \mathbf{G}_i^2 |\phi_i(\mathbf{G}_i)|^2.$$

2.2.3 The Kleinman-Bylander representation

The introduction of non-local or semi-local pseudopotentials makes the plane-wave representation of the Hamiltonian much more expensive to calculate. The kinetic-energy operator $\langle \mathbf{G}_i | \hat{T}_e | \mathbf{G}_j \rangle$ is diagonal in reciprocal space, and thus requires a number of operations proportional to the number of plane waves; the local potential, diagonal in real space, is expressed in reciprocal space via a Fast Fourier Transform, and it depends on \mathbf{G}_i and \mathbf{G}_j only as $(\mathbf{G}_i - \mathbf{G}_j)$ (this also makes the cutoff for the representation of the local pseudopotential twice as large as that necessary for the wavefunctions). The

non-local pseudopotential, on the other hand, is a non-trivial dense matrix in \mathbf{G}_i and \mathbf{G}_j , and computing this expectation value requires a number of operations which is proportional to the *square* of the number of plane waves. A shortcut around this problem is offered by rewriting the pseudopotential in the Kleinman-Bylander representation [74], making it fully non-local but at the same time separable in \mathbf{G}_i and \mathbf{G}_j , and thus making the workload just proportional to the number of plane waves. If $\phi_{lm}(\mathbf{r}) = \chi_l(r)Y_{lm}(\hat{r})$ is an eigenfunction of the atomic pseudo-Hamiltonian, the Kleinman-Bylander representation is defined by:

$$V_{KB}(\mathbf{r}, \mathbf{r}') = V_{KB}^{loc}(r) \delta(\mathbf{r} - \mathbf{r}') + \sum_{lm} \frac{|V_l^{non-loc} \phi_{lm}\rangle \langle \phi_{lm} V_l^{non-loc}|}{\langle \phi_{lm} | V_l^{non-loc} | \phi_{lm} \rangle}.$$

This representation of the potential is different from the original semi-local one, but it can be seen that it acts identically on the reference states $|\phi_{lm}\rangle$, and thus it is conceptually equally well-justified and it is expected to maintain the same degree of smoothness and transferability of the original representation. The fully non-local formulation allows for a completely separable representation in reciprocal space, where the non-local component is given by:

$$V_{KB}^{non-loc}(\mathbf{G}, \mathbf{G}') = \sum_l F_l(\mathbf{G}) F_l^*(\mathbf{G}')$$

and can be computed with a cost that is directly proportional to the number of plane waves. A further improvement towards a more economical representation of the non-local effects of the pseudopotential is obtained by taking advantage of the localization of $V_l^{non-loc}$ in real-space [75], significantly reducing the cost in the limit of very large systems, thanks to the improved scalability.

2.3 Minimizations and dynamics for electrons and ions

The approximations previously described provide an explicit functional form for the exchange-correlation functional of the Kohn-Sham formulation, and an accurate and affordable—at some computational price—form for the interaction between electrons and ions, that acts as external potential for the electron gas. This external potential is fully determined once the nature and coordinates of the pseudo-ions $\{\mathbf{R}_I\}$ are specified; the ground state is then obtained by minimizing the total energy functional (1.33) to self-consistency with respect to the set of orbitals $\{\psi_i\}$. We can consider the total energy of the ground-state as a parametric function of the ionic coordinates $\{\mathbf{R}_I\}$, or in other words as the adiabatic Potential Energy Surface (PES)

$$\mathcal{V}(\mathbf{R}_1, \dots, \mathbf{R}_N) = \min_{\{\psi_i\}} E[\{\psi_i\}; \{\mathbf{R}_I\}].$$

In addition to determining the electronic ground-state properties for a given ionic configuration, it might be important to be able to perform *structural relaxations*, thus determining the local or the global minimum of $\mathcal{V}(\{\mathbf{R}_I\})$ as a function of $\{\mathbf{R}_I\}$, or to perform *molecular dynamics* simulations, where the ions move according to the Newton's equation of motion in the potential field (the PES) determined by the instantaneous electronic ground-state [76].

The total energy depends non-linearly on the orbitals $\{\psi_i\}$, themselves constrained to be orthonormal and bound by the self-consistency requirement to reproduce at the ground state the charge-density dependent Hamiltonian for which they are solutions. In addition, the orbitals have to be represented in a basis of plane waves, and this involves a number of coefficients of the

order of few tens to several hundreds for every atom contained in the unit cell (per “atomic volume”, more appropriately). These requirements result in a problem that is computationally very intensive, and a long and focused effort of the scientific community has been geared to the development of efficient algorithms and techniques tuned to the available resources.

A brief description is given here of the more recent and successful approaches, and on how they can be applied to perform electronic-structure calculations and first-principles structural relaxations or molecular dynamics.

2.3.1 Car-Parrinello molecular dynamics

Car and Parrinello³ [49] proposed to apply the idea of an *extended Lagrangian formulation* [80] [81] to the domain of quantum mechanics: the classical Lagrangian for a system of ions in the potential $\mathcal{V}(\{\mathbf{R}_I\})$,

$$\mathcal{L} = \frac{1}{2} \sum_I M_I \dot{R}_I^2 - \mathcal{V}(\{\mathbf{R}_I\}),$$

is substituted with an effective Lagrangian in which the *electronic* degrees of freedom have been introduced in the form of *classical dynamical variables*, characterized by an appropriate fictitious mass and kinetic energy, and where the potential energy surface is defined in all generality as $E = E[\{\psi_i\}; \{\mathbf{R}_I\}]$. Following these assumptions the Car-Parrinello Lagrangian takes the form:

$$\mathcal{L} = \sum_i \frac{1}{2} \mu \int |\dot{\psi}_i(\mathbf{r})|^2 d\mathbf{r} + \frac{1}{2} \sum_I M_I \dot{R}_I^2 - E[\{\psi_i\}; \{\mathbf{R}_I\}], \quad (2.7)$$

³See Refs. [77] and [78] for comprehensive reviews of the molecular dynamics method, and Ref. [79] for a discussion of the computational efficiency in systems *with* a gap.

with the orbitals subject to the holonomic constraints of orthonormality:

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

The first term in (2.7) is the fictitious kinetic energy associated with the classical motion of the orbitals (more precisely, of the coefficients of their plane-wave representation), that are given a fictitious mass μ whose role is to tune the dynamics of these orbitals with respect to the dynamics of the ionic masses M_I (μ can also vary from orbital to orbital; for the case of fixed ions it can be simply renormalized into the integration time-step). The Euler-Lagrange equations of motion determined by the effective Lagrangian (2.7) are:

$$\mu \ddot{\psi}_i = -\frac{\delta E}{\delta \psi_i^*} + \sum_j \Lambda_{ij} \psi_j \quad (2.8)$$

$$M_I \ddot{\mathbf{R}}_I = -\frac{\partial E}{\partial \mathbf{R}_I}, \quad (2.9)$$

where the Lagrangian multipliers Λ_{ij} have been introduced explicitly to impose the holonomic constraint of orthonormality to the orbitals. It should be underlined that the functional derivative $\frac{\delta E}{\delta \psi_i^*}$ is $\hat{H}_{KS} \psi_i$, where \hat{H}_{KS} is the single-particle Kohn-Sham operator defined in (1.32). Neither Eq. (2.8) nor, in particular, Eq. (2.9), are immediately associated with some real trajectories of the physical system. The role of the present formulation is, at first, to introduce the idea of *simulated annealing* [82] as a practical method to determine the ground state. A procedure can be devised by which some kinetic energy is initially given to the system (either to the electronic degrees of freedom or also to the ions) and then the dynamical evolution in the configuration space is followed according to the Lagrangian dynamics, while the temperature gets slowly decreased towards zero. In the absence

of metastable states that can trap the system⁴ the equilibrium will be finally reached when the temperature has dropped to zero, so that also the fictitious classical kinetic energy of the orbitals and the ions are zero, and the relations $\ddot{\psi}_i = 0$ and $\ddot{\mathbf{R}}_I = 0$ hold. It is readily seen from Eqs. (2.8) and (2.9) that the first condition implies that the orbitals at this point satisfy the self-consistent Kohn-Sham equations (1.32) (a part from a unitary transformation that leaves the Hamiltonian invariant) and consequently the derivatives $\frac{\partial E}{\partial \mathbf{R}_I} = \frac{\partial \mathcal{V}}{\partial \mathbf{R}_I}$ are the true Hellmann-Feynman forces; the second condition implies that the system has reached either a local or possibly a global minimum of the potential energy surface $\mathcal{V}(\{\mathbf{R}_I\})$.

One of the significant improvements that such a formalism brings in [83] is that it does not require to ever store the full Hamiltonian matrix, which is very large in a plane-wave based formulation, since it can take full advantage of the existence of (different) diagonal representations for the kinetic energy and the pseudopotential, switching back and forth from real to reciprocal space with Fast Fourier Transforms. For this reason and since only the lowest eigenstates are needed, the evolution towards the ground state (requiring the repeated application of the operator \hat{H}_{KS} to all the wavefunctions) has an associated cost that is much lower than an iterative sequence of full diagonalizations of the Hamiltonian.

This same Car-Parrinello Lagrangian can be subtly exploited to perform some realistic molecular dynamics: if the mass associated with the electronic

⁴This is not a problem for the annealing at fixed ions, although if the ions are simultaneously allowed to relax it can be very easy to get trapped into a local ionic minimum. The procedure itself is meant to reach eventually the *global* minimum, but it requires an asymptotically slow decrease of the temperature that is not practically affordable.

motion is chosen to be much smaller than that of the ions, for appropriate initial conditions the system will be characterized by a metastable separation between a subsystem of fast oscillators (the orbitals) that have only a tiny fraction of the total (ionic and electronic) kinetic energy associated with them, and a subsystem of slow and massive ions that act adiabatically on the electrons as a driving potential. The equations of motion (2.8) favour those natural oscillations of the orbitals that are coherent with the instantaneous ionic motion; the higher frequencies and smaller fictitious mass(es) of the electronic variables assure that, in an average sense, the orbitals will keep redistributing in order to be always close to the real ground-state, parametrically evolving as a function of the ionic coordinates, and with a tolerance in the thickness of their Born-Oppenheimer confinement that is essentially determined by the fictitious electronic kinetic energy. The trade-off in the simulations is between having smaller electronic masses, thus keeping the system closer to the Born-Oppenheimer minimum and characterized by more responsive electronic degrees of freedom, and the necessity of keeping the time-step for the expensive integration of the equations of motion as large as possible, in order to simulate the dynamics for a time adequate to the atomic scales (typically from a fraction of a picosecond up to several picoseconds). Although the forces that drive the ionic dynamics are never the exact Hellmann-Feynman forces, the adiabatic separation assures that, in an average sense, their trajectory will be close to the exact one, the more so the smaller the electronic mass (and the time step) associated with the quasi-equilibrium electron dynamics.

It should be pointed out [84] that the eigen-frequencies of the oscillations

for the electronic degrees of freedom are, at fixed ions:

$$\omega_{ij} = \left(\frac{2 (\epsilon_j - \epsilon_i)}{\mu} \right)^{\frac{1}{2}},$$

where ϵ_j is a Kohn-Sham eigenvalue for a higher unoccupied state and ϵ_i is an eigenvalue for an occupied state. Thus, the lowest frequency that can appear in the electronic motion is of the order of $\omega_{min} = (2E_{gap}/\mu)^{1/2}$; for a semiconductor or insulator, with a suitable choice of μ this critical frequency can be higher than the typical frequencies of the ionic motion, and thus the condition of adiabatic separation can be satisfied for all the duration of a typical simulation (this technical detail depends on the size of the gap, and on the condition that the gap is not reduced following disordering or reconstruction). For very long simulation times, the electrons can be periodically quenched back to the Born-Oppenheimer ground-state [85], although this procedure destroys the proper evolution of the holonomic constraints [84], that requires a coherent evolution in the subspace of occupied orbitals (the total energy is invariant with respect to a unitary transformation in this subspace, but the fictitious kinetic energy is not, unless the unitary transformation is time-independent). In the case of a metal, the gap tends to zero in the limit of the periodic unit cell going to infinity (or the sampling of the Brillouin-zone becoming increasingly fine), and the metastability of the adiabatic separation breaks down immediately.

2.3.2 Conjugate-gradient minimization

A straightforward and very efficient approach can be taken by reconsidering the original problem of the functional minimization for the total energy

functional

$$E[\{\psi_i\}; \{\mathbf{R}_I\}] = \sum_{i=1}^N -\frac{1}{2} \int \psi_i^*(\mathbf{r}) \nabla^2 \psi_i(\mathbf{r}) d\mathbf{r} + E_H[n(\mathbf{r})] + \\ + E_{xc}[n(\mathbf{r})] + \int v_{ext}(\mathbf{r}; \mathbf{R}_1, \dots, \mathbf{R}_N) n(\mathbf{r}) d\mathbf{r} \quad (2.10)$$

and treating it as a complex problem of non-linear constrained optimization, either in the electronic variables alone, or together with the ionic degrees of freedom. The electronic problem will be discussed first, i.e. the problem of minimizing the total energy as a function of the coefficients of the plane waves on which the orbitals are represented. For this case, the total energy functional possesses a single, well-defined minimum, but it has a very large number of degrees of freedom and a fairly complex set of non-linear constraints.

There are two well-established analytical methods that exploit optimally the knowledge of the function to be minimized and of its first derivatives to map out improved directions towards the multi-dimensional minimum: these are respectively the *conjugate-gradient methods* (CG, either in the Fletcher-Reeves or Polak-Ribiere formulation) and the *variable-metric methods* (often found in the BFGS formulation of Broyden-Fletcher-Goldfarb-Shanno) [63]. Both use the knowledge of the multi-dimensional function and—if available—of its first derivatives to iteratively produce more efficient multi-dimensional search directions; they share the concept that it is the information on the second-derivatives, in the neighborhood of a multi-dimensional minimum, the obvious fundamental quantity that locates the position of the minimum with respect to the points that have already been sampled. The BFGS minimization operates by explicitly accumulating an estimate of the inverse Hessian matrix, while the CG method takes the indirect approach of reducing the di-

dimensionality of the search space at every iteration. The conjugate-gradient algorithm needs only a storage proportional to the number of dimensions in the full search space (i.e. the number of plane waves), while variable-metrics methods require a storage proportional to the square of the number of dimensions, and for this very reason have to be ruled out from the beginning in the context of plane-wave electronic structure calculations. In addition, there is no clear computational advantage for either method that can be generally foreseen in advance, both relying for their efficiency on the information that is, directly or indirectly, gained on the curvature of the function that is being minimized.

The ideas beneath the conjugate-gradient algorithm can be exemplified⁵ considering the problem of minimizing in the N dimensions a symmetric positive-definite function $F = \frac{1}{2}\mathbf{x} \cdot G \cdot \mathbf{x}$, where G is the gradient operator such that the gradient at \mathbf{x} is $\mathbf{g} = -G\mathbf{x}$. Starting from a generic point \mathbf{x}^1 , the best initial direction is the steepest-descent direction $\mathbf{d}^1 = \mathbf{g}^1 = -G\mathbf{x}^1$; a minimization along this multidimensional line yields $\mathbf{x}^2 = \mathbf{x}^1 + b^1\mathbf{d}^1$, where b^1 , being the minimum of F along \mathbf{d}^1 , is such that

$$\frac{dF}{db^1} = (\mathbf{x}^1 + b^1\mathbf{d}^1) \cdot G \cdot \mathbf{d}^1 = 0.$$

The straightforward application of the steepest-descent method would call now for a subsequent minimization from \mathbf{x}^2 along the direction of steepest-descent $\mathbf{d}^2 = \mathbf{g}^2 = -G\mathbf{x}^2$. This would lead to a new point along \mathbf{d}^2 , $\mathbf{x}^3 = \mathbf{x}^2 + b^2\mathbf{d}^2$, such that, as above,

$$\frac{dF}{db^2} = (\mathbf{x}^1 + b^1\mathbf{d}^1 + b^2\mathbf{d}^2) \cdot G \cdot \mathbf{d}^2 = 0.$$

⁵For this, and for an in-depth review of the total-energy pseudopotential method and its implementation as a direct minimization problem, see Ref. [50].

At this point, the problem can be reexamined; in fact, given *both* the optimal search directions \mathbf{d}^1 and \mathbf{d}^2 , the minimization problem can be regarded as a global two-dimensional problem in the space of \mathbf{d}^1 and \mathbf{d}^2 . The optimal b^1 and b^2 are those that satisfy the two coupled equations:

$$\left. \frac{dF}{db^1} \right|_{\mathbf{x}^3} = (\mathbf{x}^1 + b^1 \mathbf{d}^1 + b^2 \mathbf{d}^2) \cdot G \cdot \mathbf{d}^1 = 0$$

$$\left. \frac{dF}{db^2} \right|_{\mathbf{x}^3} = (\mathbf{x}^1 + b^1 \mathbf{d}^1 + b^2 \mathbf{d}^2) \cdot G \cdot \mathbf{d}^2 = 0.$$

These equations are clearly valid if the directions \mathbf{d}^1 and \mathbf{d}^2 are *conjugate* to each other, that is if $\mathbf{d}^1 \cdot G \cdot \mathbf{d}^2 = \mathbf{d}^2 \cdot G \cdot \mathbf{d}^1 = 0$.

This requirement can be iteratively extended [86], and it defines a sequence of search directions, starting from the initial steepest-descent one, that do identify analytically the minimum of a N -dimensional quadratic function in exactly N iterations; the search directions are given by:

$$\mathbf{d}^i = \mathbf{g}^i + \gamma^i \mathbf{d}^{i-1}, \quad (2.11)$$

where the mixing factor between the current gradient and the previous search direction, outcome of all the previous history of the minimization, is

$$\gamma^i = \frac{\mathbf{d}^i \cdot \mathbf{d}^i}{\mathbf{d}^{i-1} \cdot \mathbf{d}^{i-1}}, \quad \gamma^1 = 0. \quad (2.12)$$

The method operates by mapping out from the minimization process all the explored degrees of freedom, effectively reducing the dimensionality of the search space at every iteration. A clear example comes from visualizing a 2-dimensional paraboloid whose principal axes have very different eigenvalues; a steepest-descent strategy, starting from a generic point, would proceed via an infinite series of damped searches approaching asymptotically

the minimum (each steepest-descent direction is bound to be orthogonal to the previous one), while the conjugate-gradient algorithm reaches the minimum in two iterations, by choosing the optimal steps and directions in the combined subspace. Finally, it should be noted that, if the principal axes of the quadratic form are all equal (i.e. the eigenvalues are all identical), the descent direction is always directed towards the minimum, and one iteration is all what is needed, independently from the dimensionality, to conclude the minimization process. This illustrates the key idea of *preconditioning*, that consists of adapting the metrics of the space in which the search directions are chosen in order to make all the eigenvalues be as similar as possible (or, in other words, to make the quadratic form look as spherical as possible).

The conjugate-gradient/steepest descent method has been introduced in the context of the electronic structure of solids both as an indirect minimization technique [87] and as a direct one [88] [44]. The former implementation considers the minimization problem, similarly to the molecular dynamics approach, as a propagation in (imaginary) time, where an initial arbitrary trial state gets projected more and more accurately to the ground state by a repeated application of the operator \hat{H}_{KS} . In this, it suffers inevitably from an intrinsic instability with the growth of the system size, originating from the small wavenumber terms that appear in the denominator of the Hartree energy, and that can be controlled only by reducing a priori the time step of the fictitious evolution. The latter approach operates differently, in introducing the concept of direct minimization by taking into account the *response* of the system when a line search is performed, and choosing a variable step that minimizes the actual value of the total energy along the search direction. This

has the net effect of making the process strictly variational and very stable, in addition to greatly improving the efficiency of the minimization. This approach has been extensively developed [50] [89] [90] and successfully applied to a wide variety of semiconducting and insulating systems [91] [92] [93] [94] [95] [96] [97]. In its original formulation [88] [50] the optimization strategy was broken down into a *band-by-band* updating strategy, in order to minimize the memory requirements for the working arrays containing the gradients, the conjugate directions, and the current wavefunctions. The strategy adopted here and in the following will be an *all-bands* approach [44], that offers a much improved performance, especially in the case of metallic systems (since the cross-terms between the bands are considered in the response), together with a simpler theoretical formulation, albeit at the cost of a significant increase in the dynamic memory requirements. Given that it is currently possible to fit on a common workstation systems whose size is of the order of one hundred atoms, provided they do not need hard pseudopotentials, and given the trend of large-scale electronic structure computations towards distributed memory machines, the overhead costs associated with an all-bands scheme can be considered to be paid off by its increased performance (by at least a factor of 2 or 3, although it can be much more than that for “difficult” systems, like the very elongated cells discussed below). As an aside it should be noted that a single conjugate-gradient iteration on all the orbitals, in a band-by-band scheme, requires a much larger computational effort than one iteration in an all-bands scheme (due to the requirements of orthogonality and preconditioning along every band-by-band search), and so iteration counts are not immediately meaningful for a performance comparison. The technical details

of the implementation, specifically in a metallic system, will be sketched in the following Section and in Chapter 3, where the ionic relaxations are also discussed. As a conclusion to the points presented here, the more relevant features of conjugate-gradient methods are outlined in the following, focusing the perspective on applying the technique to large-scale or traditionally difficult cases (metallic systems being paradigmatically difficult), or to dynamical simulation, especially where the nature of the system changes during the simulation.

Conjugate-gradient direct minimization: an outline

- it is an iterative procedure driven by the functional derivatives (gradients) of the total energy $\frac{\delta E}{\delta \psi_i^*} = \hat{H}_{KS} \psi_i$, and so it shares with the Car-Parrinello method the advantages in evaluating and storing the Kohn-Sham Hamiltonian
- in the limit of a proper preconditioning formulation, it is heuristically known to converge very fast even with respect to the theoretical asymptotic behaviour [98]
- it is robust and stable, ultimately based on the directions of (steepest, conjugate) descent, and it is naturally implemented in a strictly variational fashion
- it allows the direct minimization of the self-consistent functional, invariably with just a single interpolation along the line of descent, and thus it employs the optimal (imaginary time) step at every iteration

2.3.3 Metallic systems

Severe technical difficulties arise in the treatment of metallic systems, both in static self-consistent calculations and in molecular dynamics simulations. To begin with, static calculations show a dramatic decrease of the sampling accuracy in the Brillouin Zone.

Brillouin-Zone sampling In periodic boundary conditions the translational symmetry of the external potential leads to a representation for the wavefunctions in the form of Bloch states (2.6); the sum on all the infinite states of the system (or quasi-infinite, with the Born-von Karman conditions for the wavefunctions) thus becomes an integral over the first Brillouin Zone (or a sum, done on the discrete fine mesh of wavevectors allowed by the Born-von Karman conditions), where for each wavevector there corresponds only a finite and small number of states (for a spin-degenerate calculation in an insulator, there are $\frac{N_e}{2}$ states occupied at each wavevector, with N_e the number of electrons in the unit cell). The kinetic energy, the band energy term and the charge density in (1.33), (1.34) are all defined via a full integration on the first Brillouin Zone; e.g. the charge density is

$$n(\mathbf{r}) = \frac{1}{V} \int_{BZ} \left\{ \sum_i f(\epsilon_{i\mathbf{k}}) |\psi_{i\mathbf{k}}(\mathbf{r})|^2 \right\} d\mathbf{k}, \quad (2.13)$$

where $f(\epsilon_{i\mathbf{k}}) = 2$ (if the system is spin degenerate) for the lowest $\frac{N_e}{2}$ states at each \mathbf{k} , 0 otherwise.

It was first recognized by Baldereschi [99] that the integrations over the Brillouin Zone could be performed to a very high accuracy by using just a single *special k-point*, carefully chosen according to the point group symmetry of the unit cell lattice. This analysis has been subsequently generalized

to identify a set of prescriptions for choosing relatively coarse meshes that can provide excellent accuracy in estimating the exact integrals [100] [101] [102]; the existence of symmetry operators that commute with the Hamiltonian can be further exploited via the simple relations that are satisfied by the common eigenstates at different but symmetry related \mathbf{k} -points, and effectively reducing the problem to an integration over the *irreducible wedge* of the Brillouin Zone (e.g. the time-reversal symmetry brings the relation $\psi_{\mathbf{k}}(\mathbf{r}) = \psi_{-\mathbf{k}}^*(\mathbf{r})$, straightforwardly halving the number of \mathbf{k} -points that are needed). The success of these integration techniques is analogous in nature to that of the well-known Gaussian quadrature, in that they rely on integrating accurately the lowest harmonics of the representation of the integrand, which in this case is naturally expanded as a series on the symmetrized lattice stars.

This accuracy degenerates immediately if the function to be integrated is discontinuous. This is precisely what happens in a metal, where only the states inside the Fermi surface contribute to the integrals, and the occupations drop to zero when the Fermi surface is crossed. A partial solution to this problem has been obtained by introducing a smearing technique [103] [104]: the exact density of states $n(\epsilon)$ (not to be confused with the charge density $n(\mathbf{r})$),

$$n(\epsilon) = \sum_{i\mathbf{k}} \delta(\epsilon - \epsilon_{i\mathbf{k}}), \quad (2.14)$$

is substituted with a smoother, smeared density of states

$$n(\epsilon) = \sum_{i\mathbf{k}} \frac{1}{\sigma} \tilde{\delta}\left(\frac{\epsilon - \epsilon_{i\mathbf{k}}}{\sigma}\right), \quad (2.15)$$

where the $\tilde{\delta}$ function is some broadened approximation to the original Dirac's delta (e.g. a Gaussian). The immediate effect of this broadening is to remove

the sharp discontinuity around the Fermi energy, and to *greatly* improve the sampling accuracy. This approach will be rationalized in the discussion of Chapter 4, with a careful examination of the sampling errors that the approach can cure and the systematic errors that are eventually introduced, and how they can be corrected for; for the time being it should be stressed that it is absolutely equivalent to introducing a (real or fictitious) temperature for the electronic degrees of freedom of the system.

Finite temperature formulation A finite temperature formulation in which the system is characterized by a total free energy functional, as in (1.45), (1.46), comes initially from this necessity of improving the sampling accuracy in the Brillouin Zone; the approach is actually very successful, because it offers at the same time a vastly improved sampling precision and it allows an a-posteriori rationalization of the smearing scheme that can help in correcting the systematic errors that arise from the introduction of a broadened density of states (see Chapter 4 for a discussion). Additionally, a finite temperature formulation helps in controlling the discontinuities in the energy derivatives that would be introduced by level-crossing events, i.e. by the sudden occupation or emptying of states whose energies cross the Fermi level during the iterative process towards self-consistency. Finally, a canonical formulation provides a natural way to include in the minimization process fractional occupancies and subspace rotations: in a metal, at variance with the case of a semiconductor, the charge density (2.13) and consequently the self-consistent Hamiltonian are no longer invariant with respect to unitary transformations of the occupied orbitals, which thus enter the problem and have to be considered as additional non-linear degrees of freedom.

Molecular dynamics The objective of performing stable molecular dynamics simulations in a metallic system raises additional difficulties. Unless the system size is tuned to be small enough to allow for the appearance of a *pseudo-gap* in the system [105], there is otherwise a significant overlap between the typical ionic frequencies and the electronic excitations. This breaks the adiabatic separation that is needed to perform a Car-Parrinello simulation, *even more* so if fractional occupancies and subspace rotations are introduced in the problem [106]. Ad-hoc procedures can be devised to work around this problem, e.g. by repeatedly quenching the system to the ground state [107] while coupling the ions with a Nosé thermostat [108] (provided that the system is quasi-semiconducting in character), or by forcibly coupling a second Nosé thermostat to the electronic system [109] [110], at the price of introducing in the process velocity-dependent forces in the electronic equations of motion. In both cases the treatment of the electronic degrees of freedom remains somewhat unsatisfactory [77] or restrictive.

Additionally, the treatment of the occupation numbers and subspace rotations has to be considered as an essential part in the development of an optimal minimization strategy [44] [111] [112] [113], since the non-linear constraints on these degrees of freedom, together with their intrinsic non-local character⁶, makes the control of the convergence for the energy extremely demanding even in the limit of moderate cell-sizes [114] [106][115]. This problem is even more severe in the convergence of the Hellmann-Feynman forces, whose errors are first-order in the distance between the true ground state and the trial solution, as opposed to the total (free) energy, that has

⁶A small modification of an occupation, or a subspace rotation, amounts to a non-local displacement of charge density all over the unit cell

an error that is of second-order.

Finally it should be noted that the Hellmann-Feynman theorem implies that the forces calculated in a simulation are the true derivatives of the total free energy (that is the smeared energy functional); this explains the observation of an improved conservation of the constant of motion [116] when free energies and Hellmann-Feynman forces are considered. It should be noted from now on that, unless the temperature is introduced in order to properly consider a finite-temperature Fermi-Dirac electron gas and its relative isothermal dynamics⁷, a systematic error arises in adopting the free energies and forces as an approximation to the zero-temperature (zero-smearing) limit. The errors on the energies can be corrected very easily a-posteriori [44] [117]: this allows for the choice of larger smearing widths for a given tolerance in the systematic errors, in turn allowing for accurate sampling of the Brillouin Zone with coarser grids, or for an improved accuracy at a given smearing. In order to perform consistent dynamics, the systematic smearing errors on the forces should also be corrected (this problem will be discussed, together with the solution that is here introduced, in Chapter 4; some of the relevant concepts can be found in Refs. [44] [117] [118] [119]).

The sampling technique of special \mathbf{k} -points, together with the smearing of the density of states, provides also the relevant advantage of producing consistent free energies and forces, in contrast to the improved tetrahedron methods [120], and of requiring only one single special \mathbf{k} -point in the limit of large unit cells (this will be the case for most of the results presented here).

⁷E.g. to characterize the electric conductivity of systems with small gaps, or to study processes in which the electron gas is at a much high temperature than the ionic lattice, like in laser irradiation.

Variational free energy minimization The goal of performing molecular dynamics simulations requires the development of a minimization algorithm that is both very efficient (in order to control the convergence of the Hellmann-Feynman forces, in the presence of the additional degrees of freedom of the fractional occupancies and subspace rotations) and very robust and stable (to converge to the minimum against all the ill-conditioned features of the problem). These requirements call for a minimization strategy realized by a *strictly variational conjugate-gradient method*, that is an algorithm that can ideally provide both the efficiency and, especially, the robustness that is necessary.

Its formulation requires the identification of the proper variational functional that has to be minimized, and that is the *electronic free energy* [either in a Fermi-Dirac or in a *generalized entropy* formulation (see Chapter 4), the latter being equivalent to a smeared total energy formulation [103] [119]], and the proper consideration of the dependence of the total free energy on the fractional occupancies and subspace rotations. The occupancies, in particular, can be considered either as independent variables, with all the instability coming from their non-local connection to the charge density in the unit cell, or as dependent variables, where they introduce in principle a very complicated dependence of the free energy as a function of the wavefunctions (that determine the expectation values $\langle \psi_i | \hat{H} | \psi_i \rangle$ upon which the occupancies are calculated).

The total free energy at a finite temperature T is given by (1.45); it is useful to regroup the kinetic-energy term and the non-local pseudopotential in order to highlight the dependence of the various terms on the wavefunctions

or just on the charge density:

$$A [T; \{\psi_i\}, \{f_i\}] = \sum_i f_i \langle \psi_i | \hat{T}_e + \hat{V}_{nl} | \psi_i \rangle + E_{\text{Hxc}}[n] - TS[\{f_i\}] \quad (2.16)$$

$$n(\mathbf{r}) = \sum_i f_i \psi_i^*(\mathbf{r}) \psi_i(\mathbf{r})$$

where S is either the Fermi-Dirac entropy or one of its generalizations, and where the constraints of normalization of the orbitals and charge conservation for the occupancies have also to be imposed (the latter introduces a chemical potential μ in the problem). These two constraints can be introduced via the respective Lagrange multipliers, as in (1.37), adding to (2.16) the terms

$$\sum f_i \epsilon_i (\langle \psi_i | \psi_i \rangle - 1) \quad \text{and}$$

$$\mu (N - \sum f_i).$$

Assuming that both the wavefunctions and the occupations are independent constrained variables, the steepest descent directions that lead to the minimization of the total free energy functional are given by:

$$\left\{ \begin{array}{l} -\frac{\delta A}{\delta \psi_i^*} \quad \left(\Rightarrow \hat{H} \psi_i = \epsilon_i \psi_i \right) \\ -\frac{\partial A}{\partial f_i} \quad \left(\Rightarrow \langle \psi_i | H | \psi_i \rangle - \mu = T \frac{\partial S}{\partial f_i} \right) \\ \text{subspace rotations} \end{array} \right. \quad (2.17)$$

where the additional operation of subspace rotation has to be added separately, since the search directions are projected out of the occupied subspace to conserve orthonormality (to first-order). It should be noted that at self-consistency, when all the first derivatives are zero, the first equation implies

that the orbitals satisfy the Mermin-Kohn-Sham equations (1.44), while the equilibrium distribution for the occupations is given by the second equation, and it is a Fermi-Dirac distribution (1.43) if the canonical form for the single-particle entropy (1.42) has been employed.

The equations in (2.17) define a multi-dimensional search direction, and a variational minimization strategy can be practically implemented following the prescriptions of Ref. [44] (see also Ref. [121]). In particular, the search on the wavefunctions is naturally decomposed into a search in the orthogonal and in the parallel subspaces (the latter being defined as the space spanned by the orbitals that have non-zero occupancies, and which are just a fraction of the total number of plane waves). The search in the orthogonal subspace is performed via a conjugate-gradient algorithm, while the search in the parallel subspace takes place via a subspace rotation. The rotation matrix can be defined with the help of perturbation theory, and constructed in order to always satisfy the constraint of lowering the total free energy [44]: if a transformation in the parallel subspace is defined by

$$\psi_i \Rightarrow \psi_i + \sum_j W_{ij} \psi_j,$$

the requirement of preserving orthonormality up to linear order in the rotation requires to W_{ij} to be anti-hermitian. Perturbation theory provides an explicit expression for W_{ij} ($W_{ii} = 0$ from the requirement of being anti-hermitian)

$$W_{ij} = \frac{\langle \psi_i | \hat{H} | \psi_j \rangle}{\langle \psi_i | \hat{H} | \psi_i \rangle - \langle \psi_j | \hat{H} | \psi_j \rangle}, \quad (2.18)$$

and from this the actual rotation matrix is defined by $\zeta_{ij} W_{ij}$, where ζ_{ij} acts as a cutoff if the elements in (2.18) are too large (that is to say, if the mixing between states is too big and perturbation theory does not apply), and is set

to zero if a reverse ordering is found (i.e. if two states have occupations and expectation values in reverse order). This latter condition forces the subspace rotation to be a proper minimization step, in the limit of small rotations.

This formulation, properly implemented, allows for a very robust, if not optimal, minimization strategy; as an example, the convergence of the total free energy for two very long metallic slabs is shown in Fig. 2.2. The case presented here is a paradigmatically difficult one, since the cells are very elongated, and the so-called *sloshing instabilities* can take over if the minimization procedure is not strictly variational. As it can be seen, the fundamental requirement of robustness, that comes from the variational formulation, is fully satisfied. It should be noted that an unrealistic (Gaussian) smearing of 4.5 eV has been adopted, to assure full sampling convergence with just one special point, but especially to underline the role played by a proper treatment of the temperature.

This scheme, called here the *standard scheme*, is thus composed of a sequence of variational steps in the wavefunctions, in the occupancies and in the subspace rotations. The order of these operations, as well as performing some or all of them simultaneously, is a matter of choice; it should be noted that an all-bands conjugate-gradient scheme is much more efficient than a band-by-band scheme, for difficult metallic systems, due to the interdependence of the bands. For the same reason it can be useful to perform the updating in more than one subspace at the same time (usually updating the orthogonal subspace and the occupations together), although sometimes a mixing factor between the different subspaces has to be empirically introduced. Such formulation of the standard scheme can be applied successfully

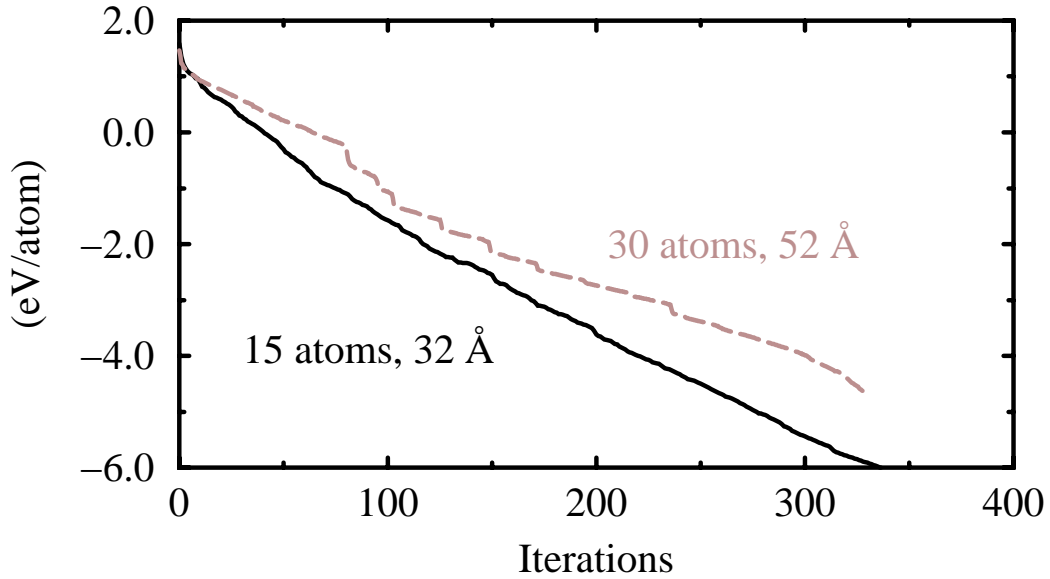


Figure 2.2: Convergence of the total free energy in a semi-logarithmic scale (base 10). The systems considered are two 1x1 Al(110) slabs, containing respectively 15 and 30 atoms (64 and 128 bands, 4.5 eV of Gaussian smearing), and 32 Å and 52 Å long, both sampled with the $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ Baldereschi point [99]. The iteration count for the longer cell has been halved, for a meaningful comparison.

to molecular dynamics simulations, especially if the systems considered are bulk or liquid metals, or if they have a mixed metallic/covalent behaviour [122] [123] [124] [125], so that either the charge density is relatively constant over the unit cell (this is not the case for metallic surfaces, represented as a slab in periodic boundary conditions) or it is more bound in the localized

covalent bonds. It should be pointed out that the formulation presented here suffers from overall poor convergence of the Hellmann-Feynman forces, which severely hampers the ability to perform efficient simulations in very large cells; additionally, the ionic forces used are the derivative of the free energy, and thus are not corrected against the systematic errors that come from a finite-temperature/smeared formulation.

Bibliography

- [1] E. Schrödinger, *Ann. Physik* **79**, 361, 489 (1926); *ibid.* **81** 109 (1926).
- [2] P. Hohenberg and W. Kohn, *Phys. Rev.* **136**, B864 (1964).
- [3] N. D. Mermin, *Phys. Rev.* **137**, A1441 (1965).
- [4] W. Kohn and L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).
- [5] J. S. Bell, *Speakable and Unsayable in Quantum Mechanics: Collected Papers on Quantum Philosophy*, Cambridge University Press, Cambridge (1987).
- [6] J. C. Phillips, *Phys. Rev.* **112**, 685 (1958); J. C. Phillips and L. Kleinman, *ibid.* **116**, 287 (1959).
- [7] M. L. Cohen and V. Heine, in *Solid State Physics* **24**, eds. H. Ehrenreich, F. Seitz and D. Turnbull, Academic Press, New York (1970).
- [8] M. Born and V. Fock, *Z. Phys.* **51**, 165 (1928).
- [9] F. J. Dyson, *J. Math. Phys.* **8**, 1538 (1967); F. J. Dyson and A. Lenard, *ibid.* **8**, 423 (1967).
- [10] E. H. Lieb, *Rev. Mod. Phys.* **48**, 553 (1976).

- [11] R. Courant and D. Hilbert, *Methods of Mathematical Physics*, p. 424, Interscience Publishers, New York (1953).
- [12] P. A. M. Dirac, *The Principles of Quantum Mechanics*, p. 45, Oxford University Press, Oxford (1958).
- [13] L. I. Schiff, *Quantum Mechanics*, McGraw-Hill, Singapore (1968).
- [14] D. R. Hartree, *Proc. Cambridge Phil. Soc.* **27**, 89 (1927); *ibid.* 111, **27** (1927); *ibid.* **27**, 426 (1928).
- [15] V. Fock, *Z. Phys.* **61**, 126 (1930); *ibid.* **62**, 795 (1930); J. C. Slater, *Phys. Rev.* **35**, 210 (1930).
- [16] H. Hellmann, *Einführung in die Quantenchemie*, Deuticke, Leipzig (1937); R. P. Feynman, *Phys. Rev.* **56**, 340 (1939).
- [17] P. Ehrenfest, *Z. Phys.* **45**, 455 (1927).
- [18] L. H. Thomas, *Proc. Cambridge Phil. Soc.* **23**, 542 (1927).
- [19] E. Fermi, *Rend. Accad. Lincei* **6**, 602 (1927); *ibid.* **7**, 342 (1928); *Z. Phys.* **48**, 73 (1928).
- [20] E. Teller, *Rev. Mod. Phys.* **34**, 627 (1962).
- [21] E. Smargiassi and P. A. Madden, *Phys. Rev. B* **49**, 5220 (1994).
- [22] B. Farid, private communication.
- [23] M. Levy, *Proc. Natl. Acad. Sci. USA* **76**, 6062 (1979).
- [24] M. Levy, *Phys. Rev. A* **26**, 1200 (1982).

- [25] E. H. Lieb, *Int. J. Quantum Chem.* **24**, 243 (1983).
- [26] W. Kohn, *Phys. Rev. Lett.* **51**, 1596 (1983).
- [27] H. Englisch and R. Englisch, *Phys. Stat. Sol. (b)* **123**, 711 (1984); *ibid.* **124**, 373 (1984).
- [28] T. L. Gilbert, *Phys. Rev. B* **12**, 2111 (1975).
- [29] L. D. Landau and E. M. Lifshitz, *Statistical Physics* Par. 5, Pergamon Press, Oxford (1980).
- [30] R. G. Parr and W. Yang, *Density Functional Theory of Atoms and Molecules*, Oxford University Press, Oxford (1988).
- [31] S. M. Valone, *J. Chem. Phys.* **73**, 1344 (1980); *ibid.* **73**, 4653 (1980).
- [32] J. Coleman, *Rev. Mod. Phys.* **35**, 668 (1963).
- [33] R. P. Feynman, *Statistical Mechanics*, Addison-Wesley, Reading Massachusetts (1972).
- [34] M. M. Valiev and G. W. Fernando, *Phys. Rev. B* **52**, 10697 (1995).
- [35] M. Levy and J. P. Perdew, in *Density Functional Methods in Physics*, R. M. Dreizler and J. da Providencia eds., Plenum, New York (1985); J. P. Perdew and M. Levy, *Phys. Rev. B* **31**, 6264 (1985).
- [36] E. H. Lieb, in *Density Functional Methods in Physics*, R. M. Dreizler and J. da Providencia eds., Plenum, New York (1985).
- [37] E. H. Lieb, *Rev. Mod. Phys.* **53**, 603 (1981).

- [38] J. F. Janak, *Phys. Rev. B* **18**, 7165 (1978).
- [39] J. Perdew and A. Zunger, *Phys. Rev. B* **23**, 5048 (1981).
- [40] M. Weinert and J. W. Davenport, *Phys. Rev. B* **45**, 13709 (1992).
- [41] J. Callaway and N. H. March, in *Solid State Physics* **38**, p. 235, eds. F. Seitz and D. Turnbull, Academic Press, New York (1984).
- [42] K. Huang, *Statistical Mechanics*, John Wiley & Sons, New York (1987).
- [43] W. Kohn and P. Vashista, in *Theory of the Inhomogeneous Electron Gas*, S. Lundqvist and N. H. March eds., Plenum, New York (1983).
- [44] M. J. Gillan, *J. Phys.: Condens. Matter* **1**, 689 (1989).
- [45] A. Alavi, J. Kohanoff, M. Parrinello, and D. Frenkel, *Phys. Rev. Lett.* **73**, 2599 (1994).
- [46] W. Yang, *Phys. Rev. A* **38**, 5504 (1988).
- [47] D. M. Ceperley and B. J. Alder, *Phys. Rev. Lett.* **45**, 566 (1980).
- [48] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, *Phys. Rev. B* **46**, 6671 (1992).
- [49] R. Car and M. Parrinello, *Phys. Rev. Lett.* **55**, 2471 (1985).
- [50] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos, *Rev. Mod. Phys.* **64**, 1045 (1992).
- [51] P. A. M. Dirac, *Proc. Cambridge Phil. Soc.* **26**, 376 (1930).
- [52] S. H. Vosko, L. Wilk, and M. Nusair, *Can. J. Phys.* **58**, 1200 (1980).

- [53] N. H. March, in *Theory of the Inhomogeneous Electron Gas*, S. Lundqvist and N. H. March eds., Plenum Press, New York (1983).
- [54] J. Harris and R. O. Jones, *I. Phys. F* **4**, 1170 (1974); O. Gunnarsson and B. I. Lundqvist, *Phys. Rev. B* **13**, 4274 (1976); D. C. Langreth and J. P. Perdew, *Phys. Rev. B* **15**, 2884 (1977).
- [55] O. Gunnarsson, M. Jonson, and B. I. Lundqvist, *Phys. Rev. B* **20**, 3136 (1979).
- [56] W. Pickett, *Comp. Phys. Rep.* **9**, 115 (1989).
- [57] J. Ihm, *Rep. Prog. Phys.* **51**, 105 (1988).
- [58] J. R. Chelikowsky, N. Troullier, and Y. Saad, *Phys. Rev. Lett.* **72**, 1240 (1994).
- [59] F. Gygi, *Europhys. Lett.* **19**, 617 (1992); F. Gygi and G. Galli, *Phys. Rev. B* **52**, R2229 (1995).
- [60] E. L. Briggs, D. J. Sullivan, and J. Bernholc, *Phys. Rev. B* **52**, R5471 (1995).
- [61] G. Makov and M. C. Payne, *Phys. Rev. B* **51**, 4014 (1995).
- [62] N. W. Ashcroft and N. D. Mermin, *Solid State Physics*, Holt Saunders, Philadelphia (1976).
- [63] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes*, Cambridge University Press, Cambridge (1992).
- [64] N. W. Ashcroft and D. C. Langreth, *Phys. Rev.* **155**, 682 (1967).

- [65] M. L. Cohen and T. K. Bergstresser, *Phys. Rev.* **141**, 789 (1966).
- [66] D. R. Hamann, M. Schlüter, and C. Chiang, *Phys. Rev. Lett.* **43**, 1494 (1979).
- [67] G. B. Bachelet, D. R. Hamann, and M. Schlüter, *Phys. Rev. B* **26**, 4199 (1982).
- [68] G. P. Kerker, *J. Phys. C* **13**, L189 (1980).
- [69] A. M. Rappe, K. M. Rabe, E. Kaxiras, and J. D. Joannopoulos, *Phys. Rev. B* **41**, 1227 (1990).
- [70] J. S. Lin, A. Qteish, M. C. Payne, and V. Heine, *Phys. Rev. B* **47**, 4174 (1993).
- [71] M.-H. Lee, *Advanced Pseudopotentials for Large Scale Electronic Structure Calculations*, Ph.D. Thesis, University of Cambridge (1995).
- [72] D. Vanderbilt, *Phys. Rev. B* **32**, 8412 (1985).
- [73] D. Vanderbilt, *Phys. Rev. B* **41**, 7892 (1990).
- [74] L. Kleinman and D. M. Bylander, *Phys. Rev. Lett.* **48**, 1425 (1982).
- [75] R. D. King-Smith, M. C. Payne, and J. S. Lin, *Phys. Rev. B* **44**, 13063 (1991).
- [76] A. Rahman, *Phys. Rev.* **136A**, 405 (1964).
- [77] G. Galli and A. Pasquarello, in *Computer Simulations in Chemical Physics*, M. P. Allen and D. J. Tildesley eds., Kluwer Academic Publishers, Dordrecht (1990).

- [78] D. K. Remler and P. A. Madden, *Mol. Phys.* **70**, 921 (1990).
- [79] F. Tassone, F. Mauri, and R. Car, *Phys. Rev. B* **50**, 10561 (1994).
- [80] H. C. Andersen, *J. Chem. Phys.* **72**, 2384 (1980).
- [81] M. Parrinello and A. Rahman, *Phys. Rev. Lett.* **45**, 1196 (1980).
- [82] S. Kirkpatrick, C. Gelatt, Jr. and M. Vecchi, *Science* **220**, 671 (1983).
- [83] R. Car and M. Parrinello, in *Simple Molecular Systems at Very High Density*, A. Polian, P. Loubeyre, and N. Boccaro eds., Plenum Press, New York (1988).
- [84] G. Pastore, E. Smargiassi, and F. Buda, *Phys. Rev. B* **44**, 6334 (1991).
- [85] Q. M. Zhang, G. Chiarotti, A. Selloni, R. Car, and M. Parrinello, *Phys. Rev. B* **42**, 5071 (1990).
- [86] P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization*, Academic, London (1981).
- [87] I. Stich, R. Car, M. Parrinello, and S. Baroni, *Phys. Rev. B* **39**, 4997 (1989).
- [88] M. P. Teter, M. C. Payne, and D. C. Allan, *Phys. Rev. B* **40**, 12255 (1989).
- [89] T. A. Arias, M. C. Payne, and J. D. Joannopoulos, *Phys. Rev. Lett.* **69**, 1077 (1992).
- [90] T. A. Arias, M. C. Payne, and J. D. Joannopoulos, *Phys. Rev. B* **45**, 1538 (1992).

- [91] I. Stich, M. C. Payne, R. D. King-Smith, J. S. Lin, and L. J. Clarke, *Phys. Rev. Lett.* **68**, 1351 (1992).
- [92] A. De Vita, M. J. Gillan, J. S. Lin, M. C. Payne, I. Stich, and L. J. Clarke, *Phys. Rev. Lett.* **68**, 3319 (1992).
- [93] V. Milman, M. C. Payne, V. Heine, R. J. Needs, J. S. Lin, and M. H. Lee, *Phys. Rev. Lett.* **70**, 2928 (1993).
- [94] A. De Vita, I. Stich, M. J. Gillan, M. C. Payne, and L. J. Clarke, *Phys. Rev. Lett.* **71**, 1276 (1993).
- [95] J. A. White, D. M. Bird, M. C. Payne, and I. Stich, *Phys. Rev. Lett.* **73**, 1404 (1994).
- [96] R. Perez, M. C. Payne, and A. D. Simpson, *Phys. Rev. Lett.* **75**, 4748 (1995).
- [97] C. Molteni, G. P. Francis, M. C. Payne, and V. Heine, *Phys. Rev. Lett.* **76**, 1284 (1996).
- [98] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore (1989).
- [99] A. Baldereschi, *Phys. Rev. B* **7**, 5212 (1973).
- [100] D. J. Chadi and M. L. Cohen, *Phys. Rev. B* **8**, 5747 (1973).
- [101] H. J. Monkhorst and J. D. Pack, *Phys. Rev. B* **13**, 5188 (1976).
- [102] J. Moreno and J. M. Soler, *Phys. Rev. B* **45**, 13891 (1992).
- [103] C.-L. Fu and K.-M. Ho, *Phys. Rev. B* **28**, 5480 (1983).

- [104] R. J. Needs, R. M. Martin and O. H. Nielsen, *Phys. Rev. B* **33**, 3778 (1986).
- [105] G. X. Quian, M. Weinert, G. W. Fernando, and J. W. Davenport, *Phys. Rev. Lett.* **64**, 1146 (1990).
- [106] M. P. Grumbach, D. Hohl, R. M. Martin, and R. Car, *J. Phys.: Condens. Matter* **6**, 1999 (1994).
- [107] G. Galli, R. M. Martin, R. Car, and M. Parrinello, *Phys. Rev. B* **42**, 7470 (1990).
- [108] S. Nosé, *Mol. Phys.* **52**, 255 (1984); S. Nosé, *J. Phys.: Condens. Matter* **2**, SA115 (1990).
- [109] P. E. Bloechl and M. Parrinello, *Phys. Rev. B* **45**, 9413 (1992).
- [110] E. S. Fois, J. I. Penman, and P. A. Madden, *J. Chem. Phys.* **98**, 6361 (1993).
- [111] G. W. Fernando, G.-X. Qian, M. Weinert, and J. W. Davenport, *Phys. Rev. B* **40**, 7985 (1989).
- [112] M. R. Pederson and K. A. Jackson *Phys. Rev. B* **43**, 7312 (1991).
- [113] Y. Yamamoto and T. Fujiwara, *Phys. Rev. B* **46**, 13596 (1992).
- [114] A. F. Wright and S. R. Atlas, *Phys. Rev. B* **50**, 15248 (1994).
- [115] N. Chetty, M. Weinert, T. S. Rahman, and J. W. Davenport, *Phys. Rev. B* **52**, 6313 (1995).

- [116] R. M. Wentzcovitch, J. L. Martins, and P. B. Allen, *Phys. Rev. B* **45**, 11372 (1992).
- [117] A. De Vita, *The Energetics of Defects and Impurities in Metals and Ionic Materials from First Principles*, Ph.D. Thesis, University of Keele (1992).
- [118] M. Methfessel and A. T. Paxton, *Phys. Rev. B* **40**, 3616 (1989).
- [119] S. de Gironcoli, *Phys. Rev. B* **51**, 6773 (1995).
- [120] P. E. Bloechl, O. Jepsen, and O. K. Andersen, *Phys. Rev. B* **49**, 16223 (1994).
- [121] J. A. White, D. M. Bird, and M. C. Payne, *Phys. Rev. B* **53**, 1667 (1996).
- [122] G. Kresse and J. Hafner, *Phys. Rev. B* **47**, 558 (1993).
- [123] G. Kresse and J. Hafner, *Phys. Rev. B* **49**, 14251 (1994).
- [124] J. M. Holender, M. J. Gillan, M. C. Payne, and A. D. Simpson, *Phys. Rev. B* **52**, 967 (1995).
- [125] J. M. Holender and M. J. Gillan, *Phys. Rev. B* **53**, 4399 (1996).
- [126] D. Vanderbilt and J. D. Joannopoulos, *Phys. Rev. B* **26**, 3203 (1982).
- [127] C. Molteni, private communication.
- [128] X. Gonze, *Phys. Rev. B* **52**, 1096 (1995).
- [129] C. Elsässer, M. Fähnle, C. T. Chan, and K. M. Ho, *Phys. Rev. B* **49**, 13975 (1994).

- [130] M. Born and J. R. Oppenheimer, *Ann. Physik* **84**, 457 (1927).
- [131] G. Chester, *Adv. Phys.* **10**, 357 (1961).
- [132] A. B. Migdal, *Sov. Phys. JETP* **7**, 996 (1958).
- [133] J. N. Andersen, H. B. Nielsen, L. Petersen, and D. L. Adams, *J. Phys. C* **17**, 173 (1984).
- [134] J. R. Noonan and H. L. Davis, *Phys. Rev. B* **29**, 4349 (1984).
- [135] H. Göbel and P. von Blankenhagen, *Phys. Rev. B* **47**, 2378 (1993).
- [136] M. P. Allen and D. J. Tildesley, in *Computer Simulation of Liquids*, Oxford University Press, Oxford (1987).
- [137] M. Bernasconi and E. Tosatti, *Surf. Sci. Rep.* **17**, 363 (1993).
- [138] F. Di Tolla, *Interplay of Melting, Wetting, Overheating and Faceting on Metal Surface: Theory and Simulations*, Ph.D. Thesis, Sissa/Isas (1995).
- [139] P. D. Ditlevsen and J. K. Norskov, *Surf. Sci.* **254**, 261 (1991).
- [140] C. S. Jayanthi, E. Tosatti, A. Fasolino, and L. Pietronero, *Surface Science* **152,153**, 155 (1985).
- [141] E. T. Chen, R. N. Barnett and U. Landman, *Phys. Rev. B* **41**, 439 (1990).
- [142] R. N. Barnett and U. Landman, *Phys. Rev. B* **44**, 3226 (1991).
- [143] L. Yang and T. S. Rahman, *Phys. Rev. Lett.* **67**, 2327 (1991).

- [144] P. D. Ditlevsen, P. Stoltze and J. K. Norskov, *Phys. Rev. B* **44**, 13002 (1991).
- [145] G. Mazzeo, *Interplay between surface in-plane ordering and roughening*, Ph.D. Thesis, Sissa/Isas (1992).
- [146] K. Rommelse and M. den Nijs, *Phys. Rev. Lett.* **59**, 2578 (1987).
- [147] G. Mazzeo, G. Jug, A. C. Levi, and E. Tosatti, *Phys. Rev. B* **49**, 7625 (1994).
- [148] J. W. M. Frenken and J. F. van der Veen, *Phys. Rev. Lett.* **54**, 134 (1985).
- [149] H. van Pinxteren, *Surfaces between melting and non-melting*, Ph.D. Thesis, FOM Amsterdam (1994).
- [150] A. Molenbroek, *Surface-melting instabilities at metal surfaces*, Ph.D. Thesis, FOM Amsterdam (1995).
- [151] P. Stoltze, J. K. Norskov, and U. Landman, *Phys. Rev. Lett.* **61**, 440 (1988).
- [152] A. W. Denier van der Gon, R. J. Smith, J. M. Gay, D. J. O'Connor, and J. F. van der Veen, *Surf. Sci.* **227**, 143 (1990).
- [153] A. W. Denier van der Gon, D. Frenkel, J. W. M. Frenken, R. J. Smith, and P. Stoltze, *Surf. Sci.* **256**, 385 (1991).
- [154] H. Dosch, T. Höfer, J. Peisl, and R. L. Johnson, *Europhysics Letters* **15**, 527 (1991).

- [155] A. Pavlovska, M. Tikhov, Y. Gun, and E. Bauer, *Surf. Sci.* **278**, 303 (1992).
- [156] M. Gester, D. Kleinhesselink, P. Ruggerone, and J. P. Toennies, *Phys. Rev. B* **49**, 5777 (1994).
- [157] A. M. Raphuthi, X. Q. Wang, F. Ercolessi, and J. B. Adams, *Phys. Rev. B* **52**, R5554 (1995).
- [158] R. Stumpf and M. Scheffler, *Phys. Rev. B* **53**, 4958 (1996).
- [159] E. V. Zaroquentsev, S. P. Kravchuk, and T. M. Tarusina, *Sov. Phys. Solid State* **18**, 239 (1976).
- [160] D. E. Sanders and A. E. DePristo, *Surf. Sci. Lett.* **264**, L169 (1992).