AB INITIO MODELLING TECHNIQUES APPLIED TO SILICON

P.R.BRIDDON
Department of Physics, University of Newcastle,
Newcastle upon Tyne, NE1 7RU, UK

August 13, 1999

1 Introduction

In this chapter we will consider what are termed *ab initio* modelling techniques which have been applied to systems involving silicon. There is not a precise consensus as to what constitutes an *ab initio* or *first principles* technique but the following are general characteristics.

1. The method should not contain any parameters that are taken from experiment. In principle all that should be required to perform a calculation would be the names of the chemical species present and possibly a starting structure.
2. The method should be capable of providing a number of different properties from the same theory (for example equilibrium structures, migration barriers, vibrational frequencies, optical excitation energies and so forth).
3. The method should be *transferable* among a variety of systems. This should mean that, in increasing order of hopefulness, a method that works for silicon should also work for other materials too such as other semiconductors, metals, ionic solids, highly correlated systems and so on.

In practice, this is not quite achieved by any single theory, and the techniques described are all approximate. In this review, the Born–Oppenheimer approximation is first discussed, as this is assumed by most techniques described here. Following this two of the common approaches are described, Hartree Fock theory and the density functional pseudopotential approach. Steps being taken towards the goal of *linear scaling* are described and finally *quantum Monte Carlo* simulations, a possible method of the future are described.

2 The Born–Oppenheimer Approximation

The solution to the full non–relativistic Schrödinger equation describing a system of atoms would be a function of the form $\Psi_T(r,R)$ where we use the symbol r to collectively label the co–ordinates of all the electrons $\{\mathbf{r}_1,\mathbf{r}_2,\ldots\}$ and R to label the co–ordinates of the nuclei $\{\mathbf{R}_1,\mathbf{R}_2,\ldots\}$. The *Born–Oppenheimer* approximation consists of looking for a separated variable type solution:

$$\Psi_T(r,R)=\Psi_R(r)\chi(R). \tag{1}$$

In this approximation a Schrödinger–type equation can be solved for the electronic degrees of freedom, Ψ_R, while the nuclei (of charge Z_a) are assumed to be stationary. The equation to be solved is

$$\hat{H}\Psi_R(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N)=E(R)\Psi_R(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N) \tag{2}$$

where

$$\hat{H}=-\frac{1}{2}\sum_i\nabla^2_{\mathbf{r}_i}+\sum_{ia}\frac{Z_a}{|\mathbf{r}_i-\mathbf{R}_a|}+\frac{1}{2}\sum_{i\neq j}\frac{1}{|\mathbf{r}_i-\mathbf{r}_j|}+\frac{1}{2}\sum_{a\neq b}\frac{Z_aZ_b}{|\mathbf{R}_a-\mathbf{R}_b|} \tag{3}$$

Note that we will be using *atomic units* unless otherwise indicated. In these units, $\hbar=m_e=e=4\pi\epsilon_0=1$. The unit of length is the Bohr radius, $a_0\approx 0.529$Å, and the unit of energy the Hartree or 27.2116 eV. Note that we have labelled the energy as $E(R)$ and will regard it as the "energy with the nuclei frozen at positions R". This function defines a *potential energy* surface which can be used for a number of purposes. For, example the minimum of this gives the equilibrium structure. Barriers to various complex processes (e.g. dissociation, diffusion) can also be found from saddle points of this surface.

The equation for $\chi(R)$ includes terms involving $\Psi_R(r)$. In the Born–Oppenheimer approximation, these are neglected, and the total energy is then just given by $E(R)$ plus the zero point energy of vibration of the nuclei which can be estimated in the harmonic approximation. The coupling terms can be important under certain circumstances (in degenerate ground states they give rise to *Jahn–Teller distortions* [1, 2]). In most other cases, this approximate separation is acceptable because of the mass difference between electrons and nucleons. One situation for which this is not adequate is that of muon–spin resonance experiments in which a *muon* binds an electron to form muonium — an atom analogous to hydrogen but with only 1/9 the mass. The zero point energy of this is several hundred meV — an energy comparable with other important parameters (e.g. diffusion barriers estimated from $E(R)$) in the system.

In truth however, equation (2) cannot readily be solved directly for systems larger than single atoms. In practice some further approximations have to be made. In the above *ab initio* spirit the approximation should be (a) non-empirical and (b) the same for all systems considered. We now move on to consider these.

3 Hartree-Fock Theory

Hartree–Fock theory was one of the first *ab initio* schemes to be put into practice. It has been widely used for atoms and small molecules, although less so for materials such as silicon, and only a brief discussion will be given here. The method is basically a variational calculation, in which the expectation value of the Hamiltonian (3) is minimised in an approximate wavefunction. The wavefunction is chosen to be a single determinant of one–electron spin–orbitals $\psi_\lambda(\mathbf{x})$ where $\mathbf{x}$ labels the spatial $\mathbf{r}$ and spin co–ordinates :

$$\Psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) = \frac{1}{\sqrt{N!}} \det|\psi_\lambda(\mathbf{x})| \tag{4}$$

Minimisation of the energy in this basis gives the famous Hartree–Fock equations [3] which are now readily solved for systems containing of order 100 atoms.

This theory has proved successful, particularly for atoms where it gives a lower energy than density functional theory to be described in the next section. It has not been widely used for work in silicon, and has one major failing : the trial wavefunction entirely ignores correlation. This is most easily seen by looking at the wavefunction for a Helium atom in its ground state which can be considered to be made up from the one-electron configuration $1s^2$:

$$\Psi_{HF}(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{\sqrt{2}} \phi_{1s}(\mathbf{r}_1)\phi_{1s}(\mathbf{r}_2)(\uparrow\downarrow - \downarrow\uparrow) \sim \exp[-\alpha(r_1 + r_2)](\uparrow\downarrow - \downarrow\uparrow) \tag{5}$$

It is clear that this wavefunction has no dependence on $|\mathbf{r}_1 - \mathbf{r}_2|$, which would be the signature of correlation between the two electrons. This is a serious omission and has disastrous consequences in the modelling of a number of different systems. Some simple molecules, for example F_2, do not have a bound state in Hartree–Fock theory, and when applied to the uniform electron gas Hartree–Fock theory gives a zero density of states at the Fermi level. This suggests that systems such as simple metals would not be well described by this approximation.

Numerous schemes have been attempted to add this correlation. Some of these apply Moeller–Plesset perturbation theory which is performed "on top" of a self–consistent Hartree Fock theory. These calculations are extremely time consuming and haven't been used extensively in the modelling of semiconductors.

In principle, the exact ground state (or even an excited state) energy can be obtained by doing a *configuration interaction* calculation in which the wavefunction is written as a linear combination of determinants. This is an even-more demanding calculation, as an enormous number (millions!) of determinants must be included, and the number required increases rapidly with system size. In practice, this that cannot really be converged for a system large enough to model a bulk solid, and as a result, this method has not been applied to model properties of silicon. We will now pass onto methods that are more routinely used to study processes in semiconductors.

4 Density Functional Theory

The principal feature of *density functional* methods is that the many problem is solved directly for the charge density, $n(\mathbf{r})$ rather than for the many-electron wavefunction Ψ. This is a massive simplification, as we only need consider a function of three variables x, y and z, rather than the $3N$ variable problem above.

4.1 The Hohenberg–Kohn Theorem

Prior to the work of Hohenberg and Kohn in 1964 the use of the electron density as a fundamental variable was thought of as a somewhat *ad hoc* approach. However, Hohenberg and Kohn [4] showed that this method may in principle enable the calculation of the *exact* ground-state energy. The Hohenberg–Kohn theorem states that *the external potential* $V^{ext}(\mathbf{r})$ *is determined to within an additive constant, by the electron density* $n(\mathbf{r})$. This is a quite remarkable assertion, and the proof presented was astonishingly simple [4].

The implication of this is that, because the usual machinery of quantum mechanics (i.e. solving the Schrödinger equation) enables the calculation of the ground state energy from the external potential, if this potential is uniquely determined by the charge density, it must follow that it is possible to write down an expression for the total energy directly in terms of the electron density. Hohenberg and Kohn wrote

$$E_{TOT} = E_{HK}[n] = \int n(\mathbf{r}) V^{ext}(\mathbf{r}) d\mathbf{r} + F[n]$$

where the functional $F[n]$ is *universal* in the sense that it only depends on the electron density and and not on the background potential.

The second theorem included in the work of Hohenberg and Kohn stated that for a trial density $\tilde{n}(\mathbf{r})$, correctly normalised and satisfying $\tilde{n}(\mathbf{r}) > 0$,

$$E_{HK}[\tilde{n}] \geq E_{GS}$$

so that it may be possible to find the ground state charge density using a minimisation technique.

A number of elaborations have been proposed since this pioneering work. An important generalisation has been to the case where the external field includes a spin–dependent potential (as is the case with a magnetic field, where a term proportional to $\sum_i \mathbf{B}(\mathbf{r}_i) \cdot \mathbf{s}_i$ must be added to the Hamiltonian (3). This time we need two fundamental variables $n_\uparrow$ and $n_\downarrow$ the spin–up and spin–down densities which together fix the external potential operating [5, 6]. This is sometimes referred to as local spin–density functional theory. The main advantage of this formalism has turned out not to be that magnetic fields can be described, but rather that when coupled with a local density approximation (see section 4.3) the description of open–shell atoms, molecules and more complex systems such as paramagnetic defects in semiconductors is vastly improved. Spin–orbit coupling terms can also be included within this framework [7].

Another important generalisation was to finite temperature systems [8] in which all properties of a system in equilibrium with fixed temperature T and chemical potential μ are determined by the charge density. This has also proved to be important when attempting to model metals.

One initial objection to a practical use of the second theorems was that not all functions are *v–representable*. In other words not all functions can necessarily be associated with an anti–symmetric ground–state wavefunction of a Hamiltonian. It is not known what conditions are necessary or sufficient for this. Fortunately, in developing the *constrained search* derivation of the Hohenberg–Kohn theorem, Levy showed that (i) the first theorem also applies to degenerate ground states, a loophole in the original Hohenberg–Kohn derivation and (ii) it suffices for the variational principle that the trial densities be *N–representable*, that is that they be obtainable from an antisymmetric wavefunction. This is a much weaker condition and is satisfied by all continuous, non–negative and correctly normalised functions [9, 10].

4.2 The Kohn–Sham equations

The above theorems paved the way for a new approach to structural calculations. However, it has not proved possible to write down an explicit form for $F[n]$ which is accurate enough for materials such as silicon, although some progress has been achieved in systems in which the density does not vary too rapidly [11]. Kohn and Sham (KS) [12] addressed this problem in 1966, by introducing an auxilliary system containing N non–interacting electrons (in states ψ_λ) in a background potential $v_s(\mathbf{r})$, chosen such that the charge density in this auxilliary system is exactly the same as that in the full interacting system :

$$\sum_\lambda |\psi_\lambda(\mathbf{r})|^2 = n(\mathbf{r}).$$

It is simple to calculate the kinetic energy of the *non-interacting* system of electrons where the electrons are in these states using

$$T_s[n] = \sum_{\lambda=1}^{N} \langle \psi_\lambda | -\frac{1}{2}\nabla^2 | \psi_\lambda \rangle \tag{6}$$

The true kinetic energy of the interacting system will of course differ from this, and the difference between this term (the largest contribution) and the exact result is treated separately. The total energy is then written as

$$E_{TOT} = T_s[n] + \int V^{ext}(\mathbf{r}) n(\mathbf{r}) d\mathbf{r} + \frac{1}{2} \int \frac{n(\mathbf{r}) n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' + E_{xc}[n]$$

where a suitable approximation is to be sought for E_{xc}. This term contains the exchange and correlation energies and the correction to the kinetic energy. We will consider this term in the following section.

Kohn and Sham showed that the ψ_λ functions can be obtained from a self–consistent solution of the equations :

$$\begin{aligned} -\frac{1}{2}\nabla\psi_\lambda + v_s(\mathbf{r})\psi_\lambda(\mathbf{r}) &= \epsilon_\lambda \psi_\lambda(\mathbf{r}) \\ v_s(\mathbf{r}) &= V^{ext}(\mathbf{r}) + \int \frac{n(\mathbf{r}')\, d\mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|} + \frac{\delta E_{xc}}{\delta n(\mathbf{r})} \\ n(\mathbf{r}) &= \sum_\lambda |\psi_\lambda(\mathbf{r})|^2 \end{aligned} \tag{7}$$

A more detailed discussion of this (and the various extensions to it, particularly the extensions to spin polarisation, finite temperatures and variable occupancies) is contained in, for example, [13] or [14]. These equations are a tremendous simplification to the original many body problem. Before we turn to the details of how this is done in practice, we will first consider the common approximations for E_{xc} and then give a brief discussion of the reliability of this formalism.

4.3 Approximations for E_{xc}

Up to this point, no approximation has yet been made — if the functional $E_{xc}[n]$ were known, an application of the above procedure would produce the correct ground state energy and charge density. We will now review some of the common approximations made for E_{xc}.

The most common approach is the local density approximation (LDA). This treats the inhomogeneous electron case as uniform locally :

$$E_{xc} = \int n(\mathbf{r}) \epsilon_{xc}\left(n(\mathbf{r})\right) d\mathbf{r}$$

where $\epsilon_{xc}(n)$ is the exchange–correlation energy per electron for a uniform electron gas of density n. Clearly this is very much in the spirit of the Thomas–Fermi treatment of kinetic energy, but this time the approximation is much more accurate as the exchange–correlation energy is smaller and more slowly varying.

An improved description is that of the local spin–density approximation (LSDA) [5]. In this one has two quantities, the spin-up and spin down charge densities, $n_\uparrow$ and $n_\downarrow$ and this time we use the expression $\epsilon_{xc}(n_\uparrow, n_\downarrow)$ for the uniform electron gas.

We now discuss the determination of ϵ_{xc} for the homogeneous electron gas. Typically we split this term into exchange and correlation effects $E_{xc}[n_\uparrow, n_\downarrow] = E_x[n_\uparrow, n_\downarrow] + E_c[n_\uparrow, n_\downarrow]$. The exchange part is easily evaluated via an analytic Hartree–Fock treatment of the uniform electron gas :

$$E_x[n_\uparrow, n_\downarrow] = -\frac{3}{2}\left(\frac{3}{4\pi}\right)^{1/3} [n_\uparrow^{4/3} + n_\uparrow^{4/3}]. \tag{8}$$

The correlation part is more complex, and is evaluated in the high density limit using many body perturbation theory [15] and in the low density limit by Green function quantum Monte Carlo calculations [16]. Typically these numerical results are fitted to a simple parametrised form. One of the most common parametrisations is that of Perdew and Zunger (PZ) [17] who first fitted two limiting the non–polarised result $\epsilon_c(n/2, n/2)$ and the completely spin–polarised result $\epsilon_c(n, 0)$. If we introduce the Wigner-Seitz radius $r_s = (4\pi n/3)^{-1/3}$ then ϵ_c for the non-polarised and fully polarised electron gases are given by :

$$\epsilon_c = \begin{cases} \gamma\{1 + \beta_1\sqrt{r_s} + \beta_2 r_s\}^{-1}, & \text{for } r_s \geq 1 \\ B + (A + Cr_s)\ln(r_s) + Dr_s, & \text{for } r_s < 1 \end{cases}$$

The values of the coefficients are given for both cases in TABLE 1.

A partially polarised gas can be characterised by the polarisation factor $\xi = [n_\uparrow - n_\downarrow]/n$ where $0 < \xi < 1$. In this case the correlation energy is averaged over the polarised and non-polarised cases using the procedure due to von Barth and Hedin [5] :

$$\epsilon_c(n_\uparrow, n_\downarrow) = \epsilon_c(n/2, n/2) + f(\xi)\big[\epsilon_c(n, 0) - \epsilon_c(n/2, n/2)\big] \tag{9}$$

Table 1: Perdew–Zunger parametrisation of exchange-correlation energy

	γ	β_1	β_2	A	B	C	D
Non-polarised	-.1423	1.0529	0.3334	0.0311	-0.0480	0.0020	-0.0116
Polarised	-.0843	1.3981	0.2611	0.0155	-0.0269	0.0007	-0.0048

where the interpolation function is given by

$$f(\xi) = \frac{(1+\xi)^{4/3} + (1-\xi)^{4/3} - 2}{2^{4/3} - 2}. \tag{10}$$

An alternative parametrisation has been given by Vosko–Wilk–Nusair (VWN) [18]. More recently, a new parametrisation has been suggested by Perdew and Wang [19] with a number of improvements over the previous work. This may be the most accurate representation available at present, but for most computational purposes, PZ and VWN parametrisations give very similar results.

These formulae have been used widely over two decades to model an enormous variety of systems ranging over atoms, molecules, clusters and solids and have made a significant contribution to a number of areas of physics, chemistry and materials science. The accuracy has been remarkable, and more than expected from such a simple approach.

Early attempts to improve on the LDA were not successful. The most obvious step on from LDA is to develop a function $\epsilon_{xc}(n, |\nabla n|)$ which would take into account the inhomogeneity of the gas. The first attempt at this was the gradient expansion approximation (GEA). On dimensional grounds, the this is seen to be of the form

$$E_{xc} = \int n(\mathbf{r})\epsilon_{xc}(n(\mathbf{r}))d\mathbf{r} + \int C(n(\mathbf{r}))\frac{|\nabla n|^2}{n^{4/3}}d\mathbf{r}.$$

However, when applied to real systems such as atoms, the charge density varies far too quickly for perturbation theory to be valid, and as a result results were made significantly worse by this supposed improvement. For example, correlation energies of atoms were predicted to be positive.

From this initial work, the incorporation of gradient corrections has gradually developed. It has been realised for some time [7] that the exchange–correlation hole around each electron satisfies certain sum rules. In particular, the exchange hole integrates to one electron, is strictly negative and the correlation hole integrates to zero. The LDA corresponds to a real physical system and therefore satisfies these criteria automatically whereas the GEA being a leading term in a series expansion does not. Gradually gradient corrections were incorporated in ways that enforced these sum rules [20, 21]. In additional various co-ordinate scaling laws [22] and limits [23] that should be obeyed by functionals were developed. The resulting approximations, known as generalised gradient approximations (GGAs) are now coming into widespread use.

An early functional referred to as Becke–Perdew (BP) consisted of the Perdew formula for correlation [21] combined with an extremely accurate but empirical formula for exchange developed by Becke [24, 25]. A more widely used functional was developed by Perdew and Wang in 1991 [26] and is referred to as PW91 in the literature. Applications of this show improvements over the LDA, particularly for atoms and molecules and alkali metals [27]. However, when applied to semiconductors such as silicon, mixed results were obtained. In these materials, the density gradients are never too large and the LDA does exceptionally well. The GGA tends to increase bond lengths and as a consequence reduce bulk moduli giving somewhat worse values than the LDA. However, significant problems were encountered with the construction of pseudopotentials within the GGA [28] which may have been a factor behind the variable results obtained. Most recently, a new functional referred to as PBE96 has been developed [29, 30] and this is both a simplification and an improvement on the PW91 functional, removing a number of undesirable features, although numerically both PW91 and PBE96 give very similar results.

4.4 Reliability

Calculations performed in the local density approximation have been shown to be remarkably accurate for ground state properties. For example the structure of small molecules are given to within 1-2% and vibrational frequencies to within 5-10%. The one disappointment has been the binding energies of molecules which are typically overestimated by 10% or so. The GGA improves this as shown in table 2, taken from [29].

Early calculations focussed on properties of bulk semiconductors. For example, it was soon found that lattice constants and bulk moduli were given accurately. Zero temperature phase stability diagrams were obtained [31, 32]. Calculation of phonon spectra followed — a review of early work is given by Kunc [33] and more recent results [34] are given in

Table 2: Atomisation energies of some small molecules in kcal/mol [29]

Molecule	HF	LSDA	PW91	PBE96	Expt.
H_2	84	113	105	105	109
CH_4	328	462	421	420	419
H_2O	155	267	235	234	232
O_2	33	175	143	144	121
F_2	-37	78	54	53	39

Table 3: Calculated (c) and experimental (e) lattice constant and phonon frequencies [34].

	Si(c)	Si(e)	Ge(c)	Ge(e)	GaAs(c)	GaAs(e)
a_0 (Å)	10.20	10.26	10.60	10.68	10.61	10.68
$\Gamma_{TO}(cm^{-1})$	517	517	306	304	271	271
$\Gamma_{LO}(cm^{-1})$	517	517	306	304	291	293
$X_{TA}(cm^{-1})$	146	150	80	80	82	82
$X_{LA}(cm^{-1})$	414	410	243	241	223	225
$X_{TO}(cm^{-1})$	466	463	275	276	254	257
$X_{LO}(cm^{-1})$	414	410	243	241	240	240
$L_{TA}(cm^{-1})$	111	114	62	63	63	63
$L_{LA}(cm^{-1})$	378	378	224	222	210	207
$L_{TO}(cm^{-1})$	494	487	291	290	263	264
$L_{LO}(cm^{-1})$	419	417	245	245	238	242

TABLE 3. It is seen that these frequencies are extremely reliable. Localised vibrational modes of impurities have also been obtained [35] and compare well with experiment. Surfaces have also been extensively modelled — for example, the 7x7 Takayanagi reconstruction of the Si [111] surface has been modelled [36]. Point defects have been extensively treated [37] and even hyperfine couplings have been extracted [38]. The sheer volume and variety of applications illustrate the applicability the robustness of this method.

5 Pseudopotentials

The use of pseudopotentials has proved an extremely important step in using *ab initio* methods to model large systems. It has long been realised that, although a silicon atom has 14 electrons, only 4 of them play a significant role in chemical bonding. The core electrons remain localised close to the atom and the shapes of the $1s$, $2s$ and $2p$ orbitals are largely unaffected by the chemical environment of the silicon atom. The basic idea is that instead of using the full Coulomb potential, $-Z/r$, to describe the nuclear interaction, we use a *pseudopotential*, $V^{ps}(r)$, which effectively eliminates the core states from the calculation. An excellent review of this technique has been compiled by Pickett [39].

Using the full Coulomb potential can cause considerable problems. The total energy then becomes extremely large and since one is interested in relatively small differences in energies this places great demands on the precision of calculation, especially as the energy is overwhelmingly dominated by contributions from electrons localised near the nucleus, which are the least important as far as physics, chemistry or materials science is concerned. Secondly, the fitting of core wavefunctions with either plane waves or Gaussian orbitals is extremely difficult and small errors can make large differences in the core eigenvalues. FIGURE 1 shows the $4s$ wavefunction and pseudo-wavefunctions in Ni. It is clear that the pseudo-wavefunction is a much simpler and smoother function to approximate than the all-electron wavefunction. Thirdly, for the heavier atoms, relativistic effects are important and the Dirac equation is required. However, the valence electrons can continue to be treated non-relativistically and a spin-orbit potential can be introduced which describes polarised valence electrons.

One important feature of pseudopotentials is that there is no–unique method of calculating them — at least to a casual glance, different pseudopotentials for the same element can appear to differ more than pseudopotentials for different elements! This is illustrated in FIGURE 2 in which three prescriptions for the $l = 1$ component of the carbon pseudopotential are shown. All give excellent results for the properties of carbon!

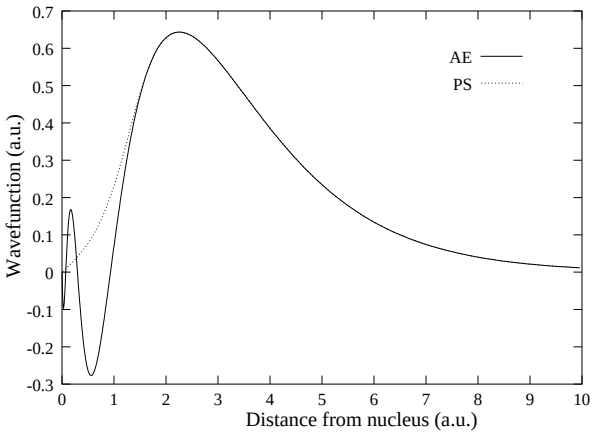

Figure 1: The 4s all–electron (AE) and pseudo (PS) radial wavefunctions for the Ni atom.

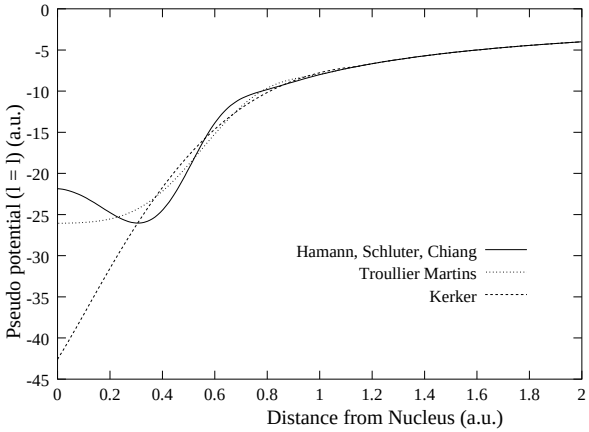

Figure 2: Three different constructions of the $l = 1$ pseudopotential for carbon

5.1 Ionic Pseudopotentials

Empirical pseudopotentials were first developed to produce a good description of some physical property. For example, a functional form for a pseudopotential $V^{ps}(r)$ was specified. This was varied until the solutions E_λ of

$$(\hat{T} + V^{ps})\psi_\lambda = E_\lambda \psi_\lambda \tag{11}$$

agreed with properties of some known system (e.g. band gaps, deformation potentials of bulk silicon). The potential was then used to model a related system. The problem with such pseudopotentials is that they are not very *transferable* — they only work well in systems which are very similar to those in which they have been fitted.

One obvious reason for this is that these pseudopotentials include the potential from the valence states which will certainly vary with chemical environment. A transferable pseudopotential must be de-screened, or after determining a potential V^{ps} which when used in equation 11 gives acceptable solutions, an *ionic* potential must be constructed by subtracting off the potential arising from the valence states.

$$V^{ps}_{ion}(\mathbf{r}) = V^{ps}(\mathbf{r}) - \int \frac{n^{ps}(\mathbf{r}')\, d\mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|} - V_{xc}(n^{ps}(\mathbf{r})) \tag{12}$$

where

$$n^{ps}(\mathbf{r}) = \sum_\lambda |\psi^{ps}_\lambda(\mathbf{r})|^2 \tag{13}$$

and the sum is over the occupied valence pseudo eigenstates. Carrying out this procedure opens up the possibility of generating pseudopotentials from a first–principles calculation on the atom and then using it in a solid state environment, thereby regaining the *ab initio* spirit of the procedure.

5.2 Approximations

A number of approximations are implicitly used in the construction of pseudopotentials. First a *one–electron* picture is being used when splitting electrons into core and valence sets. This however fits naturally with the Kohn–Sham scheme.

A second approximation is the *frozen core approximation* which is fundamental to the concept of a pseudopotential — that certain of the one–electron states do not change significantly, but remain frozen when transferred from one chemical environment to another. It is not always obvious which these are. For example, in the case of the II-VI compound semiconductor ZnSe, it may be hoped that the Zn 3d electrons can be treated as core states, as they are not expected to participate significantly in chemical bonding. However, this is not a successful approximation as the resulting predicted lattice constant is far too short at 5.19Åcompared to the experimental 5.67Å. This error is removed if the 3d electrons are treated as valence states. In fact the shape of these states does change and the charge density becomes noticeably non–spherical [40].

A third approximation is made in the de-screening step described above — the *small core approximation*. This assumes that the core and valence states do not overlap significantly. The problem is that the exchange–correlation potential is non–linear and that it is certainly not true that

$$V_{xc}(n_c + n_v) = V_{xc}(n_c) + V_{xc}(n_v)$$

This is approximately true when the core and valence density are localised in different regions of space, but is certainly violated by the d states of group III elements such as Ga and In which, although they do not participate in bonding, do overlap appreciably with the s and p valence states. One result of this is that standard "d in the core" pseudopotential calculations of the lattice constant of GaAs give a result 2.5% shorter than experiment, and agreement is worse in InAs. This is not the case in an all–electron calculation. This can be improved by the use of *non–linear core corrections* [41]. The idea here is that the de-screening step is carried out using the exchange–correlation potential of the valence charge density *plus* a component of the core density. This component of core density needs to be carried along in the pseudopotential calculation and added onto the valence density before evaluating the exchange–correlation potential and energy.

5.3 Transferability and Norm conservation

A major step forward in the attempts to produce transferable first principles pseudopotentials came in the early 1980s with the advent of norm conserving pseudopotentials. This term is used as outside a certain *cutoff radius*, these potentials give rise to pseudo-wavefunctions which are *identical* to the true all–electron wavefunction, not just proportional to it as was the case previously. The cutoff radius is not an adjustable parameter, but rather a *quality* parameter – the smaller this radius, the closer the pseudo-wavefunction is to the true wavefunction and therefore the more transferable the pseudopotential. Typically the cutoff radius is chosen to be between the outermost node and outermost radial maximum of the true orbital.

Hamann, Schlüter and Chiang [42] showed that norm–conservation also guaranteed that the energy derivative of the logarithmic derivative of the wavefunction, ψ_l'/ψ_l, is the same as that for the true wavefunction. This means that the scattering properties should match over a large energy range, giving these potentials their transferability. More recently it has been shown that higher energy derivatives can also be matched [43]. Indeed, a comparison between the energy dependence of the logarithmic derivative of the all–electron and pseudopotential wavefunctions is one common test of the quality of a pseudopotential. A second test is a comparison between energies of different electronic configurations. For example, the energy difference between C in the s^2p^2 and sp^3 configurations is 8.23 eV when the all-electron theory is used and 8.25 eV using the BHS pseudopotential described below. This gives us confidence that the energetics of different bonding patterns may be well represented in the solid state. A more recent study has proposed a systematic method of measuring transferability [44].

Norm–conserving potentials are *non–local* and operate differently on s, p and d type wavefunctions. They usually have the form $\sum_l V_l(r)\hat{P}_l$ where $\hat{P}_l$ is an angular momentum projection operator. A number of different methods for constructing such potentials have been proposed and we will now review these. An early scheme was proposed by Kerker [45], although these potentials have not been widely used. We start with the work of Hamann, Schlüter and Chiang [42].

5.4 The pseudopotentials of Bachelet, Hamann, Schlüter and Chiang

Hamann, Schlüter and Chiang (HSC) [42] proposed a very *systematic* method for construction, and this was exploited by Bachelet, Hamann and Schlüter (BHS) [46] who applied this methodology to all atoms from hydrogen to plutonium. They fitted the pseudopotentials to a simple analytic form and published the resulting parameters for all these elements. The following stages are involved:

1. The Dirac equation (in the LSDA) is solved for a specimen atomic configuration. The determination of the all–electron Kohn–Sham eigenvalues ϵ_λ and eigenfunctions u_λ. The following steps are then carried out for each angular momentum, $l = 0, 1, \ldots$.

2. A "first–step" pseudopotential is constructed by smoothly cutting off the $r = 0$ singularity in the screened all–electron potential just found. This is carried out, whilst at the same time ensuring that the eigenvalue is not changed.

3. The pseudo–orbital corresponding to this new potential is modified to impose norm–conservation.

4. The "stage–two" pseudopotential that gives rise to this new orbital is determined by inversion of the radial Schrödinger equation.

5. This pseudopotential is descreened to produce an ionic pseudopotential.

6. This numerical representation is fitted to a simple analytical form thereby making it simple for other workers to use.

The BHS pseudopotentials have proved reliable in a large number of calculations. However, they are quite *hard*. In practical terms, this means that a very large number of plane waves is required to represent them in Fourier space. This has been addressed by Vanderbilt [47] who suggested modifications to certain steps of the BHS procedure. This led to pseudopotentials of comparable accuracy which have more rapidly convergent Fourier transforms.

5.5 The Troullier–Martins construction

Another form of pseudopotential which has come into common usage is that of Troullier and Martins [48]. These pseudopotentials were designed to be smoother and more amenable to expansion in plane waves. The work followed on from that of Kerker [45], but additional conditions were imposed. In essence,

1. the wavefunction and its first four derivatives were forced to be continuous at the cut-off radius. This is a distinct improvement on both Kerker's work and that HSC pseudopotentials in which only two derivatives were continuous.
2. The odd derivatives of the pseudopotentials at $r = 0$ were forced to be zero
3. The screened pseudopotential was forced to have zero curvature at $r = 0$.

Troullier and Martins argued that to produce a pseudopotential of equal quality to the Hamann–Schlüter–Chiang method, these additional constraints allowed a much larger cutoff radius to be chosen, giving a significant increase in convergence rate with basis set size.

5.6 The ultrasoft pseudopotentials of Vanderbilt

A much more radial scheme for the construction of soft–pseudopotentials has been suggested by Vanderbilt and co–workers [49, 50]. Vandebilt noted that elements such as oxygen require large numbers of plane waves to describe valence states tightly bound to the nucleus and this cannot be improved in conventional norm–conserving schemes as a significant weight of these states lines *inside* the cut–off radius. Norm–conservation was relaxed, allowing charge to flow out of the core region, giving a much more slowly varying charge density that can be treated with fewer plane waves and the charge balance between core and valence regions later restored by a charge augmentation step. This enables oxygen to be teated with a cut–off as low as 25 Ry. These potentials are now coming into more widespread usage.

5.7 The Kleinman–Bylander form

In general, the action of a non–local potential $\hat{V}$ on a function $\phi(\mathbf{r})$ may be written as

$$\langle \mathbf{r}|\hat{V}|\phi\rangle = \hat{V}\phi(\mathbf{r}) = \int V(\mathbf{r},\mathbf{r}')\phi(\mathbf{r}')d\mathbf{r}'. \tag{14}$$

where the kernel $V(\mathbf{r},\mathbf{r}')$ may be written as

$$V(\mathbf{r},\mathbf{r}') = \sum_{lm} V_l(r,r')Y^*_{lm}(\theta,\phi)Y_{lm}(\theta',\phi') \tag{15}$$

The non–local pseudopotentials discussed above are however non–local only in the sense that they are angular–momentum dependent and may be written as

$$\hat{V}^{ps} = \sum_{lm} |Y_{lm}\rangle V_l(\mathbf{r})\langle Y_{lm}| \tag{16}$$

so that in this case the function $V_l(r,r')$ of equation 15 may be written as

$$V(r,r') = V_l(r)\delta(r-r') \tag{17}$$

This form is known as *semi–local* as it is non–local only in the angular co–ordinates. In working out the matrix elements of this in N plane waves, $N(N+1)/2$ integrals need to be done, a very heavy workload. Kleinman and Bylander [51, 52] suggested that by working with a *completely non–local* potential (i.e. non–local in the radial as well as angular co–ordinates) the workload can be reduced. In this, the quantity $V_l(r,r')$ in equation (17) is written in separable form as

$$V_l(r,r') = F^*_l(r)F_l(r')$$

so that when evaluating of matrix elements $V_{ij} = \langle i|\hat{V}|j\rangle$ in a set of N basis functions $|i\rangle$, the $\mathbf{r}$ integral (which involves only the basis function i), and the $\mathbf{r}'$ integral involving only function j can be done separately, reducing the number of integrals to N, an enormous computational advantage.

The precise prescription put forward by Kleinman and Bylander was

$$\hat{V}^{ps} = V_{loc} + \sum_{lm} \frac{|\delta V_l \psi^0_{lm}\rangle\langle\psi^0_{lm}\delta V_l|}{\langle\psi^0_{lm}|\delta V_l|\psi^0_{lm}\rangle} \tag{18}$$

where an arbitrary local potential V_{loc} is separated out the quantity $\delta V_l = V_l(r) - V_{loc}(r)$ is introduced. The functions ψ^0_{lm} are the pseudo-eigenfunctions from which the original pseudopotential was created.

The action of this KB form of the pseudopotential on an atomic pseudo–orbital is just the same as the original potential, but in general the two forms will not produce identical results in a molecular or solid state problem. Indeed, care has to be taken when deciding on the choice of V_{loc} or what are termed *ghost states* may appear. These are unphysical states having very low energies. The origin of these states has been studied by Gonze *et al.,*[53, 54] and criteria developed by which the appearance of these states can be avoided. In brief, the δV_l should be as small and short–ranged as possible, and large positive values should be avoided. Generalisation of the KB form have been made by Vanderbilt [55] and Bloechl [56].

6 Solution of Kohn–Sham equations

Here we turn to some of the practical details of the solution of the Kohn–Sham equations. Usually the equations have been solved by expanding the solution in a basis. This transforms the partial differential equation to a discrete matrix problem that can then be solved. This is in itself a huge area of work, with many different techniques being employed. Here only a few general points will be made.

6.1 Choice of Boundary Condition

Bulk silicon is of course a periodic system, and as such is best modelled using a unit cell applying periodic boundary conditions, that the charge density is periodic, and that the Kohn–Sham orbitals satisfy Bloch's theorem. When modelling a defect in a solid, there is no longer periodicity and so a unit cell is no longer ideal. Two approaches are commonly adopted using supercells or clusters.

In the supercell approach, the defect is placed in a large unit cell, typically containing 50-100 atoms. It is hoped that this is sufficiently large that the interaction between defects in adjacent supercells is small. This is difficult to achieve, and even with 50 atoms, defect related states which are quite localised are found to have up to 0.5 eV of spurious dispersion across the zone. When combined with the typical LDA underestimate of the band gap, this can result in states that should be strongly localised mixing strongly with band states. This dispersion width is often comparable with other energies of interest in the problem. When a defect gives rise to a dipole moment, the interaction is greater, and methods for reducing such interactions have been proposed [57]. Interaction effects can be more serious if a defect such as a dislocation is modelled. This time, to restore periodicity, two dislocations must be placed in the unit cell and the interaction between them is of course very large.

In contrast, the cluster approach does model only one copy of the defect, but this time a new penalty is paid — the interaction between the defect and the surface of the cluster. This should not be too significant if the cluster is large, but this is not always easy to demonstrate. One advantage is that defects that are intrinsically non–periodic such as stacking faults or dislocations can be modelled easier. The surface of the cluster is generally passivated with hydrogen so that no dangling bonds are left. This is important as otherwise the electronic states associated with the unpaired electrons would fall into the band gap and interfere with the states associated with the defect being studied. It is best to relax these hydrogen atoms to the equilibrium bond length as this reduces the strain on the cluster of silicon atoms. Calculations using clusters with full self–consistent density function theory routinely treat 2-300 atoms [58].

A third approach is the Green function approach — in principle this combines the best features of both clusters and supercells, that only one defect is being studied, but that defect is embedded into an infinite solid. This approach was used twenty years ago to study point defects [59, 60] but no lattice relaxation was incorporated at that time, and since then this technique has not been used frequently. A slightly different Green–function method, the "quasi–band method" has been used by Lindefelt and Zunger [61] to study $3d$ transition metal impurities in silicon.

The choice of approach is related to the choice of basis and possibly to the individual worker's background (it is not unusual to see periodic solids modelled with clusters and molecules modelled in unit cells!).

6.2 Choice of Basis set

Two common choices of basis set are plane waves and Gaussian type orbitals.

The usual choice has been plane waves as these fit naturally with periodic boundary conditions, which are generally used by condensed matter theorists. In this case, the Kohn–Sham orbitals for some point $\mathbf{k}$ in the Brillouin zone are expanded in terms of plane waves:

$$\psi_{n\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} c_{n\mathbf{k}}(\mathbf{G}) \exp[\mathbf{i}(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}] \tag{19}$$

This expansion has the advantage that it is very transparent (there are no additional parameters that may vary from one worker to another), it is very stable (plane waves with different $\mathbf{G}$ are of course orthogonal, so no instability can creep into the calculation through near–linear dependencies in the basis set), it is clear how to improve the expansion (usually, all plane waves with $G^2/2 < E_{cut}$ are included - to improve this we just increase E_{cut}), it is not biased (charge can move without restriction to any point of the unit cell) and finally it is very simple to program (if an element with f orbitals is modelled, no further coding is needed, just a higher E_{cut} and more CPU time).

There are also some disadvantages with plane wave expansions. The principal of these is that an extremely large number of functions need to be used. For example, with an element such as silicon, a minimum of 100 planes waves per atom need to be used. Some elements are particularly difficult - for example first period elements such as carbon and oxygen or the $3d$ transition metals. This means that the time and memory requirements of such a code will be considerable. Furthermore, if a unit cell contains 100 silicon atoms and just one oxygen atom, the presence of one difficult atom (as is often present in a defect) controls the size of the basis and results in an unnecessarily flexible descriptive power having to be employed at all points in the unit cell. One way of avoiding this is to use a technique pioneered by Gygi [62] and applied to solids by Hamann [63]. The alternative is to use localised orbitals:

$$\psi_\lambda(\mathbf{r}) = \sum_i c_i^\lambda \phi_i(\mathbf{r}). \tag{20}$$

A common choice for the functions ϕ_i is Gaussian type orbitals:

$$\phi_i(\mathbf{r}) = (x-R_{ix})^{n_1}(y-R_{iy})^{n_2}(z-R_{iz})^{n_3} e^{-a_i(\mathbf{r}-\mathbf{R}_i)^2},$$

where n_1, n_2 and n_3 are integers. If these are all zero they correspond to s-orbitals of spherical symmetry. Orbitals of p-symmetry correspond to one of these integers being unity and the others zero, whereas five d-like and one s-like orbital can be generated if $\sum_i n_i = 2$.

This expansion has the advantage that it is very efficient (results can be obtained with only 8 orbitals per atom, and quite well converged results with 16 functions), not dependent on atom type (the first–period elements such as carbon, nitrogen and oxygen can be treated just as easily as silicon, gallium, arsenic etc.), it is flexible (if we have one difficult atom, additional orbitals can be placed on just that atom so the overall speed of the calculation is not significantly affected). Disadvantages include the fact that the functions can become over–complete (numerical noise can enter a calculation if two functions with similar exponents are placed on the same atom), that they are difficult to program (especially if high angular momentum functions are needed), that it is difficult to test or to demonstrate absolute convergence (many things can be changed – the number of functions, the exponents, the location of the function centres).

One final advantage of localised orbitals is that the Hamiltonian matrix becomes sparse as the system size increases, and this is one feature that is important for the development of linear scaling methodologies which may eventually replace todays conventional methods.

6.3 Sampling of the Brillouin Zone

When working within periodic boundary conditions, the Kohn–Sham orbitals are evaluated for a given point $\mathbf{k}$ within the Brillouin zone and satisfy the Bloch condition. In terms of these, the charge density is given by

$$n(\mathbf{r}) = \sum_{n\mathbf{k}} f_{n\mathbf{k}} |\psi_{n\mathbf{k}}(\mathbf{r})|^2 \tag{21}$$

where is sum is over bands and allowed k-points in the Brillouin zone (and as we are modelling an infinite system there are an infinite number). $f_{n\mathbf{k}}$ is the occupancy of the band n at the k–point $\mathbf{k}$. In practice this sum can be well approximated by a small set of carefully chosen points, especially for systems with a band gap such as the semiconductors considered in this volume. These points, known as *special points*, are placed in the *irreducible* Brillouin

zone (IBZ), which for our purposes is the smallest volume of the Brillouin zone, which when operated on by all the space group operations of the system, covers the whole zone. For bulk silicon, the IBZ is 1/48th of the whole zone. Baldereschi [64] looked for the single best point, if only one were chosen. A more general scheme was proposed by Chadi and Cohen [65] who produced a set of equations that can be solved for a chosen density of k-points in the zone. Monkhorst and Pack produced a scheme based on equally spaced points [66]. All these schemes converge rapidly for insulators, but slowly for metals where much work has to be done to define the shape of the Fermi surface.

6.4 Traditional Diagonalisation Techniques

Early implementations of the LDA preceded by taking matrix elements of the Kohn–Sham Hamiltonian in basis functions and solving the resulting eigenvalue problem. The first condensed matter application looked at properties of bulk semiconductors and therefore used a basis of plane waves. The formalism for this is presented by Ihm, Zunger and Cohen [67] and Yin and Cohen [68]. Further reviews have been given by Ihm [69], Srivastava and Weire [70] and the direct approach is discussed briefly by Payne *et al.*,[71]. The main points are

1. An initial input charge density is chosen. This could be a linear combination of atomic charge densities and is required in Fourier components :
$$n(\mathbf{r}) = \sum_{\mathbf{G}} n(\mathbf{G}) \exp[i\mathbf{G} \cdot \mathbf{r}]. \tag{22}$$
2. The Fourier components of the Hartree potential are readily found from this by the analytic solution of Poisson's equation. The components are $V^H(\mathbf{G}) = 4\pi n(\mathbf{G})/G^2$.
3. The Fourier components of the exchange–correlation potential are found. First the Fourier components of charge density $n(\mathbf{G})$ are found on a equally spaced grid in real space using a fast Fourier transform. This is an extremely rapid step requiring only $O(N \ln N)$ operations for N plane waves. The exchange correlation potential is then evaluated on this grid, using the LDA locally at each point (some additional problems that occur when using a GGA are discussed by White and Bird [72]) and a second FFT is performed to obtain the Fourier components.
4. The matrix elements of all local potentials (the local part of the pseudopotential and the two many–body terms described above) and the kinetic energy operator are trivially found in a plane wave basis.
5. The matrix element of the non–local part of the pseudopotential is more involved, and the traditional evaluation is described in [67].
6. All these terms are added together to give the full Hamiltonian, which is then diagonalised giving the Kohn–Sham eigenvalues and eigenvectors.
7. The *output* charge density corresponding to these solutions is found using special k-point sampling described above. This will not be the same as the input charge density we started with above.
8. An iterative procedure is used to move to a self–consistent solution. Some methods for this are reviewed in [39].
9. Care needs to be taken when finding the total energy, especially with the $\mathbf{G} = 0$ Fourier component. This is discussed in [67].
10. The forces acting on the atoms are found using the Hellmann–Feynman theorem [73, 74] which in this case involves finding the expectation value of the derivative of the pseudopotential, and adding on the derivative of the ion–ion interaction term. This is very much faster than finding a total energy, but care has to be taken in monitoring the convergence as the force is quite sensitive to errors in the charge density (more more so than the energy, which is of course variationally protected, giving a quadratic dependence on errors).

A similar procedure can be followed when using localised orbitals, but this time different groups have developed different methodologies. The main differences are

1. The matrix problem is now a *generalised eigenvalue problem* as the basis functions are not in general orthogonal.
2. Matrix elements of the kinetic energy and pseudopotential are more difficult to program, but are less demanding than for the plane wave case above as the work done is only $O(N)$ because of the localisation of the basis functions.

3. The Hartree potential raises a significant challenge. A direct coding of this introduces four–centre integrals

$$V_{ij}^H = \int \phi_i V^H(\mathbf{r})\phi_j d\mathbf{r} = \sum_{kl} b_{kl} \int \frac{\phi_i(\mathbf{r})\phi_j(\mathbf{r})\phi_k(\mathbf{r}')\phi_l(\mathbf{r}')\, d\mathbf{r}d\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|} \tag{23}$$

where b_{ij} is the charge density matrix in this basis. Much work has been invested into the evaluation of these integrals, as they also occur in implementations of Hartree–Fock calculations, but their evaluation is still a demanding task. These can be avoided by either solving Poisson's equation numerically [75] or, most efficiently, by introducing an intermediate fitting basis [58]. The latter approach has the advantage that the total energy, Hartree potential and force and all be found consistently in the same approximation, so exact internal consistency is maintained in the calculation.

4. The exchange correlation potential also causes problems. This can be found either using a discrete grid or again more efficiently by using an analytic approximation [58]. Again, the analytic approach has the advantage that exact internal consistency is maintained, but the disadvantage that a further approximation has been made.

5. The matrix has to be set up and diagonalised as above.

6. The evaluation of forces is more complex than in the case of a plane wave basis, as localised orbitals are placed on atoms and move with the atoms. To find the forces, all terms in the total energy therefore need to be differentiated as this hidden dependence on atom position is passed to the charge density. This means that force evaluation, although still much faster than the self–consistent cycle is a more significant contribution to the calculation than it is in plane wave calculations.

Direct diagonalisation is not now used for plane wave calculations as it would be far too costly. A single workstation can comfortably diagonalise matrices of size around 1000. As a minimum of 100 plane waves per atom are needed, this corresponds to only 10 atoms and this is especially inefficient as only 1-2% of the wavefunctions are needed. The first improvement was the use of iterative diagonalisation in which only the occupied Kohn–Sham states were found, although this has been superseded by the methods described below. With localised orbitals, direct diagonalisation does not imply such a large overhead, as the number of occupied states is a much greater percentage of the total number. This is still used by many groups.

6.5 Car-Parrinello Molecular Dynamics

The Car-Parrinello methodology has become a widespread method of solving the Kohn–Sham equations. Molecular dynamics has been a popular way of modelling systems using empirical potentials for many years. The principle has been to form an interatomic potential that returns the total energy as a function of the atom positions $V(\{\mathbf{R}_i\})$. Motion of the atoms in governed by a Lagrangian

$$L = \frac{1}{2}\sum_i M_i \dot{\mathbf{R}}_i^2 - V(\{\mathbf{R}_i\})$$

and equations of motion

$$M_i \ddot{\mathbf{R}}_i = -\frac{\partial V}{\partial \mathbf{R}_i}$$

These can then be integrated and properties of the system evaluated. For example the temperature of the system can be increased, leading to melting, the liquid studied, the system could then be rapidly cooled and the defects frozen in studied and so on.

Of course, this calculation is not quantum–mechanical. However, Car and Parrinello [76] have put forward a formalism which enables the function $V(\{\mathbf{R}_i\})$ to be treated quantum mechanically within the LDA. The Lagrangian is of the form

$$L = \frac{1}{2}\sum_i M_i \dot{R}_i^2 + \frac{1}{2}\mu \sum_\lambda \int |\dot{\psi}_\lambda|^2 d\mathbf{r} - E[\{\mathbf{R}_i\},\{\psi_\lambda\}]$$

where E is the Kohn–Sham energy functional, μ is an arbitrary fictitious mass the ψ_λ are subject to the holonomic constraints of orthonormality. The resulting equations of motion are

$$\begin{aligned} \mu\ddot{\psi}_\lambda(\mathbf{r},t) &= -\hat{H}\psi_\lambda(\mathbf{r},t) + \sum_\nu \Lambda_{\lambda\nu}\psi_\nu \\ M_i\ddot{\mathbf{R}}_i &= -\nabla_{\mathbf{R}_i} E \end{aligned} \tag{24}$$

where the Λ's are Lagrange multipliers. These equations are integrated, and if the kinetic energy of the nuclei and fictitious kinetic energy of the electronic degrees of freedom are gradually reduced (either by directly scaling the velocities, by including a dissipative term or by attaching a Nosé–Hoover thermostat) the equilibrium state of minimum energy is attained. Car and Parrinello showed that this method reproduced the values of bulk properties such as phonon modes given by static calculations. One interesting feature of this method is that the calculation can move off the Born–Oppenheimer surface discussed above. Advantages of this include the fact that the Hamiltonian does not need to be stored, it is not explicitly diagonalised, reducing the workload from $O(N^3)$ to $O(NM^2)$ where N is the number of basis functions and M the number of occupied levels. Further discussion of practical details (how orthonormality of the orbitals is maintained and how the equations of motion are best integrated) is given by Payne *et al.*,[71].

The applications of this are very numerous, with many notable successes. An early study was of amorphous silicon [77]. Calculation of diffusion constants can has also been attempted [78]. More recently, finite temperature thermodynamic properties have been found using an extension of this technique [79]. More details of the method and its successes are given in [80].

6.6 Direct minimisation of the Kohn–Sham functional

An alternative strategy is to regard the Kohn–Sham energy as a function of all the wavefunction coefficients and to use a standard technique from numerical analysis to carry out the minimisation process. One significant advantage of this is that we can guarantee that the energy always decreases from iteration to iteration, removing instabilities which it has been claimed [71] makes the Car–Parrinello molecular dynamics method unsuitable for application to large systems [71]. The conjugate gradient method achieves this minimisation by making a series of one–dimensional line minimisations. Again, a significant advantage is that the Hamiltonian doesn't need to be stored, and that only the action of the Hamiltonian on the wavefunctions is required. The details of the procedure applied with a plane wave basis is given in [71] and in localised orbitals in [75]. Key issues are how the bands are updated, how orthogonality constraints are imposed and what form of preconditioning is applied.

7 Linear Scaling Methodologies

7.1 Introduction

All of the conventional implementations of both density functional theory and Hartree–Fock theory described above have a memory requirement that is proportional to N^2 and a computing time that is proportional to N^3 where N is a measure of the system size being modelled (i.e. it is related to either the number of atoms, the number of electrons or, in some calculations, to the volume of the unit cell). Both of these are serious problems.

In the case of plane–wave calculations in which the wavefunction is expanded in a basis of plane waves, between 100 and 1000 plane waves per atom may be required depending on the problem being considered. If the number of occupied states is say twice the number of atoms, then if we are performing a spin–polarised calculation using complex arithmetic, the memory requirement is scales as N^2 and is of the order 100MB for 100 atoms and 10GB for 1000 atoms. This is very considerable requirement and is in itself a serious problem. Furthermore, the computing time scales as N^3 and this also puts a serious limit on the number of atoms that can be modelled.

It seems clear that the future of computing lies with parallel processors. As a simple example, if we say that a single workstation can model a 50 atom system in an acceptable time then a 256 processor machine can model a system containing $50 \times \sqrt[3]{256}$ or approximately 300, an increase of of only a factor six. Using this argument, it seems unlikely that much more than 1000 atoms can be treated in a *routine* manner with these methods, even with an order of magnitude increase in available resources. There is a significant research effort underway at present which attempts to surmount this problem and achieve *linear scaling* in which both the memory requirement and the processing time scale linearly with the system size.

It is well known that this N^3 scaling can be overcome. This has been emphasised by Heine and co–workers [81] who have expressed properties in terms of Green functions or density matrices and have developed the *recursion method* to exploit this in the context of tight binding theory. This has been implemented within the LDA [82]. More recently a number of alternative approaches have been attempted. We will look at each of these in turn.

7.2 Divide and Conquer methods

This method was implemented by Yang [83, 84] and was probably the first linear scaling method. The system in question is divided up into a number of overlapping subsystems, that are treated semi-independently. The Hamiltonian for each sub-system includes the potential from the other subsystems and is diagonalised independently using a conventional LCAO method (in $O(N^3)$ operations). The charge density is then extracted and the potential updated. This is repeated until self-consistency is achieved.

7.3 Localisation of Orbitals

In this approach two significant modifications are made to the standard approaches outlined above.

1. The orthogonalisation step required by both Car–Parrinello and direct minimisation approaches is omitted. In place of this, an *unconstrained* minimisation is performed on a modified energy functional [85, 86], which is designed to push the electron states towards orthogonality, achieving this at the minimum. In this, the sum of one–electron Kohn–Sham eigenvalues (referred to as the *band structure energy* in tight binding approaches) is given by
$$\tilde{E}_{BS} = 2\left[\sum_{i}^{N} H_{ii} - \sum_{ij}(S_{ij} - \delta_{ij})H_{ji}\right] \tag{25}$$
which is defined for $2N$ electrons in states $|\psi_i\rangle$ and $S_{ij} = \langle\psi_i|\psi_j\rangle$ is the overlap matrix, which is different from the unit matrix δ_{ij} if the states are non–orthogonal.
2. Linear scaling is obtained when the set of functions $|\psi_i\rangle$ defining the ground state are chosen to be localised. This set of functions is not unique, since any unitary transformation of these has the same energy. In particular, we can describe the ground state using orthonormal localised wavefunctions or *Wannier functions*. These are exponentially localised in an insulator and have power–law localisation in a metal. In practice, non–orthogonal Wannier–like functions (WLF) are used. These are forced to decay rapidly with distance by expanding each $|\psi_i\rangle$ in terms of an underlying basis of *localised* orbitals, where only the basis functions within a certain cut–off radius are included. Localisation of the wavefunctions can also be achieved with an underlying plane wave basis [87] but at the expense of an inversion of the overlap matrix, a step best avoided.

This approach has been applied in a self–consistent form to large systems containing up to 1000 atoms [88], and in non–self–consistent implementations up to 3840 atoms [89].

The functionals defined above are for exactly $2N$ states, and it has been suggested that this leads to spurious local minima in the functional. A generalised functional which allows for an arbitrary number of orbitals has been proposed by Kim, Mauri and Galli [90].

7.4 Charge density matrix method

The next method we consider is based on the one–particle charge density matrix $\rho(\mathbf{r}, \mathbf{r}')$. The total energy can be written in terms of this quantity as :

$$E_{TOT} = -\int [\nabla^2_{\mathbf{r}}\rho(\mathbf{r}, \mathbf{r}')]_{\mathbf{r}=\mathbf{r}'}\, d\mathbf{r} + 2\int V_{ps}(\mathbf{r}, \mathbf{r}')\rho(\mathbf{r}, \mathbf{r}')d\mathbf{r}d\mathbf{r}' + \frac{1}{2}\int \frac{n(\mathbf{r})n(\mathbf{r}')d\mathbf{r}d\mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|} + \int n(\mathbf{r})\epsilon_{xc}[n(\mathbf{r})]d\mathbf{r} + E_{i-i}$$

where the charge density may be written as $n(\mathbf{r}) = 2\rho(\mathbf{r}, \mathbf{r})$.

The quantity $\rho(\mathbf{r}, \mathbf{r}')$ has the property that it is short–ranged as a function of $|\mathbf{r} - \mathbf{r}'|$, a property referred to as *near–sightedness* by Kohn [91]. Typically, a set of auxiliary basis functions are introduced :

$$\rho(\mathbf{r}, \mathbf{r}') = \sum_{ij} \rho_{ij}\phi_i(\mathbf{r})\phi_j(\mathbf{r}') \tag{26}$$

where the functions $\phi_i(\mathbf{r})$ are *localised*, so for example $\phi_i(\mathbf{r}) = f(\mathbf{r} - \mathbf{R}_i)$ where $f(\mathbf{r}) = 0$ if $|\mathbf{r}| > R_c$. This is extremely important when evaluation the energy, as only $O(N)$ operations need to be performed. A detailed description of how this is done is contained in [92] and [93].

The most significant challenge comes with the minimisation of the total energy with regard to the charge density matrix. This is not straightforward as the minimisation should be carried out subject to certain constraints — the charge density matrix must be Hermitian, correctly normalised and also *idempotent*, that is $\hat{\rho}^2 = \hat{\rho}$. This last requirement is analogous to demanding that the Kohn-Sham orbitals be orthogonal and presents a serious challenge to progress. Two main approaches have been made:

1. An alternative functional $E[\rho]$ is introduced such that unconstrained minimisation of this automatically gives an idempotent ρ which is the same as would be obtained by the constrained minimisation of the above functional. One such functional was introduced by Kohn [91], although the functional is complicated and includes a penalty function that has square–root behaviour at the minimum. Other minimum energy principles have been recently published by Yang [94].

2. Another method is to use a purifying transformation. This was suggested by Li, Nunes and Vanderbilt [95] in the context of tight–binding theory and later by Hernández, Gillan and Goringe [92] in the LDA. In this, we assume that we have a matrix $\tilde{\rho}(\mathbf{r},\mathbf{r}')$ which is approximately idempotent and then perform the transformation

$$\rho(\mathbf{r},\mathbf{r}') = 3\int d\mathbf{r}''\tilde{\rho}(\mathbf{r},\mathbf{r}'')\tilde{\rho}(\mathbf{r}'',\mathbf{r}') - 2\int d\mathbf{r}''d\mathbf{r}'''\tilde{\rho}(\mathbf{r},\mathbf{r}'')\tilde{\rho}(\mathbf{r}'',\mathbf{r}''')\tilde{\rho}(\mathbf{r}''',\mathbf{r}')$$

   which produces a quantity $\rho(\mathbf{r},\mathbf{r}')$ which is closer to being idempotent. The total energy is evaluated with this quantity, but the minimisation is carried our with respect to $\tilde{\rho}$. The locality of $\tilde{\rho}$ ensures that the above transformation can also be carried out in $O(N)$ operations.

7.5 Grid based methods

The solution of the Kohn–Sham equations has invariably been attempted via expansion in a basis, and surprisingly little work has used grid–based approaches common in many other branches of computational physics. A real–space multi–grid based approach which has linear scaling has recently been proposed by Briggs *et. al.* [96]. This method used a real space grid as a basis and used the multigrid technique to accelerate the convergence over the different length scales present in the systems simulated. Some different technical problems need to be solved in this approach compared to the basis expansion methods more routinely adopted. In particular, care must be taken with the action of the kinetic energy operator as this cannot be evaluated exactly; some form of *filtering* must be included to reduce undesirable effects when atoms move over grid points (this introduces unphysically high wave-vector components into potentials which must be filtered out); local enhancement of the grid should be made in regions where the charge density is rapidly varying. The implementation is described in detail in ref [96] together with references to other real space methods.

8 Quantum Monte–Carlo Methods

Quantum–Monte Carlo methods have began to be implemented on real materials only in the last few years as they are extremely computationally intensive.

The first form of quantum Monte–Carlo is variational quantum Monte–Carlo and this is essentially a variational calculation. Typically, the *fixed–node* approximation is used. In this the (many–body) wavefunction is written as

$$\Psi(\mathbf{r}_1,\ldots,\mathbf{r}_N) = \exp[J]\det(\uparrow)\det(\downarrow) \tag{27}$$

where $\det(\uparrow)$ and $\det(\downarrow)$ are the determinantal functions constructed from the spin–up and spin–down solutions of the Kohn-Sham equations for the system in question. In the absence of the factor $\exp[J]$ the minimisation of the expectation value of the Hamiltonian in this function would produce the Hartree–Fock energy. The purpose of the prefactor is to introduce correlation into the wavefunction. One form used has been [97]:

$$J = \sum_{i,s}\chi_s(\mathbf{r}_i^s) - \sum_{ijss'}u_{ss'}(|\mathbf{r}_i - \mathbf{r}_j|) \tag{28}$$

where i,j label the electrons and s the spin state. The functions χ and u are parametrised and have a form which implements the cusp condition when $\mathbf{r}_i = \mathbf{r}_j$ [97, 98].

Monte Carlo methods are used to evaluate the expectation value of the Hamiltonian with this wavefunction (a 3N-dimensional integral). A discussion of this is given in [97]. This is then minimised with respect to the parameters in the Jastrow factor. The result is the ground state energy (which will contain statistical noise) and wavefunction.

The lattice constant and bulk modulus of carbon and silicon have been given to an accuracy comparable with the LDA (which is already excellent), but the cohesive energy is far superior. The result for diamond is 7.45 eV/atom (to be compared with 8.63 eV in the LDA, 5.85 eV in HF theory and 7.37 eV from experiment). The cohesive energy of silicon is given as 4.88eV/atom (to be compared with 5.29 eV/atom in the LDA, 3.66 eV/atom in HF theory and a series of experimental values varying from 4.62–4.88 eV/atom.

The second form of Monte–Carlo is known as *diffusion quantum Monte–Carlo* (DQMC) and in principle does indeed find the ground state energy. This works by solving the time-dependent Schrödinger equation

$$i\frac{\partial\Psi}{\partial t} = [\hat{H} - E_T]\Psi \tag{29}$$

where $\hat{H} = \hat{T}+U$ is the same as given in equation (3) apart from that fact that the Coulomb interaction with the nuclei is replaced by a non–local pseudopotential, and E_T is a constant which merely changes the phase of the wavefunction. Now, changing to imaginary time ($s = it/\hbar$), we have:

$$\frac{\partial\Psi}{\partial s} = \frac{1}{2}\sum_i \nabla_i^2\Psi + (E_T - U)\Psi \tag{30}$$

Clearly, this resembles the diffusion equation with an additional term which provides structure to the solution. This resemblance to the diffusion equation gives the method its name. The solution to this may be expanded in energy eigenstates :

$$\Psi(\mathbf{r}, s) = \sum_n \psi_n(\mathbf{r}) \exp[(E_n - E_T)s] \tag{31}$$

From this it is seen that if the shift, E_T, is less than the ground state energy, Ψ decays to zero. On the other hand, if $E_T > E_0$, the solution grows exponentially. Only if $E_T = E_0$ do we get a stable wavefunction which, after all the excited states have died out, is exactly the ground state wavefunction. In this way the ground state wavefunction and energy can be solved. Again, stochastic methods are used to solve the equation. More background to this is contained in Refs. [99] and [100].

This procedure is extremely time consuming and is only possible for small systems with of order 10 atoms. The energies are however significantly lower than achieved with variational quantum Monte–Carlo. This was first applied to properties of bulk silicon by Li, Ceperley and Martin [101], obtaining results that were slightly better than those obtained by variational Monte Carlo. Several advances have been made since this work, particularly with regard to the elimination of the various finite size effects present in the calculation, by incorporating standard k–point sampling techniques [102] or by using modified interaction potentials [103].

Programmes of work are in place to study defects with these methods. These calculations are already within reach as in Monte–Carlo theory, large unit cells of 16 or 54 atoms are necessary, even to model the bulk material. As this is already possible, we may not have the long time–delay that was experienced with other techniques while computers increased in power to enable the size of system to increase from the two-atom unit cell of silicon to the 54 atoms required to model defects and more complex processes. Today, a 50 atom quantum Monte–Carlo calculation is no more demanding with current computing power than was a 50 atom LDA calculation 15 years ago on the machines available then.

References

[1] M. Stoneham, [*Defects in Solids*, Oxford University Press, (London, 1975.)]

[2] J. Bourgoin and M. Lannoo, [*Point Defects in Semiconductors II*, Springer–Verlag series **35**, (New York, 1983)]

[3] J. C. Slater, [*The Self–Consistent Field for Molecules and Solids vol 4*, McGraw–Hill, (New York, 1974)]

[4] P. Hohenberg and W. Kohn, [*Phys. Rev.* vol. 136 (1964) p.864B]

[5] U. Von Barth and L. Hedin, [*J. Phys. C: Sol. State Phys.* vol. 5 (1972) p.1629]

[6] M. N. Pant and A. K. Rajagopal, [*Solid State Commun.* vol. 10 (1972) p.1157]

[7] O. Gunnarsson and B. I. Lundqvist, [*Phys. Rev. B* vol. 13 (1976) p.4274]

[8] N. D. Mermin, [*Phys. Rev.* vol. 137 (1965) p.A1441]

[9] J. E. Harriman, [*Phys. Rev. A* vol. 24 (1981) p.680]

[10] G. Zumbach, K. Maschke, [*Phys. Rev. A* vol. 28 (1983) p.544]

[11] M. Pearson, E. Smargiassi and P. A. Madden, [*J. Phys. C: Sol. State Phys.0M* vol. 5 (1993) p.3321] ; E.Smargiassi and P.A.Madden, [*Phys. Rev. B* vol. 51 (1995) p.117]

[12] W. Kohn and L. J. Sham, [*Phys. Rev.* vol. 140 (1966) p.1133A]

[13] R. G. Parr and W. Yang, [*Density functional theory of atoms and molecules*, Oxford University Press, (New York, 1989)]

[14] R. M. Dreizler and E. K. U. Gross, [*Density Functional Theory*, Springer–Verlag, (Berlin, 1990)]

[15] M. Gell–Mann and K. A. Brueckner, [{*PRV* vol. 106 (1957) p.364]

[16] D. M. Ceperley, [*Phys. Rev. B* vol. 18 (1978) p.3126] D. M. Ceperley and B. J. Alder, [*Phys. Rev. Lett.* vol. 45 (1980) p.566]

[17] J. P. Perdew and A. Zunger, [*Phys. Rev. B* vol. 23 (1981) p.5048]

[18] S. H. Vosko, L. Wilk and M. Nusair, [*Can. J. Phys.* vol. 58 (1980) p.1200]

[19] J. P. Perdew and Y. Wang, [*Phys. Rev. B* vol. 45 (1992) p.13244]

[20] J. P. Perdew and Y. Wang, [*Phys. Rev. B* vol. 33 (1986) p.8800]

[21] J. P. Perdew, [*Phys. Rev. B* vol. 33 (1986) p.8822]

[22] M. Levy, [in *Density Functional Theory* Ed. , (E.K.U.Gross and R.M.Dreizler, Plenum Press, New York1995)]

[23] M. Levy and J. P. Perdew, [*Phys. Rev. B* vol. 48 (1993) p.11638]

[24] A. D. Becke, [*J. Chem. Phys* vol. 85 (1986) p.7184]

[25] A. D. Becke, [*Phys. Rev. A* vol. 38 (1988) p.3098]

[26] J. P. Perdew, [in *Electronic Structure of Solids '91* Ed. P.Ziesche and H.Eschrig (Akademie Verlag, Berlin, 1991)]

[27] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, C. Fiolhais, [*Phys. Rev. B* vol. 46 (1992) p.6671]

[28] G. Oritz and P. Ballone, [*Phys. Rev. B* vol. 43 (1991) p.6376]

[29] J. P. Perdew, K. Burke and M. Ernzerhof, [*Phys. Rev. Lett.* vol. 77 (1996) p.3865]

[30] J. P. Perdew, K. Burke and M. Ernzerhof, [*Phys. Rev. B* vol. 54 (1996) p.16533]

[31] M. T. Yin and M. L. Cohen, [*Phys. Rev. B* vol. 26 (1982) p.5668]

[32] R. J. Needs and A. Mujica, [*Phys. Rev. B* vol. 51 (1995) p.9652]

[33] K. Kunc and R. M. Martin, [in *Ab Initio Calculation of Phonon Spectra* Ed. J.T.Devreese, V.E.van Doren, P.E.van Camp (Plenum, New York, 1983)]

[34] P. Giannozzi, S. de Gironcoli, P. Pavone and S. Baroni, [*Phys. Rev. B* vol. 43 (1991) p.7231]

[35] P. R. Briddon and R. Jones, [*Phys. Rev. B* vol. 64 (1990) p.2535]

[36] I. Stich *et al.*,, [*Phys. Rev. Lett.* vol. 71 (1993) p.3613]

[37] C. P. Ewels, R. Jones, S. Öberg, J. Miro, P. Deák, [*Phys. Rev. Lett.* vol. 77 (1996) p.865]

[38] C. G. Van de Walle and P. E. Blöchl, [*Phys. Rev. B* vol. 47 (1993) p.4244]

[39] W. E. Pickett, [*Comput. Phys. Rep.* vol. 9 (1989) p.115]

[40] G. Lee, M. H. Lee and J. Ihm, [*Phys. Rev. B* vol. 52 (1995) p.1459]

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

[42] D. R. Hamann, M. Schlüter and C. Chiang, [*Phys. Rev. Lett.* vol. 48 (1982) p.1425]

[43] E. L. Shirley, D. C. Allan, R. M. Martin and J. D. Joannopoulos, [*Phys. Rev. B* vol. 40 (1989) p.3652]

[44] A. Fillippetti, D. Vanderbilt, W. Zhong, Y. Cai, G. B. Bachelet, [*Phys. Rev. B* vol. 52 (1995) p.11793]

[45] G. P. Kerker, [*J. Phys. C: Sol. State Phys.* vol. 13 (1980) p.L189]

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

[47] D. Vanderbilt, [*Phys. Rev. B* vol. 32 (1985) p.8412]

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

[49] D. Vanderbilt, [*Phys. Rev. B* vol. 41 (1990) p.7892]

[50] L. Laasonen, R. Car, C. Lee and D. Vanderbilt, [*Phys. Rev. B* vol. 43 (1991) p.6796]

[51] L. Kleinman and D. M. Bylander, [*Phys. Rev. Lett.* vol. 48 (1982) p.1425]

[52] D. M. Bylander and L. Kleinman, [*Phys. Rev. B* vol. 41 (1990) p.907]

[53] X. Gonze, P. Kackell and M. Scheffler, [*Phys. Rev. B* vol. 42 (1990) p.12264]

[54] X. Gonze, R. Strumpf and M. Scheffler, [*Phys. Rev. B* vol. 44 (1991) p.8503]

[55] D. H. Vanderbilt, [*Phys. Rev. B* vol. 41 (1990) p.7892]

[56] P. E. Bloechl, [*Phys. Rev. B* vol. 41 (1990) p.5414]

[57] G. Makov and M. C. Payne, [*Phys. Rev. B* vol. 51 (1995) p.4014]

[58] R. Jones, and P. R. Briddon, *The Ab Initio Cluster Method and the Dynamics of Defects in Semiconductors*, in *Identification of Defects in Semiconductors*, ed. M. Stavola, *Semiconductors and Semimetals*, treatise editors, R. K. Willardson, A. C. Beer, and E. R. Weber, Academic Press.), in press (1998).

[59] G. A. Baraff and M. Schlüter, [*Phys. Rev. B* vol. 19 (1979) p.4965]

[60] J. Bernholc, N. O. Lipari and S. T. Pantelides, [*Phys. Rev. B* vol. 21 (1980) p.3545]

[61] U. Lindefelt and A. Zunger, [*Phys. Rev. B* vol. 24 (1981) p.5913]

[62] F. Gygi, [*Phys. Rev. B* vol. 48 (1993) p.11692]

[63] D. R. Hamann, [*Phys. Rev. B* vol. 51 (1995) p.9508]

[64] A. Baldereschi, [*Phys. Rev. B* vol. 7 (1973) p.5212]

[65] D. J. Chadi and M. L. Cohen, [*Phys. Rev. B* vol. 8 (1973) p.5747]

[66] H. J. Monkhorst and J. D. Pack, [*Phys. Rev. B* vol. 13 (1976) p.5188]

[67] J. Ihm, A. Zunger and M. L. Cohen, [*J. Phys. C: Sol. State Phys.* vol. 12 (1979) p.4409]

[68] M. T. Yin and M. L. Cohen, [*Phys. Rev. B* vol. 25 (1982) p.7403]

[69] J. Ihm, [*Rep. Prog. Phys.* vol. 51 (1988) p.105]

[70] G. P. Srivastava and D. Weaire, [*Adv. Phys.* vol. 36 (1987) p.463]

[71] M. C. Payne, ¡. P. Teter, D. C. Allan, T. A. Arias, J. Joannopoulos, [*Rev. Mod. Phys.* vol. 64 (1992) p.1045]

[72] J. A. White and D. Bird, [*Phys. Rev. B* vol. 50 (1994) p.4954]

[73] Eifuhrung in die Quantumchemie, [*Devticke*, Leipzeig, (1937,)]

[74] R. P. Feynman, [*Phys. Rev.* vol. 56 (1939) p.340]

[75] X. J. Chen, J. M. Langlois and W. A. Goddard, [*Phys. Rev. B* vol. 52 (1995) p.2348]

[76] R. Car and M. Parrinello, [*Phys. Rev. Lett.* vol. 55 (1985) p.2471]

[77] R. Car, M. Parrinello and M. C. Payne, [*Phys. Rev. B* vol. 44 (1991) p.11092]

[78] F. Buda, G. L. Chiarotti, R. Car and M.Parrinello, [*Phys. Rev. Lett.* vol. 63 (1989) p.294]

[79] O. Sugino and R. Car, [*Phys. Rev. Lett.* vol. 74 (1995) p.1823]

[80] G. Galli and A. Pasquarello, [in *Computer Simulation in Chemical Physics* Ed. M. P. Allen and D. J. Tildesley (Kluwer, Amsterdam, 1993)]

[81] F. Seitz, C. Turnball and H. Ehrenreich (Eds) [*Solid State Physics*, vol.35 (Academic Press, New York, 1980)]

[82] S. Baroni and P. Giannozzi, [*Euro. Phys. Lett.* vol. 17 (1992) p.547]

[83] W. Yang, [*Phys. Rev. Lett.* vol. 66 (1991) p.1438]

[84] W. Yang, [*Phys. Rev. A* vol. 44 (1991) p.7823]

[85] P. Ordejón, D. A. Drabold, R. M. Martin and M. P. Grumbach, [*Phys. Rev. B* vol. 51 (1995) p.1456]

[86] F. Mauri, G. Galli and R. Car, [*Phys. Rev. B* vol. 47 (1993) p.9973]

[87] G. Galli and M. Parrinello, [*Phys. Rev. Lett.* vol. 69 (1992) p.1077]

[88] P. Ordejón, E. Artacho and J. M. Soler, [*Phys. Rev. B* vol. 53 (1996) p.R10441]

[89] S. Itoh, P. Ordejón, D. A. Drabold and R. M. Martin, [*Phys. Rev. B* vol. 53 (1996) p.2132]

[90] J. Kim, F. Mauri and G. Galli, [*Phys. Rev. B* vol. 52 (1995) p.1640]

[91] W. Kohn, [*Phys. Rev. Lett.* vol. 76 (1996) p.3168]

[92] E. Hernández, M. J. Gillan and C. M. Goringe, [*Phys. Rev. B* vol. 53 (1996) p.7147]

[93] E. Hernández, M. J. Gillan and C. M. Goringe, [*Phys. Rev. B* vol. 55 (1997) p.13485]

[94] W. Yang, [*Phys. Rev. B* vol. 56 (1997) p.9294]

[95] X. P. Li, R. W. Nunes and D.Vanderbilt, [*Phys. Rev. Lett.* vol. 47 (1993) p.10891]

[96] E. L. Briggs, D. J. Sullivan and J. Bernholc, [*Phys. Rev. B* vol. 54 (1996) p.14362]

[97] S. Fahy, X. W. Wang, S. G. Louie, [*Phys. Rev. B* vol. 42 (1990) p.3503]

[98] A. K. Williamson *et al.*,, [*Phys. Rev. B* vol. 53 (1996) p.9640]

[99] D. M. Ceperley and M. H. Kalos, [in *Monte Carlo Methods in Statistical Physics* Ed. K. Binder (Springer, Berlin, 1979)]

[100] R. M. Martin and V. D. Natoli, [in *Computational Approaches to Novel Condensed Matter Systems* Ed. D. Neilson and M. P. Das (Plenum Press, New York, 1995)]

[101] X. P. Li, D. M. Ceperley and R. M. Matrin, [*Phys. Rev. B* vol. 44 (1991) p.10929]

[102] G. Rajagopal, R. J. Needs, S. Kenney, W. M. C. Foulkes and A. James, [*Phys. Rev. Lett.* vol. 73 (1994) p.1959]

[103] A. J. Williamson, G. Rajagopal, R. J. Needs, L. M. Fraser, W. M. C. Foulkes, Y. Wang, M. Y. Chou, [*Phys. Rev. B* vol. 55 (1997) p.R4851]