Ab initio molecular dynamics: basic concepts, current trends and novel applications

2002 J. Phys.: Condens. Matter 14 R1297

(http://iopscience.iop.org/0953-8984/14/50/202)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **14** (2002) R1297–R1355 PII: S0953-8984(02)30651-9

TOPICAL REVIEW

Ab initio molecular dynamics: basic concepts, current trends and novel applications

Mark E Tuckerman

Department of Chemistry and Courant Institute of Mathematical Sciences, New York University, NY 10003, USA

Received 25 September 2002, in final form 20 November 2002
Published 6 December 2002
Online at stacks.iop.org/JPhysCM/14/R1297

Abstract

The field of *ab initio* molecular dynamics (AIMD), in which finite temperature molecular dynamics (MD) trajectories are generated with forces obtained from accurate 'on the fly' electronic structure calculations, is a rapidly evolving and growing technology that allows chemical processes in condensed phases to be studied in an accurate and unbiased way. This article is intended to present the basics of the AIMD method as well as to provide a broad survey of the state of the art of the field and showcase some of its capabilities. Beginning with a derivation of the method from the Born–Oppenheimer approximation, issues including the density functional representation of electronic structure, basis sets, calculation of observables and the Car–Parrinello extended Lagrangian algorithm are discussed. A number of example applications, including liquid structure and dynamics and aqueous proton transport, are presented in order to highlight some of the current capabilities of the approach. Finally, advanced topics such as inclusion of nuclear quantum effects, excited states and scaling issues are addressed.

(Some figures in this article are in colour only in the electronic version)

Contents

1. Introduction

Modern theoretical methodology, aided by the advent of high speed computing, has advanced to a level where the microscopic details of chemical processes in condensed phases can now be treated on a relatively routine basis. One of the most commonly used theoretical approaches for such studies is the MD method, in which the classical Newtonian equations of motion for a system are solved numerically starting from a prespecified initial state and subject to a set of boundary conditions appropriate to the problem. MD methodology allows both equilibrium thermodynamic and dynamical properties of a system at finite temperature to be computed, while simultaneously providing a 'window' onto the microscopic motion of individual atoms in the system. One of the most challenging aspects of an MD calculation is the specification of the forces. In many applications, these are computed from an empirical model or *force field*, in which simple mathematical forms are employed to describe bond, bend and dihedral angle potentials as well as van der Waals and electrostatic interactions between atoms, and the model is parametrized by fitting to experimental data or high level *ab initio* calculations on small clusters or fragments. This approach has enjoyed tremendous success in the treatment of systems ranging from simple liquids and solids to polymers and biological systems such as proteins and nucleic acids.

Despite their success, force fields have a number of serious limitations. First, charges appear as static parameters in the force field, and therefore electronic polarization effects are not included. This limitation has long been recognized, and attempts to rectify the problem have been proposed in the form of so-called polarizable models (see [1] for a review), in which charges and/or induced dipoles are allowed to fluctuate in response to a changing

environment. While these models have enjoyed considerable success, they also have a number of serious limitations, including a lack of transferability and standardization. Second, force fields generally assume a pre-specified connectivity among the atoms and, therefore, suffer from an inability to describe chemical bond breaking and forming events. The latter problem can be treated in an approximate manner using techniques such as the empirical valence bond method [2] or other semi-empirical approaches. However, such methods are also not transferable and, therefore, need to be reparametrized for each type of reaction and may end up biasing the reaction path in undesirable ways.

Because of the limitations of force field based approaches, one of the most important recent developments in MD, which addresses these problems, is the so-called *ab initio* molecular dynamics (AIMD) method [3–11], which combines finite temperature dynamics with forces obtained from electronic structure calculations performed 'on the fly' as the MD simulation proceeds (see, in particular, the recent excellent review by Marx and Hutter [10] for detailed discussions). Because the electronic structure is treated explicitly in AIMD calculations, many-body forces, electronic polarization and bond-breaking and forming events are described to within the accuracy of the electronic structure representation. Moreover, the AIMD methodology can be easily extended to incorporate nuclear quantum effects via the Feynman path integral approach [12, 13], leading to the *ab initio* path integral (AIPI) technique [14–17].

The AIMD and AIPI methods have been used to study a wide variety of chemically interesting and important problems in areas such as liquid structure, acid–base chemistry, industrial and biological catalysis, and materials. Applications include (but are certainly not limited to) calculations of the structure and dynamics of water [18–30] and other hydrogen-bonded liquids [31–35], structure and dynamics in acidic [36–41] and basic solutions [42, 43], proton transport in aqueous [44–48] and other environments [33, 34, 49–52] and in clusters [53–57], structure, proton order/disorder and dynamical properties of ice [27, 58–66], structure of liquid silicates and glasses [67–75], mechanisms of Ziegler–Natta industrial catalysis [76–78] and other surface catalytic processes [65, 79–83] and polymer knotting [84–87], to give just a sampling of the application areas. More recently, AIMD methods have started to impact the biological sciences and have been applied in calculations of NMR chemical shifts in drug–enzyme complexes [88], structure of nucleic acids [89], exploration of the design of possible biomimetics [90, 91] and structure, dynamics and binding mechanisms in myoglobin [92, 93]. In many of these applications, new physical phenomena have been revealed, which could not have been uncovered using empirical models, often leading to new interpretations of experimental data and even suggesting new experiments to perform.

Not unexpectedly, the power and flexibility of the AIMD (and AIPI) methodology come at the price of a significant increase in computational overhead compared to force field based approaches. Whereas the latter can currently be applied routinely to systems consisting of 10^4–10^6 atoms and access timescales on the order of tens of nanoseconds or longer, AIMD calculations can currently be applied routinely to systems of just a few tens or hundreds of atoms and access timescales on the order of tens of picoseconds. Moreover, whereas force fields can be finely tuned for specific situations, the accuracy of AIMD calculations is limited by the accuracy of the electronic structure method employed. Currently, the most commonly used electronic structure theory in AIMD is the Kohn–Sham formulation of density functional theory (DFT) [94–96], in which the electronic orbitals are expanded in a plane-wave basis set. This protocol provides a reasonably accurate description of the electronic structure for many types of chemical environment while maintaining an acceptable computational overhead and constitutes the original Car–Parrinello (CP) formulation of the method [3]. Clearly, then, AIMD calculations are limited by the accuracy of currently available density functionals. However, it is important to note that AIMD is a general approach that can be used with any

electronic structure method, and a number of examples exist in the literature which employ more accurate or more empirical electronic structure representations [97, 98] as well as different basis sets [99–101].

The organization of this article is as follows. In section 2, we shall begin with a derivation of the AIMD method starting from the Born–Oppenheimer approximation and arriving at the classical equations of motion on the exact ground electronic surface. We shall then discuss, in section 3, the use of the DFT representation of the electronic structure and various approximation schemes. Next, in section 4, we will discuss different basis set expansions of the electronic orbitals and indicate the advantages and disadvantages of different choices. Following this, in section 5, the adiabatic dynamics method and its relation to the CP algorithm [3] for AIMD calculations will be described. In section 6, the calculation of observable properties will be outlined and specific examples considered. In particular, we will focus on properties for which having direct access to the electronic structure, a unique feature of AIMD calculations, is important. These include IR and Raman spectra and NMR chemical shifts. In sections 7 and 8, we will present a number of applications of the AIMD methodology, including structure and dynamics of liquids and aqueous solutions, biological systems and industrial catalysis. Following this, in section 9, the inclusion of nuclear quantum effects via the Feynman path integral approach will be reviewed, and an AIMD based scheme for the calculation of path integrals will be detailed. Applications of the AIPI technique to intramolecular proton transfer and aqueous proton transport will be presented. Finally, in sections 10 and 11, we will briefly discuss a number of advanced AIMD topics including methods for incorporation of excited states and linear scaling approaches.

2. The Born–Oppenheimer approximation and *ab initio* molecular dynamics

We begin our discussion of AIMD by considering a system of N nuclei described by coordinates $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_N \equiv \boldsymbol{R}$, momenta $\boldsymbol{P}_1, \ldots, \boldsymbol{P}_N \equiv \boldsymbol{P}$ and masses $M_1, \ldots, M_N$, and N_e electrons described by coordinates $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{N_e} \equiv \boldsymbol{r}$, momenta $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_{N_e} \equiv \boldsymbol{p}$ and spin variables $s_1, \ldots, s_{N_e} \equiv s$. Nuclear spin is ignored in the present discussion. The non-relativistic Hamiltonian of the system is given by

$$H = \sum_{I=1}^{N} \frac{\boldsymbol{P}_I^2}{2M_I} + \sum_{i=1}^{N_e} \frac{\boldsymbol{p}_i^2}{2m} + \sum_{i>j} \frac{e^2}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} + \sum_{I>J} \frac{Z_I Z_J e^2}{|\boldsymbol{R}_I - \boldsymbol{R}_J|} - \sum_{i,I} \frac{Z_I e^2}{|\boldsymbol{R}_I - \boldsymbol{r}_i|}$$
$$\equiv T_{\mathrm{N}} + T_{\mathrm{e}} + V_{\mathrm{ee}}(\boldsymbol{r}) + V_{\mathrm{NN}}(\boldsymbol{R}) + V_{\mathrm{eN}}(\boldsymbol{r}, \boldsymbol{R}) \tag{2.1}$$

where m is the mass of the electron and $Z_I e$ is the charge on the Ith nucleus. In the second line, T_{N}, T_{e}, V_{ee}, V_{NN} and V_{eN} represent the nuclear and electron kinetic energy operators and electron–electron, nuclear–nuclear and electron–nuclear interaction potential operators, respectively. In order to solve the complete quantum mechanical problem, we start by seeking the eigenfunctions and eigenvalues of this Hamiltonian, which will be given by solution of the time-independent Schrödinger equation

$$[T_{\mathrm{N}} + T_{\mathrm{e}} + V_{\mathrm{ee}}(\boldsymbol{r}) + V_{\mathrm{NN}}(\boldsymbol{R}) + V_{\mathrm{eN}}(\boldsymbol{r}, \boldsymbol{R})]\Psi(\boldsymbol{x}, \boldsymbol{R}) = E\Psi(\boldsymbol{x}, \boldsymbol{R}) \tag{2.2}$$

where $\boldsymbol{x} \equiv (\boldsymbol{r}, s)$ denotes the full collection of electron position and spin variables, and $\Psi(\boldsymbol{x}, \boldsymbol{R})$ is an eigenfunction of H with eigenvalue E. Clearly, an exact solution of equation (2.2) is not possible and approximations must be made. We first invoke the Born–Oppenheimer approximation by recognizing that, in a dynamical sense, there is a strong separation of timescales between the electronic and nuclear motion, since the electrons are

lighter than the nuclei by three orders of magnitude. In terms of equation (2.2), this can be exploited by assuming a quasi-separable ansatz of the form

$$\Psi(\boldsymbol{x}, \boldsymbol{R}) = \phi(\boldsymbol{x}, \boldsymbol{R})\chi(\boldsymbol{R}) \tag{2.3}$$

where $\chi(\boldsymbol{R})$ is a nuclear wavefunction and $\phi(\boldsymbol{x}, \boldsymbol{R})$ is an electronic wavefunction that depends parametrically on the nuclear positions. We note, at this point, that an alternative derivation using a fully separable ansatz to the time-dependent Schrödinger equation was presented by Marx and Hutter in [10]. Substitution of equation (2.3) into (2.2) and recognition that the nuclear wavefunction $\chi(\boldsymbol{R})$ is more localized than the electronic wavefunction, i.e. $\nabla_I \chi(\boldsymbol{R}) \gg \nabla_I \phi(\boldsymbol{x}, \boldsymbol{R})$, yields

$$\frac{[T_\mathrm{e} + V_\mathrm{ee}(\boldsymbol{r}) + V_\mathrm{eN}(\boldsymbol{r}, \boldsymbol{R})]\phi(\boldsymbol{x}, \boldsymbol{R})}{\phi(\boldsymbol{x}, \boldsymbol{R})} = E - \frac{[T_\mathrm{N} + V_\mathrm{NN}(\boldsymbol{R})]\chi(\boldsymbol{R})}{\chi(\boldsymbol{R})}. \tag{2.4}$$

From the above, it is clear that the left-hand side can only be a function of $\boldsymbol{R}$ alone. Let this function be denoted $\varepsilon(\boldsymbol{R})$. Thus,

$$\begin{aligned} &\frac{[T_\mathrm{e} + V_\mathrm{ee}(\boldsymbol{r}) + V_\mathrm{eN}(\boldsymbol{r}, \boldsymbol{R})]\phi(\boldsymbol{x}, \boldsymbol{R})}{\phi(\boldsymbol{x}, \boldsymbol{R})} = \varepsilon(\boldsymbol{R}) \\ &[T_\mathrm{e} + V_\mathrm{ee}(\boldsymbol{r}) + V_\mathrm{eN}(\boldsymbol{r}, \boldsymbol{R})]\phi(\boldsymbol{x}, \boldsymbol{R}) = \varepsilon(\boldsymbol{R})\phi(\boldsymbol{x}, \boldsymbol{R}). \end{aligned} \tag{2.5}$$

Equation (2.5) is an electronic eigenvalue equation for an electronic Hamiltonian, $H_\mathrm{e}(\boldsymbol{R}) = T_\mathrm{e} + V_\mathrm{ee}(\boldsymbol{r}) + V_\mathrm{eN}(\boldsymbol{r}, \boldsymbol{R})$, which will yield a set of normalized eigenfunctions $\phi_n(\boldsymbol{x}, \boldsymbol{R})$ and eigenvalues $\varepsilon_n(\boldsymbol{R})$, which depend parametrically on the nuclear positions, $\boldsymbol{R}$. For each solution, there will be a nuclear eigenvalue equation:

$$[T_\mathrm{N} + V_\mathrm{NN}(\boldsymbol{R}) + \varepsilon_n(\boldsymbol{R})]\chi(\boldsymbol{R}) = E\chi(\boldsymbol{R}). \tag{2.6}$$

Moreover, each electronic eigenvalue, $\varepsilon_n(\boldsymbol{R})$, will give rise to an electronic surface on which the nuclear dynamics is determined by a time-dependent Schrödinger equation for the time-dependent nuclear wavefunction $X(\boldsymbol{R}, t)$:

$$[T_\mathrm{N} + V_\mathrm{NN}(\boldsymbol{R}) + \varepsilon_n(\boldsymbol{R})]X(\boldsymbol{R}, t) = \mathrm{i}\hbar\frac{\partial}{\partial t}X(\boldsymbol{R}, t). \tag{2.7}$$

The physical interpretation of equation (2.7) is that the electrons respond instantaneously to the nuclear motion; therefore, it is sufficient to obtain a set of instantaneous electronic eigenvalues and eigenfunctions at each nuclear configuration, $\boldsymbol{R}$ (hence the parametric dependence of $\phi_n(\boldsymbol{x}, \boldsymbol{R})$ and $\varepsilon_n(\boldsymbol{R})$ on $\boldsymbol{R}$). The eigenvalues, in turn, give a family of (uncoupled) potential surfaces on which the nuclear wavefunction can evolve. Of course, these surfaces can (and often do) become coupled by so-called non-adiabatic effects, contained in the terms that have been neglected in the above derivation.

In many cases, non-adiabatic effects can be neglected, and we may consider motion *only* on the ground electronic surface described by

$$\begin{aligned} &[T_\mathrm{e} + V_\mathrm{ee}(\boldsymbol{r}) + V_\mathrm{eN}(\boldsymbol{r}, \boldsymbol{R})]\phi_0(\boldsymbol{x}, \boldsymbol{R}) = \varepsilon_0(\boldsymbol{R})\phi_0(\boldsymbol{x}, \boldsymbol{R}) \\ &[T_\mathrm{N} + \varepsilon_0(\boldsymbol{R}) + V_\mathrm{NN}(\boldsymbol{R})]X(\boldsymbol{R}, t) = \mathrm{i}\hbar\frac{\partial}{\partial t}X(\boldsymbol{R}, t). \end{aligned} \tag{2.8}$$

Moreover, if nuclear quantum effects can be neglected, then we may arrive at the classical nuclear evolution by assuming $X(\boldsymbol{R}, t)$ is of the form

$$X(\boldsymbol{R}, t) = A(\boldsymbol{R}, t)\mathrm{e}^{\mathrm{i}S(\boldsymbol{R},t)/\hbar} \tag{2.9}$$

and neglecting all terms involving $\hbar$, which yields an approximate equation for $S(\boldsymbol{R}, t)$:

$$H_\mathrm{N}(\nabla_1 S, \ldots, \nabla_N S, \boldsymbol{R}_1, \ldots, \boldsymbol{R}_N) + \frac{\partial S}{\partial t} = 0. \tag{2.10}$$

This is just the classical Hamiltonian–Jacobi equation with

$$H_{\mathrm{N}}(\boldsymbol{P}_1, \ldots, \boldsymbol{P}_N, \boldsymbol{R}_1, \ldots, \boldsymbol{R}_N) = \sum_{I=1}^{N} \frac{\boldsymbol{P}_I^2}{2M_I} + V_{\mathrm{NN}}(\boldsymbol{R}) + \varepsilon_0(\boldsymbol{R}) \tag{2.11}$$

denoting the classical nuclear Hamiltonian. The Hamilton–Jacobi equation is equivalent to classical motion on the ground-state surface, $E_0(\boldsymbol{R}) = \varepsilon_0(\boldsymbol{R}) + V_{\mathrm{NN}}(\boldsymbol{R})$, given by

$$\begin{aligned} \dot{\boldsymbol{R}}_I &= \frac{\boldsymbol{P}_I}{M_I} \\ \dot{\boldsymbol{P}}_I &= -\nabla_I E_0(\boldsymbol{R}). \end{aligned} \tag{2.12}$$

Note that the force $-\nabla_I E_0(\boldsymbol{R})$ contains a term from the nuclear–nuclear repulsion and a term from the derivative of the electronic eigenvalue, $\varepsilon_0(\boldsymbol{R})$. Because of the Hellman–Feynman theorem, the latter can be expressed as

$$\nabla_I \varepsilon_0(\boldsymbol{R}) = \langle \phi_0(\boldsymbol{R}) | \nabla_I H_{\mathrm{e}}(\boldsymbol{R}) | \phi_0(\boldsymbol{R}) \rangle. \tag{2.13}$$

Equations (2.12) and (2.13) form the theoretical basis of the AIMD approach. The practical implementation of the AIMD method requires an algorithm for the numerical solution of equation (2.12) with forces obtained from equation (2.13) at each step of the calculation. Moreover, since an exact solution for the ground-state electronic wavefunction, $|\phi_0(\boldsymbol{R})\rangle$, and eigenvalue, $\varepsilon_0(\boldsymbol{R})$, are not available, in general, it is necessary to introduce an approximation scheme for obtaining these quantities. This is the topic of the next section.

3. Representation of the electronic structure

At this point, a simple form for $E_0(\boldsymbol{R})$ could be introduced, giving rise to a force field based approach. Such a form would necessarily be specific to a particular system and, therefore, not be transferable to other situations. If, on the other hand, one derives forces directly from very accurate electronic structure calculations, the computational overhead associated with the method will be enormous. It is clear, therefore, that the practical utility of the AIMD approach relies on a compromise between accuracy and efficiency of the electronic structure representation based on available computing resources. One approach that has proved particularly successful in this regard is density functional theory (DFT). DFT is based on the Hohenberg–Kohn theorem [102], which states that a one-to-one mapping exists between ground-state electronic densities and external potentials. The ground-state density, $n_0(\boldsymbol{r})$, is given in terms of the ground-state wavefunction by

$$n_0(\boldsymbol{r}) = \sum_{s, s_2, \ldots, s_{N_e}} \int \mathrm{d}\boldsymbol{r}_2 \cdots \mathrm{d}\boldsymbol{r}_{N_e} |\phi_0(\boldsymbol{r}, s, \boldsymbol{r}_2, s_2, \ldots, \boldsymbol{r}_{N_e}, s_{N_e})|^2. \tag{3.1}$$

(Here, $\boldsymbol{r}$ and s represent a single position and spin variable, respectively.) A consequence of the Hohenberg–Kohn theorem is that the exact ground-state energy, $\varepsilon_0(\boldsymbol{R})$, can be obtained by minimizing a certain functional, $\varepsilon[n]$, over all electronic densities $n(\boldsymbol{r})$ that can be associated with an antisymmetric ground-state wavefunction, $|\phi_0\rangle$, of a Hamiltonian H_{e} for some potential V_{eN} (the so-called v-representability condition) subject to the restriction that $\int \mathrm{d}\boldsymbol{r}\, n(\boldsymbol{r}) = N_{\mathrm{e}}$. The theorem can also be extended to so-called N-representable densities (obtained from any antisymmetric wavefunction) via the Levy prescription [95, 96]. The functional $\varepsilon[n]$ is given as a sum, $T[n] + W[n] + V[n]$, where $T[n]$ and $W[n]$ represent the kinetic energy and Coulomb repulsion energies, respectively, and $V[n] = \int \mathrm{d}\boldsymbol{r}\, V_{\mathrm{eN}}(\boldsymbol{r}) n(\boldsymbol{r})$. Although the functional $T[n] + W[n]$ is universal for all systems of N_{e} electrons, its form is not known. Thus, in order that DFT be of practical utility, Kohn and Sham (KS) introduced the idea of a

non-interacting reference system with a potential $V_{\rm KS}(\boldsymbol{r}, \boldsymbol{R})$ such that the ground-state energy and density of the non-interacting system equal those of the true interacting system [94]. Within the KS formulation of DFT, a set of $n_{\rm occ}$ orthonormal single-particle orbitals, $\psi_i(\boldsymbol{r})$, $i = 1, \ldots, n_{\rm occ}$, with occupation numbers f_i, where $\sum_{i=1}^{n_{\rm occ}} f_i = N_e$, is introduced. These are known as the KS orbitals. In terms of the KS orbitals, the density is given by

$$n(\boldsymbol{r}) = \sum_{i=1}^{n_{\rm occ}} f_i |\psi_i(\boldsymbol{r})|^2 \tag{3.2}$$

and the functional takes the form

$$\begin{aligned}\varepsilon[\{\psi_i\}] &= -\frac{\hbar^2}{2m}\sum_{i=1}^{n_{\rm occ}} f_i \langle\psi_i|\nabla^2|\psi_i\rangle + \frac{e^2}{2}\int \mathrm{d}\boldsymbol{r}\,\mathrm{d}\boldsymbol{r}'\,\frac{n(\boldsymbol{r})n(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|} + \varepsilon_{\rm xc}[n] + \int \mathrm{d}\boldsymbol{r}\, n(\boldsymbol{r}) V_{\rm eN}(\boldsymbol{r}, \boldsymbol{R}) \\ &\equiv T_{\rm nonint}[\{\psi\}] + J[n] + \varepsilon_{\rm xc}[n] + V[n]. \end{aligned} \tag{3.3}$$

The first term in the functional represents the quantum kinetic energy, the second is the direct Coulomb term from Hartree–Fock theory, the third term is the exact exchange–correlation energy, whose form is unknown, and the fourth term is the interaction of the electron density with the external potential due to the nuclei. Thus, the KS potential is given by

$$V_{\rm KS}(\boldsymbol{r}, \boldsymbol{R}) = \frac{e^2}{2}\int \mathrm{d}\boldsymbol{r}'\,\frac{n(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|} + \frac{\delta\varepsilon_{\rm xc}}{\delta n(\boldsymbol{r})} + V_{\rm eN}(\boldsymbol{r}, \boldsymbol{R}) \tag{3.4}$$

and the Hamiltonian of the non-interacting system is, therefore,

$$H_{\rm KS} = -\frac{\hbar^2}{2m}\nabla^2 + V_{\rm KS}(\boldsymbol{r}, \boldsymbol{R}). \tag{3.5}$$

The KS orbitals will be the solutions of a set of self-consistent equations known as the *Kohn–Sham equations* [95, 96]:

$$H_{\rm KS}\psi_i(\boldsymbol{r}) = \varepsilon_i\psi_i(\boldsymbol{r}) \tag{3.6}$$

where ε_i are the KS energies. Equation (3.7) constitutes a self-consistent problem because the KS orbitals are needed to compute the density, which is needed to specify the KS Hamiltonian. However, the latter must be specified in order to determine the orbitals and orbital energies.

The preceding discussion makes clear the fact that DFT is, in principle, an exact theory for the ground state of a system. However, because the exchange–correlation functional, defined to be $\varepsilon_{\rm xc}[n] = T[n] - T_{\rm nonint}[\{\psi\}] + W[n] - J[n]$, is unknown, in practice, approximations must be made. One of the most successful approximations is the so-called *local density approximation* (LDA), in which the functional is taken to be the spatial integral over a local function that depends only on the density:

$$\varepsilon_{\rm xc}[n] \approx \int \mathrm{d}\boldsymbol{r}\, f_{\rm LDA}(n(\boldsymbol{r})). \tag{3.7}$$

The LDA is physically motivated by the notion that the interaction between the electrons and the nuclei creates only weak inhomogeneities in the electron density. Therefore, the form of $f_{\rm LDA}$ is obtained by evaluating the exact expressions for the exchange and correlation energies of a homogeneous electron gas of uniform density n at the inhomogeneous density $n(\boldsymbol{r})$. The LDA has been successfully used in numerous applications of importance in solid state physics, including studies of semiconductors and metals. In many instances of importance in chemistry, however, the electron density possesses sufficient inhomogeneities that the LDA breaks down. This is particularly true in hydrogen-bonded systems, for example. In such cases, the LDA can be improved by adding an additional dependence on the lowest order gradients of the density:

$$\varepsilon_{\rm xc}[n] \approx \int \mathrm{d}\boldsymbol{r}\, f_{\rm GGA}(n(\boldsymbol{r}), |\nabla n(\boldsymbol{r})|, \nabla^2 n(\boldsymbol{r})) \tag{3.8}$$

which is known as the *generalized gradient approximation* (GGA). Among the most widely used GGAs are those of Becke [103], Lee and Parr [104], Perdew and Wang [105], Perdew *et al* [106] and Cohen and Handy [107–109]. Typically, these can be calibrated to reproduce some subset of the known properties satisfied by the exact exchange–correlation functional. GGAs such as these have been used successfully in nearly all of the application areas discussed in section 1. However, GGAs are also known to underestimate transition state barriers and cannot adequately treat dispersion. Attempts to incorporate dispersion interactions in an empirical way have recently been proposed [110]. In order to improve reaction barriers, new approximation schemes such as Becke's 1992 functional [111], which incorporates exact exchange, and the so-called *meta-GGA* functionals [112–115], which include an additional dependence on the electron kinetic energy density

$$\tau(\boldsymbol{r}) = \sum_{i=1}^{n_{\mathrm{occ}}} f_i |\nabla \psi_i(\boldsymbol{r})|^2 \tag{3.9}$$

have been proposed with reasonable success. However, the problem of designing accurate approximate exchange–correlation functionals remains one of the greatest challenges in DFT.

Finally, in order to overcome the limitations of DFT in the context of AIMD, it is, of course, possible to employ a more accurate electronic structure method, and approaches using full configuration-interaction representations have been proposed [97]. Typically, these have a higher computational overhead and, therefore, can only be used to study much smaller systems such as very small clusters. However, as computing platforms become more powerful and new algorithms are developed, it is conceivable that other electronic structure methods may be used more routinely in AIMD studies.

4. Basis set expansions

4.1. Plane-wave basis sets

In MD calculations, the most commonly employed boundary conditions are periodic boundary conditions, in which the system is replicated infinitely in space. This is clearly a natural choice for solids and is particularly convenient for liquids. In an infinite periodic system, the KS orbitals become Bloch functions of the form

$$\psi_{i\boldsymbol{k}}(\boldsymbol{r}) = \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}} u_{i\boldsymbol{k}}(\boldsymbol{r}) \tag{4.1}$$

where $\boldsymbol{k}$ is a vector in the first Brillouin zone and $u_{i\boldsymbol{k}}(\boldsymbol{r})$ is a periodic function. A natural basis set for expanding a periodic function is the Fourier or plane-wave basis set, in which $u_{i\boldsymbol{k}}(\boldsymbol{r})$ is expanded according to

$$u_{i\boldsymbol{k}}(\boldsymbol{r}) = \frac{1}{\sqrt{\Omega}} \sum_{\boldsymbol{g}} c_{i,\boldsymbol{g}}^{\boldsymbol{k}} \mathrm{e}^{\mathrm{i}\boldsymbol{g}\cdot\boldsymbol{r}} \tag{4.2}$$

where Ω is the volume of the cell, $\boldsymbol{g} = 2\pi \boldsymbol{h}^{-1}\hat{\boldsymbol{g}}$ is a reciprocal lattice vector, $\boldsymbol{h}$ is the cell matrix, whose columns are the cell vectors ($\Omega = \det(\boldsymbol{h})$), $\hat{\boldsymbol{g}}$ is a vector of integers and $\{c_{i,\boldsymbol{g}}^{\boldsymbol{k}}\}$ are the expansion coefficients. An advantage of plane waves is that the sums needed to go back and forth between reciprocal space and real space can be performed efficiently using fast Fourier transforms (FFTs). In general, the properties of a periodic system are only correctly described if a sufficient number of $\boldsymbol{k}$-vectors are sampled from the Brioullin zone. However, for the applications to be considered herein, which are largely concerned with nonmetallic

systems, it is generally sufficient to consider a single $\boldsymbol{k}$ point, ($\boldsymbol{k} = (0, 0, 0)$) known as the Γ-point, so that the plane wave expansion reduces to

$$\psi_i(\boldsymbol{r}) = \frac{1}{\sqrt{\Omega}} \sum_{\boldsymbol{g}} c_{i,\boldsymbol{g}} \mathrm{e}^{\mathrm{i}\boldsymbol{g}\cdot\boldsymbol{r}}. \tag{4.3}$$

At the Γ-point, the orbitals can always be chosen to be real functions. Therefore, the plane-wave expansion coefficients satisfy the following property:

$$c^*_{i,\boldsymbol{g}} = c_{i,-\boldsymbol{g}} \tag{4.4}$$

which requires keeping only half of the full set of plane-wave expansion coefficients. In actual applications, only plane waves up to a given cut-off, $\hbar^2|\boldsymbol{g}|^2/2m < E_{\mathrm{cut}}$, are kept. Similarly, the density $n(\boldsymbol{r})$ given by equation (3.2) can also be expanded in a plane-wave basis:

$$n(\boldsymbol{r}) = \frac{1}{\Omega} \sum_{\boldsymbol{g}} n_{\boldsymbol{g}} \mathrm{e}^{\mathrm{i}\boldsymbol{g}\cdot\boldsymbol{r}}. \tag{4.5}$$

However, since $n(\boldsymbol{r})$ is obtained as a square of the KS orbitals, the cut-off needed for this expansion is $4E_{\mathrm{cut}}$ for consistency with the orbital expansion.

Using equations (4.3) and (4.5) and the orthogonality of the plane waves, it is straightforward to compute the various energy terms. From here on, we will employ atomic units ($\hbar = 1$, $e = 1$ and $m = 1$). Thus, the kinetic energy can be easily shown to be

$$\varepsilon_{\mathrm{KE}} = -\tfrac{1}{2} \sum_i \int \mathrm{d}\boldsymbol{r}\, \psi_i^*(\boldsymbol{r}) \nabla^2 \psi_i(\boldsymbol{r}) = \tfrac{1}{2} \sum_i \sum_{\boldsymbol{g}} g^2 |c^i_{\boldsymbol{g}}|^2 \tag{4.6}$$

where $g = |\boldsymbol{g}|$. Similarly, the Hartree energy becomes

$$\varepsilon_{\mathrm{H}} = \frac{1}{2} \int \mathrm{d}\boldsymbol{r}\, \mathrm{d}\boldsymbol{r}' \frac{n(\boldsymbol{r}) n(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} = \frac{1}{\Omega} {\sum_{\boldsymbol{g}}}' \frac{4\pi}{g^2} |n_{\boldsymbol{g}}|^2 \tag{4.7}$$

where the summation excludes the $\boldsymbol{g} = (0, 0, 0)$ term.

The exchange and correlation energy, $\varepsilon_{\mathrm{xc}}[n]$ in the LDA or GGA, is evaluated on the real-space FFT grid so that it can be expressed as

$$\varepsilon_{\mathrm{xc}}[n] = \frac{\Omega}{N_{\mathrm{grid}}} \sum_{\boldsymbol{r}} f_{\mathrm{GGA}}(n(\boldsymbol{r}), |\nabla n(\boldsymbol{r})|, \nabla^2 n(\boldsymbol{r})) \tag{4.8}$$

where N_{grid} is the number of real-space grid points. As shown by White and Bird [116], the use of the grid eliminates the complexity of functional differentiation by allowing the contribution to the KS potential from $\varepsilon_{\mathrm{xc}}$ be computed from

$$\frac{\mathrm{d}\varepsilon_{\mathrm{xc}}}{\mathrm{d}n(\boldsymbol{r})} = \frac{\Omega}{N_{\mathrm{grid}}} \frac{\partial f_{\mathrm{GGA}}}{\partial n(\boldsymbol{r})} + \frac{\Omega}{N_{\mathrm{grid}}} \sum_{\boldsymbol{r}'} \left[\frac{\partial f_{\mathrm{GGA}}}{\partial |\nabla n(\boldsymbol{r}')|} \frac{\partial |\nabla n(\boldsymbol{r}')|}{\partial n(\boldsymbol{r})} + \frac{\partial f_{\mathrm{GGA}}}{\partial \nabla^2 n(\boldsymbol{r}')} \frac{\partial \nabla^2 n(\boldsymbol{r}')}{\partial n(\boldsymbol{r})} \right]. \tag{4.9}$$

The gradient and (if needed) the Laplacian of the density can be computed efficiently using FFTs:

$$\begin{aligned} \nabla n(\boldsymbol{r}) &= \sum_{\boldsymbol{g}} \mathrm{i}\boldsymbol{g} \mathrm{e}^{\mathrm{i}\boldsymbol{g}\cdot\boldsymbol{r}} \sum_{\boldsymbol{r}'} n(\boldsymbol{r}') \mathrm{e}^{-\mathrm{i}\boldsymbol{g}\cdot\boldsymbol{r}'} \\ \nabla^2 n(\boldsymbol{r}) &= -\sum_{\boldsymbol{g}} g^2 \mathrm{e}^{\mathrm{i}\boldsymbol{g}\cdot\boldsymbol{r}} \sum_{\boldsymbol{r}'} n(\boldsymbol{r}') \mathrm{e}^{-\mathrm{i}\boldsymbol{g}\cdot\boldsymbol{r}'}; \end{aligned} \tag{4.10}$$

equation (4.10) also shows how the derivatives needed in equation (4.9) can be easily computed using combinations of forward and inverse FFTs.

The external energy is made somewhat complicated by the fact that, in a plane-wave basis, very large basis sets are needed to treat the rapid spatial fluctuations of core electrons.

Therefore, core electrons are often replaced by atomic pseudopotentials [117–119] or treated via augmented plane-wave techniques [120]. Here, we shall discuss the former. In the atomic pseudopotential scheme, the nucleus plus the core electrons are treated in a frozen core type approximation as an 'ion' carrying only the valence charge. In order to make this approximation, the valence orbitals, which, in principle, must be orthogonal to the core orbitals, must see a different pseudopotential for each angular momentum component in the core, which means that the pseudopotential must generally be nonlocal. In order to see this, we consider a potential operator of the form

$$\hat{V}_{\text{pseud}} = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} v_l(r)|lm\rangle\langle lm| \tag{4.11}$$

where r is the distance from the ion, and $|lm\rangle\langle lm|$ is a projection operator onto each angular momentum component. In order to truncate the infinite sum over l in equation (4.11), we assume that for some $l \geqslant \bar{l}$, $v_l(r) = v_{\bar{l}}(r)$ and add and subtract the function $v_{\bar{l}}(r)$ in equation (4.11):

$$\begin{aligned}\hat{V}_{\text{pseud}} &= \sum_{l=0}^{\infty} \sum_{m=-l}^{l} (v_l(r) - v_{\bar{l}}(r))|lm\rangle\langle lm| + v_{\bar{l}}(r) \sum_{l=0}^{\infty} \sum_{m=-l}^{l} |lm\rangle\langle lm| \\ &= \sum_{l=0}^{\infty} \sum_{m=-l}^{l} (v_l(r) - v_{\bar{l}}(r))|lm\rangle\langle lm| + v_{\bar{l}}(r) \approx \sum_{l=0}^{\bar{l}-1} \sum_{m=-l}^{l} \Delta v_l(r)|lm\rangle\langle lm| + v_{\bar{l}}(r)\end{aligned} \tag{4.12}$$

where the second line follows from the fact that the sum of the projection operators is unity, $\Delta v_l(r) = v_l(r) - v_{\bar{l}}(r)$ and the sum in the third line is truncated before $\Delta v_l(r) = 0$. The complete pseudopotential operator will be

$$\hat{V}_{\text{pseud}}(r; \boldsymbol{R}_1, \ldots, \boldsymbol{R}_N) = \sum_{I=1}^{N} \left[v_{\text{loc}}(|\boldsymbol{r} - \boldsymbol{R}_I|) + \sum_{l=0}^{\bar{l}-1} \Delta v_l(|\boldsymbol{r} - \boldsymbol{R}_I|)|lm\rangle\langle lm| \right] \tag{4.13}$$

where $v_{\text{loc}}(r) \equiv v_{\bar{l}}(r)$ is known as the local part of the pseudopotential (having no projection operator attached to it). Now, the external energy, being derived from the ground-state expectation value of a one-body operator, will be given by

$$\varepsilon_{\text{ext}} = \sum_i f_i \langle \psi_i | \hat{V}_{\text{pseud}} | \psi_i \rangle. \tag{4.14}$$

The first (local) term gives simply a local energy of the form

$$\varepsilon_{\text{loc}} = \sum_{I=1}^{N} \int \mathrm{d}\boldsymbol{r}\, n(\boldsymbol{r}) v_{\text{loc}}(|\boldsymbol{r} - \boldsymbol{R}_I|) \tag{4.15}$$

which can be evaluated in reciprocal space as

$$\varepsilon_{\text{loc}} = \frac{1}{\Omega} \sum_{I=1}^{N} \sum_{\boldsymbol{g}} n_{\boldsymbol{g}}^{*} \tilde{v}_{\text{loc}}(\boldsymbol{g}) \mathrm{e}^{-\mathrm{i}\boldsymbol{g}\cdot\boldsymbol{R}_I} \tag{4.16}$$

where $\tilde{V}_{\text{loc}}(\boldsymbol{g})$ is the Fourier transform of the local potential. Note that at $\boldsymbol{g} = (0, 0, 0)$, only the nonsingular part of $\tilde{v}_{\text{loc}}(\boldsymbol{g})$ contributes. In the evaluation of the local energy, it is often convenient to add and subtract a long range term of the form $Z_I \mathrm{erf}(\alpha_I r)/r$, where $\mathrm{erf}(x)$ is the error function, for each ion in order to obtain the nonsingular part explicitly and a residual short range function $\bar{v}_{\text{loc}}(|\boldsymbol{r} - \boldsymbol{R}_I|) = v_{\text{loc}}(|\boldsymbol{r} - \boldsymbol{R}_I|) - Z_I \mathrm{erf}(\alpha_I |\boldsymbol{r} - \boldsymbol{R}_I|)/|\boldsymbol{r} - \boldsymbol{R}_I|$ for each ionic core. For the nonlocal contribution, equation (4.3) is substituted into equation (4.13), and an

expansion of the plane waves in terms of spherical Bessel functions and spherical harmonics is made. After some algebra, one obtains

$$\varepsilon_{\mathrm{NL}} = \sum_i f_i \sum_I \sum_{\boldsymbol{g},\boldsymbol{g}'} \mathrm{e}^{-\mathrm{i}\boldsymbol{g}\cdot\boldsymbol{R}_I} c^*_{i,\boldsymbol{g}} v_{\mathrm{NL}}(\boldsymbol{g},\boldsymbol{g}') c_{i,\boldsymbol{g}'} \mathrm{e}^{\mathrm{i}\boldsymbol{g}'\cdot\boldsymbol{R}_I} \tag{4.17}$$

where

$$v_{\mathrm{NL}}(\boldsymbol{g},\boldsymbol{g}') = (4\pi)^2 \sum_{l=0}^{\bar{l}-1} \sum_{m=-l}^{l} \int \mathrm{d}r\, r^2 j_l(gr) j_l(g'r) \Delta v_l(r) Y_{lm}(\theta_g,\phi_g) Y^*_{lm}(\theta_{g'},\phi_{g'}) \tag{4.18}$$

and θ_g ($\theta_{g'}$) and ϕ_g ($\phi_{g'}$) are the spherical polar angles associated with the vector $\boldsymbol{g}$ ($\boldsymbol{g}'$), Y_{lm} are the spherical harmonics and $j_l(x)$ is a spherical Bessel function. Equation (4.18), which is known as the *semi-local* form, shows that the evaluation of the nonlocal energy can be quite computationally expensive. It also shows, however, that the matrix element is *almost* separable into $\boldsymbol{g}$- and $\boldsymbol{g}'$-dependent terms. A fully separable approximation can be obtained by writing

$$\begin{aligned} v_{\mathrm{NL}}(\boldsymbol{g},\boldsymbol{g}') = (4\pi)^2 \sum_{l=0}^{\bar{l}-1} \sum_{m=-l}^{l} \int \mathrm{d}r\, r^2 \int \mathrm{d}r'\, r'^2 j_l(gr) j_l(g'r') \Delta v_l(r) \frac{\delta(r-r')}{rr'} \\ \times Y_{lm}(\theta_g,\phi_g) Y^*_{lm}(\theta_{g'},\phi_{g'}) \end{aligned} \tag{4.19}$$

where a radial δ-function has been introduced. Next, the δ-function is expanded in terms of a set of radial eigenfunctions (usually taken to be those of the Hamiltonian from which the pseudopotential is obtained) for each angular momentum channel

$$\frac{\delta(r-r')}{rr'} = \sum_{n=0}^{\infty} \phi^*_{nl}(r) \phi_{nl}(r'). \tag{4.20}$$

If this expansion is now substituted into equation (4.19), the result is

$$\begin{aligned} v_{\mathrm{NL}}(\boldsymbol{g},\boldsymbol{g}') = (4\pi)^2 \sum_{n=0}^{\infty} \sum_{l=0}^{\bar{l}-1} \sum_{m=-l}^{l} \left[\int \mathrm{d}r\, r^2 j_l(gr) \Delta v_l(r) \phi^*_{nl}(r) Y_{lm}(\theta_g,\phi_g) \right] \\ \times \left[\int \mathrm{d}r'\, r'^2 j_l(g'r') Y^*_{lm}(\theta_{g'},\phi_{g'}) \phi_{nl}(r') \right] \end{aligned} \tag{4.21}$$

which is now fully separable at the expense of another infinite sum that needs to be truncated. The sum over n can be truncated after a finite number of terms, although some care is required in performing the truncation; the so-called *Kleinman–Bylander approximation* [121] is the result of truncating it at just a single term. The result of this truncation can be shown to yield the approximate form

$$\begin{aligned} v_{\mathrm{NL}}(\boldsymbol{g},\boldsymbol{g}') \approx (4\pi)^2 \sum_{l=0}^{\bar{l}-1} \sum_{m=-l}^{l} N^{-1}_{lm} \left[\int \mathrm{d}r\, r^2 j_l(gr) \Delta v_l(r) \phi^*_l(r) Y_{lm}(\theta_g,\phi_g) \right] \\ \times \left[\int \mathrm{d}r'\, r'^2 \Delta v_l(r') j_l(g'r') Y^*_{lm}(\theta_{g'},\phi_{g'}) \phi_l(r') \right] \end{aligned} \tag{4.22}$$

where

$$N_{lm} = \int \mathrm{d}r\, r^2 \phi^*_l(r) \Delta v_l(r) \phi_l(r) \tag{4.23}$$

and $\phi_l(r) \equiv \phi_{0l}(r)$. Finally, substituting equation (4.22) into (4.17) gives the nonlocal energy as

$$\varepsilon_{\mathrm{NL}} = \sum_{i=1}^{N_e} \sum_{I=1}^{N} \sum_{l=0}^{\bar{l}-1} \sum_{m=-l}^{l} Z^*_{iIlm} Z_{iIlm} \tag{4.24}$$

where

$$Z_{iIlm} = \sum_{\boldsymbol{g}} c_{\boldsymbol{g}}^{i} \mathrm{e}^{\mathrm{i}\boldsymbol{g}\cdot\boldsymbol{R}_I} \tilde{F}_{lm}(\boldsymbol{g}) \tag{4.25}$$

and

$$\tilde{F}_{lm}(\boldsymbol{g}) = 4\pi N_{lm}^{-1/2} \int \mathrm{d}r\, r^2 j_l(gr) \Delta v_l(r) \phi_l(r) Y_{lm}(\theta_{\boldsymbol{g}}, \phi_{\boldsymbol{g}}). \tag{4.26}$$

For certain elements, it has been shown that the simple Kleinman–Bylander form can lead to spurious or unphysical bound states known as *ghost levels*. [122] and [123] contain analyses and techniques for treating spurious ghost states. Alternatively, ghosts can be eliminated by taking more terms than just the first in equation (4.21) [124] or working directly with the semilocal form.

The last issue we shall discuss is that of boundary conditions within the plane-wave description. Plane waves naturally describe fully periodic systems, such as solids, or systems that can be effectively treated with periodic boundary conditions, such as liquids. What if we wish to study a system, such as a cluster, surface or wire, in which one or more boundaries is *not* periodic? It turns out that such situations can be described rigorously within the plane-wave formalism. One approach is based on a direct solution to the Poisson equation in a box containing the cluster [125, 126]. Here, we shall discuss a simpler and more direct approach developed by Martyna and Tuckerman [127–129], which involves the use of a screening function in the long range energy terms, i.e. the Hartree and local pseudopotential terms. The idea is to use the so-called first image form of the average energy in order to form an approximation to a cluster, wire or surface system, whose error can be controlled by the dimensions of the simulation cell. Thus, given any density, $n(\boldsymbol{r})$, and any interaction potential, $\phi(\boldsymbol{r} - \boldsymbol{r}')$, the average potential energy in this approximation is given by

$$\langle \phi \rangle^{(1)} = \frac{1}{2\Omega} \sum_{\boldsymbol{g}} |n_{\boldsymbol{g}}|^2 \bar{\phi}(-\boldsymbol{g}) \tag{4.27}$$

where $\bar{\phi}(\boldsymbol{g})$ is a Fourier expansion coefficient of the potential given by

$$\bar{\phi}(\boldsymbol{g}) = \begin{cases} \displaystyle\int_{-L_c/2}^{L_c/2} \mathrm{d}z \int_{-L_b/2}^{L_b/2} \mathrm{d}y \int_{-L_a/2}^{L_a/2} \mathrm{d}x\, \phi(\boldsymbol{r}) \mathrm{e}^{-\mathrm{i}\boldsymbol{g}\cdot\boldsymbol{r}} & \text{(cluster)} \\ \displaystyle\int_{-L_c/2}^{L_c/2} \mathrm{d}z \int_{-L_b/2}^{L_b/2} \mathrm{d}y \int_{-\infty}^{\infty} \mathrm{d}x\, \phi(\boldsymbol{r}) \mathrm{e}^{-\mathrm{i}\boldsymbol{g}\cdot\boldsymbol{r}} & \text{(wire)} \\ \displaystyle\int_{-L_c/2}^{L_c/2} \mathrm{d}z \int_{-\infty}^{\infty} \mathrm{d}y \int_{-\infty}^{\infty} \mathrm{d}x\, \phi(\boldsymbol{r}) \mathrm{e}^{-\mathrm{i}\boldsymbol{g}\cdot\boldsymbol{r}} & \text{(surface).} \end{cases} \tag{4.28}$$

Here, L_a, L_b and L_c are the dimensions of the simulation cell (assumed to be orthorhombic for simplicity) in the x, y and z directions ($\boldsymbol{h} = \mathrm{diag}(L_a, L_b, L_c)$). Note that, unlike equation (4.7), the $\boldsymbol{g} = (0, 0, 0)$ term is not excluded. In order to obtain an expression that is easily computed within the plane-wave description, consider two functions $\phi^{(\mathrm{long})}(\boldsymbol{r})$ and $\phi^{(\mathrm{short})}(\boldsymbol{r})$, which are assumed to be the long and short range contributions to the total potential, i.e.

$$\begin{aligned} \phi(\boldsymbol{r}) &= \phi^{(\mathrm{long})}(\boldsymbol{r}) + \phi^{(\mathrm{short})}(\boldsymbol{r}) \\ \bar{\phi}(\boldsymbol{g}) &= \bar{\phi}^{(\mathrm{long})}(\boldsymbol{g}) + \bar{\phi}^{(\mathrm{short})}(\boldsymbol{g}). \end{aligned} \tag{4.29}$$

We require that $\phi^{(\mathrm{short})}(\boldsymbol{r})$ vanish exponentially quickly at large distances from the centre of the parallelepiped and that $\phi^{(\mathrm{long})}(\boldsymbol{r})$ contain the long range dependence of the full potential,

$\phi(\boldsymbol{r})$. With these two requirements, it is possible to write

$$\bar{\phi}^{(\text{short})}(\boldsymbol{g}) = \int_{D(\Omega)} \mathrm{d}\boldsymbol{r}\ \exp(-\mathrm{i}\boldsymbol{g}\cdot\boldsymbol{r})\phi^{(\text{short})}(\boldsymbol{r}) = \int_{\text{all space}} \mathrm{d}\boldsymbol{r}\ \exp(-\mathrm{i}\boldsymbol{g}\cdot\boldsymbol{r})\phi^{(\text{short})}(\boldsymbol{r}) + \epsilon(\boldsymbol{g}) = \tilde{\phi}^{(\text{short})}(\boldsymbol{g}) + \epsilon(\boldsymbol{g}) \tag{4.30}$$

with exponentially small error, $\epsilon(\boldsymbol{g})$, provided the range of $\phi^{(\text{short})}(\boldsymbol{r})$ is small compared with the size of the parallelepiped. In order to ensure that equation (4.30) is satisfied, a convergence parameter, α, is introduced which can be used to adjust the range of $\phi^{(\text{short})}(\boldsymbol{r})$ such that $\epsilon(\boldsymbol{g}) \sim 0$ and the error, $\epsilon(\boldsymbol{g})$, will be neglected in the following.

The function, $\tilde{\phi}^{(\text{short})}(\boldsymbol{g})$, is the Fourier transform of $\phi^{(\text{short})}(\boldsymbol{r})$. Therefore,

$$\bar{\phi}(\boldsymbol{g}) = \bar{\phi}^{(\text{long})}(\boldsymbol{g}) + \tilde{\phi}^{(\text{short})}(\boldsymbol{g}) = \bar{\phi}^{(\text{long})}(\boldsymbol{g}) - \tilde{\phi}^{(\text{long})}(\boldsymbol{g}) + \tilde{\phi}^{(\text{short})}(\boldsymbol{g}) + \tilde{\phi}^{(\text{long})}(\boldsymbol{g}) = \hat{\phi}^{(\text{screen})}(\boldsymbol{g}) + \tilde{\phi}(\boldsymbol{g}) \tag{4.31}$$

where $\tilde{\phi}(\boldsymbol{g}) = \tilde{\phi}^{(\text{short})}(\boldsymbol{g}) + \tilde{\phi}^{(\text{long})}(\boldsymbol{g})$ is the Fourier transform of the full potential, $\phi(\boldsymbol{r}) = \phi^{(\text{short})}(\boldsymbol{r}) + \phi^{(\text{long})}(\boldsymbol{r})$ and

$$\hat{\phi}^{(\text{screen})}(\boldsymbol{g}) = \bar{\phi}^{(\text{long})}(\boldsymbol{g}) - \tilde{\phi}^{(\text{long})}(\boldsymbol{g}). \tag{4.32}$$

Thus, equation (4.32) leads to

$$\langle\phi\rangle = \frac{1}{2\Omega}\sum_{\hat{g}} |\bar{n}(\boldsymbol{g})|^2[\tilde{\phi}(-\boldsymbol{g}) + \hat{\phi}^{(\text{screen})}(-\boldsymbol{g})]. \tag{4.33}$$

The new function appearing in the average potential energy, equation (4.33), is the difference between the Fourier series and Fourier transform of the long range part of the potential energy and will be referred to as the screening function because it is constructed to 'screen' the interaction of the system with an infinite array of periodic images.

The specific case of the Coulomb potential,

$$\phi(\boldsymbol{r}) = \frac{1}{r} \tag{4.34}$$

can be separated into short and long range components via

$$\frac{1}{r} = \frac{\operatorname{erf}(\alpha r)}{r} + \frac{\operatorname{erfc}(\alpha r)}{r} \tag{4.35}$$

where the first term is long range. The screening function for the cluster case is easily computed by introducing an FFT grid and performing the integration numerically [127]. For the wire [129] and surface [128] cases, analytical expressions can be worked out and are given by

$$\begin{aligned}\bar{\phi}^{(\text{screen})}(\boldsymbol{g}) = -\frac{4\pi}{g^2}\Bigg\{&\cos\left(\frac{g_c L_c}{2}\right)\Bigg[\exp\left(-\frac{g_s L_c}{2}\right) - \frac{1}{2}\exp\left(-\frac{g_s L_c}{2}\right)\operatorname{erfc}\left(\frac{\alpha^2 L_c - g_s}{2\alpha}\right) \\ &- \frac{1}{2}\exp\left(\frac{g_s L_c}{2}\right)\operatorname{erfc}\left(\frac{\alpha^2 L_c + g_s}{2\alpha}\right)\Bigg] \\ &+ \exp\left(-\frac{g^2}{4\alpha^2}\right)\operatorname{Re}\left[\operatorname{erfc}\left(\frac{\alpha^2 L_c + \mathrm{i}g_c}{2\alpha}\right)\right]\Bigg\} \qquad \text{(surface)}\end{aligned} \tag{4.36}$$

$$\begin{aligned}\bar{\phi}^{(\text{screen})}(\boldsymbol{g}) = \frac{4\pi}{g^2}\Bigg[&\exp\left(-\frac{g^2}{4\alpha^2}\right)E(\alpha, L_b, g_b)E(\alpha, L_c, g_c) \\ &+ \cos\left(\frac{g_b L_b}{2}\right)\frac{4\sqrt{\pi}}{\alpha L_b}\exp\left(-\frac{g_c^2}{4\alpha^2}\right)I(\alpha, L_b, L_c, g_c)\end{aligned}$$

$$+\ \cos\left(\frac{g_c L_c}{2}\right)\frac{4\sqrt{\pi}}{\alpha L_c}\exp\left(-\frac{g_b^2}{4\alpha^2}\right)I(\alpha, L_c, L_b, g_b)\Bigg] - \frac{4\pi}{g^2}\mathrm{e}^{-g^2/4\alpha^2}$$

(wire) (4.37)

where

$$I(\alpha, L_1, L_2, g) = \int_0^{\alpha L_1/2} \mathrm{d}x\, x\mathrm{e}^{-g_a^2 L_1^2/16x^2}\mathrm{e}^{-x^2} E\left(\frac{2x}{L_1}, L_2, g\right) \tag{4.38}$$

and

$$E(\lambda, L, g) = \mathrm{erf}\left(\frac{\lambda^2 L + \mathrm{i}g}{2\lambda}\right) \tag{4.39}$$

where $\boldsymbol{g} = (g_a, g_b, g_c)$ and $g_s = \sqrt{g_a^2 + g_b^2}$. The one-dimensional integrals in equation (4.38) are well suited to be performed by Gaussian quadrature techniques. It should be noted that a simplified expression for the surface screening function can be obtained in the limit $\alpha \to \infty$ [10, 128, 130]:

$$\bar{\phi}^{(\mathrm{screen})}(\boldsymbol{g}) \longrightarrow -\frac{4\pi}{g^2}\left[\cos\left(\frac{g_c L_c}{2}\right)\mathrm{e}^{-g_s/L_c/2}\right]. \tag{4.40}$$

However, as is discussed in [128], some care is needed for $g_s = 0$.

Similarly, for the wire screening function, a simplified expression can be obtained in the limit $\alpha \to \infty$ [129]:

$$\bar{\phi}^{(\mathrm{screen})} \longrightarrow \frac{16}{g^2}\left[\cos\left(\frac{g_b L_b}{2}\right)\frac{J(g_c, g_a, L_c, L_b)}{L_b} + \cos\left(\frac{g_c L_c}{2}\right)\frac{J(g_b, g_a, L_b, L_c)}{L_c}\right] - \frac{4\pi}{g^2} \tag{4.41}$$

where

$$J(g_1, g_2, L_1, L_2) = \int_0^{L_1/2} \mathrm{d}x\, \mathrm{e}^{\mathrm{i}g_1 x}\sqrt{\theta(x, g_2, L_2)}K_1\left(\sqrt{4\theta(x, g_2, L_2)}\right) \tag{4.42}$$

$$\theta(x, g, L) = \frac{g^2 L^2/16}{1 + 4x^2/L^2} \tag{4.43}$$

and $K_1(z)$ is a modified Bessel function.

4.2. Gaussian basis sets

There is a great advantage to be gained by the use of localized basis sets over a delocalized basis like plane waves. In particular, the computations scale better for well localized orbitals. One of the most widely used localized basis sets is the Gaussian basis. In a Gaussian basis, the KS orbitals are expanded according to

$$\psi_i(\boldsymbol{r}) = \sum_{\alpha,\beta,\gamma} C^i_{\alpha\beta\gamma} G_{\alpha\beta\gamma}(\boldsymbol{r}; \boldsymbol{R}) \tag{4.44}$$

where the basis functions, $G_{\alpha\beta\gamma}(\boldsymbol{r}; \boldsymbol{R})$, are centred on atoms and, therefore, are dependent on the positions of the atoms. The basis functions generally take the form

$$G_{\alpha\beta\gamma}(\boldsymbol{r}; \boldsymbol{R}_I) = \mathcal{N}_{\alpha\beta\gamma} x^\alpha y^\beta z^\gamma \exp[-|\boldsymbol{r} - \boldsymbol{R}_I|^2/2\sigma^2_{\alpha\beta\gamma}] \tag{4.45}$$

with integer α, β and γ for the Gaussian centred on atom I. The advantage of Gaussians is that many of the integrals appearing in the DFT functional (and in other electronic structure methods) can be done analytically. The main disadvantage, however, is that Gaussian basis sets are nonorthogonal and, hence, the overlap matrix between basis functions needs to be

included in the various energy terms. Moreover, being dependent on atomic positions, the derivatives of the basis functions with respect to positions need to be computed. This leads to a considerable degree of complication for MD. Finally, with Gaussian bases, it is difficult to reproduce the correct asymptotic behaviour of the density, and one must always be aware of the effects of basis set superposition errors. Many of these problems can be eliminated by choosing to work with a simpler localized orthonormal basis set to be discussed in the next subsection.

4.3. Discrete variable representations

While plane-wave basis sets have the advantage of simplicity and lack the spatial bias inherent in Gaussian basis sets, they lead to $\mathcal{O}(N^2M)$ scaling for DFT calculations, where N is the number of electronic states and M is the number of basis functions. Moreover, because they are spatially delocalized, they are not optimal for use on emerging massively parallel computing architectures because of their high communication overhead. As noted above, the problem of delocalization is largely eliminated with Gaussian basis sets at the expense of considerably increased complexity due to the nonorthogonality of the basis functions and the introduction of basis set superposition error. A significant advantage might be gained if the localized character of Gaussian basis sets could be achieved using simple, orthonormal basis functions that are not centred on atomic positions, thereby avoiding the complexity of Pulay forces and basis set super position error. In fact, such a basis set is possible in the form of the discrete variable representation (DVR) [131–135], and implementation of AIMD with DVR basis sets was recently introduced [101]. A one-dimensional DVR is composed of a set of N functions, $u_i(x)$, $i = 1, \ldots, N$, and a set of N grid points, x_i, such that the basis functions satisfy a Kroenecker δ property:

$$u_i(x_j) = \frac{\delta_{ij}}{a_i} \tag{4.46}$$

on the grid points. Here, a_i is a (generally) complex number such that $w_i = |a_i|^2$ defines a set of quadrature weights. Therefore, the DVR functions behave as coordinate eigenfunctions on a particular quadrature grid (see figure 1). In this sense, they are the real-space analogues of plane waves, which are momentum eigenfunctions. Moreover, they can be constructed in such a way that each DVR function is well localized about a point on the grid. DVRs are commonly used in accurate bound-state and scattering calculations [131–135] and have been recently adapted for use in DFT based electronic structure calculations [101]. The basis functions then satisfy orthogonality and completeness relations of the form

$$\begin{aligned} \sum_{k=1}^{N} w_k u_i^*(x_k) u_j(x_k) &= \delta_{ij} \\ \sum_{k=1}^{N} w_k u_k^*(x_i) u_k(x_j) &= \delta_{ij}. \end{aligned} \tag{4.47}$$

An important property of a DVR is that any position-dependent operator $\Omega(x)$ is an approximately diagonal DVR basis set. By defining a projection operator, $\mathcal{P}_N = \sum_{i=1}^{N} |u_i\rangle\langle u_j|$, it can be shown that

$$\begin{aligned} &\langle u_i|\mathcal{P}_N \Omega(x) \mathcal{P}_N|u_j\rangle = \Omega(x_i)\delta_{ij} \\ &\langle u_i|\Omega(x)|u_j\rangle \approx \Omega(x_i)\delta_{ij} \\ &\lim_{N\to\infty} \langle u_i|\Omega(x)|u_j\rangle = \Omega(x_i)\delta_{ij} \end{aligned} \tag{4.48}$$

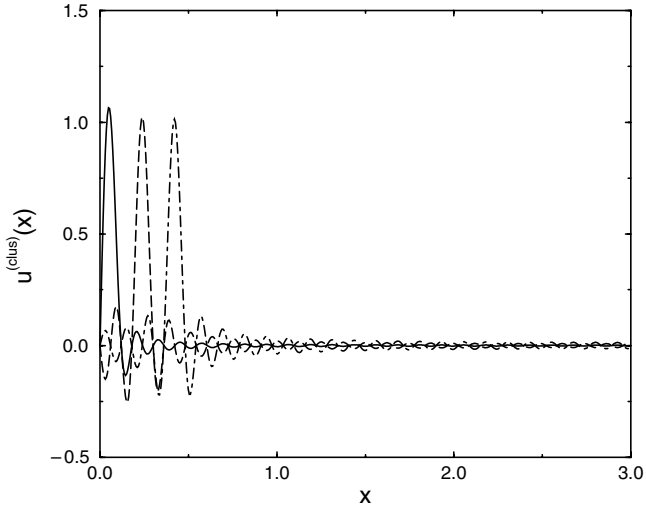

Figure 1. Example DVR basis functions generated using equation (4.50) showing the localization of the basis functions at grid points.

so that as the basis set size approaches infinity, position-dependent operators become exactly diagonal.

DVR functions that satisfy equation (4.46) can be constructed from simpler basis functions according to the boundary conditions of the problem. For example, a DVR appropriate for periodic boundary conditions on a one-dimensional grid of N points can be constructed from a sum of cosine functions according to

$$u_l^{(\mathrm{per})}(x) = \sqrt{\frac{1}{NL}} \sum_{\lambda=1}^{N} \cos[k_\lambda (x - x_l)] \tag{4.49}$$

where $k_\lambda = 2\pi(\lambda - N' - 1)/L$, $\lambda, l = 1, \ldots, N$, and $N' = (N-1)/2$, while a fixed-node DVR appropriate for a one-dimensional cluster system can be constructed from sine functions according to

$$u_l^{(\mathrm{clus})}(x) = \frac{2}{\sqrt{(N+1)L}} \sum_{\lambda=1}^{N} \sin k_\lambda x \sin k_\lambda x_l \tag{4.50}$$

where $k_\lambda = \pi\lambda/L$, $\lambda = 1, \ldots, N$. Finally, a useful basis set for expansion of the KS orbitals can be constructed from a direct product of these one-dimensional DVRs according to

$$\Phi_{lmn}(\boldsymbol{r}) = u_l^{(a)}(x) v_m^{(b)}(y) q_n^{(c)}(z) \tag{4.51}$$

where a, b and c indicate the type of function used, i.e. periodic (per) or cluster (clus). Thus, the basis functions can be tailored for different types of boundary condition, e.g. a periodic DVR in one dimension and cluster DVRs in the other two could yield a basis set appropriate for surface calculations etc. Once a direct product basis set is chosen, the KS orbitals are expanded according to

$$\psi_i(\boldsymbol{r}) = \sum_{l,m,n} C_{lmn}^i \Phi_{lmn}(\boldsymbol{r}) \tag{4.52}$$

where C_{lmn}^i is a set of expansion coefficients.

Specific terms in the energy functional follow directly from the form of the basis functions. Thus, the kinetic energy can be expressed analytically as

$$\varepsilon_{\mathrm{KE}} = \sum_{i} \sum_{l,l'} \sum_{m,m'} \sum_{n,n'} C^{i*}_{lmn} T_{ll',mm',nn'} C^{l}_{l'm'n'} \tag{4.53}$$

where the kinetic energy matrix has the highly sparse form

$$T_{ll',mm',nn'} = t^{(a)}_{ll'} \delta_{mm'} \delta_{nn'} + \delta_{ll'} t^{(b)}_{mm'} \delta_{nn'} + \delta_{ll'} \delta_{mm'} t^{(c)}_{nn'}. \tag{4.54}$$

The one-dimensional matrices are given by

$$t^{(\mathrm{per})}_{nn'} = \begin{cases} -\left(\dfrac{2\pi}{L}\right)^2 \dfrac{N'}{3}(N'+1) & n = n' \\ \dfrac{-\left(\frac{2\pi}{L}\right)^2 (-1)^{n-n'} \cos\left[\frac{\pi(n-n')}{N}\right]}{2\sin^2\left[\frac{\pi(n-n')}{N}\right]} & n \neq n' \end{cases} \tag{4.55}$$

for periodic and

$$t^{(\mathrm{clus})}_{nn'} = \begin{cases} -\dfrac{1}{L^2}\dfrac{\pi^2}{2}\left[\dfrac{2(N+1)^2+1}{3} - \dfrac{1}{\sin^2\left(\frac{\pi n}{N+1}\right)}\right] & n = n' \\ -\dfrac{(-1)^{n-n'}}{L^2}\dfrac{\pi^2}{2}\left[\dfrac{1}{\sin^2\left(\frac{\pi(n-n')}{2(N+1)}\right)} - \dfrac{1}{\sin^2\left(\frac{\pi(n+n')}{2(N+1)}\right)}\right] & n \neq n' \end{cases} \tag{4.56}$$

for fixed-node DVRs, respectively. Terms in the functional involving the electron density can be computed by substituting equation (4.52) into (3.2):

$$n(\boldsymbol{r}) = \sum_{i=1}^{n_{\mathrm{occ}}} \left[\sum_{l,m,n} C^{i}_{lmn} \Phi_{lmn}(\boldsymbol{r})\right]^2. \tag{4.57}$$

The resulting expression simplifies considerably when the density is subsequently evaluated at the points of the DVR grid, and the latter are all that is required for computing energies and forces. The gradient of the density, which is needed for GGA functionals, is computed by differentiating equation (4.57) and evaluating the result at the DVR grid points. Resulting derivatives of the DVR functions are computed once at the beginning and stored on one-dimensional grids. A simplifying approximation to this approach can also be introduced by postulating a similar DVR expansion for the density

$$n(\boldsymbol{r}) = \sum_{l,m,n,} N_{lmn} \Phi_{lmn}(\boldsymbol{r}) = \sum_{l,m,n} N_{lmn} u_l(x) v_m(y) q_n(z). \tag{4.58}$$

This allows the gradient of the density to be computed according to

$$\nabla n(x_l, y_m, z_n) = \sum_{l'} N_{l'mn} u'_{l'}(x_l) + \sum_{m'} N_{lm'n} v'_{m'}(y_m) + \sum_{n'} N_{lmn'} q'_{n'}(z_n). \tag{4.59}$$

Here, again, the derivatives of the DVR functions can be computed once at the beginning of a simulation and the values stored on one-dimensional grids. Our (as yet limited) experience, thus far, is that the approximation in equation (4.59) tends to converge the total energy with the same number of grid points as the exact method; however, the errors in the exact method are smaller for DVR grids with fewer points. The nonlocal part of the pseudopotential is also evaluated in real space using the expansion coefficients of the KS orbitals. In principle, the Hartree energy could also be evaluated in real space using fast multipole moment (FMM) techniques. However, for simplicity, we choose to employ a hybrid approach in which the Hartree energy and long range part of the local pseudopotential energy are evaluated in reciprocal space using the screening function methodology described in section 4.1. The short range part of the local

Table 1. Convergence of the total energy for the periodic eight-Si-atom system (see text) for the DVR and PW basis sets. Total energies are in Hartrees, and the PW energy cut-off is in Ryd. The energy difference ($|\Delta E|$) measures the difference between the energy at each grid size/PW cut-off and the converged value in kcal mol^{-1}.

Grid size	E(DVR)	$\|\Delta E\|$	PW $E_{\rm cut}$	E(PW)	$\|\Delta E\|$
16^3	−31.8094	2.3845	4	−30.8337	609.8770
20^3	−31.8057	0.0628	6	−31.2026	378.3885
32^3	−31.8056	0	20	−31.7936	7.5301
48^3	−31.8056	0	50	−31.8050	0.3765
60^3	−31.8056	0	80	−31.8051	0.3138

pseudopotential is still evaluated in real space. The hybrid method involves the evaluation of only two FFTs, one to obtain the density in reciprocal space and the other to transform the long range contributions to the KS potential back to real space.

In order to demonstrate the validity of the DVR approach and compare its convergence with basis set size to the standard PW scheme, we consider a system consisting of eight silicon atoms in a simple cubic arrangement in a cubic periodic box of length 5.3 Å. The Kohn–Sham orbitals are expanded about the Γ-point, exchange and correlation are treated within the LDA and a Bachelet–Hamann–Schlüter type pseudopotential [117] is employed. Table 1 shows that convergence to within 1 kcal mol^{-1} of the total energy with DVR grid size and PW energy cut-off (chosen to give a real-space grid of the same size as the DVR grid) is achieved with a factor of 14 fewer grid points in the DVR basis compared to the size of the FFT grid used in the PW calculation. Note that, while the number of DVR coefficients per orbital on a 20^3 grid is 8000 compared to 5573 PW coefficients per orbital with a 50 Ryd PW cut-off, the number of PW density coefficients is 22 513 and the size of the FFT grid is 110 592 points, while for DVR on a 20^3 grid, the same numbers are 1936 and 8000, respectively. Thus, it can be seen that the number of operations needed to converge the total energy in the DVR basis is considerably fewer than in the PW basis. Table 2 shows the convergence of the atomic forces as measured by the quantity

$$\bar{F} = \sqrt{\frac{1}{N}\sum_{I=1}^{N} \boldsymbol{F}_I \cdot \boldsymbol{F}_I} \tag{4.60}$$

where $\boldsymbol{F}_I$ is the force on atom I. The ability to generate accurate forces is of critical importance in AIMD simulations. Again, it can be seen that complete convergence of the forces occurs with a factor of 14 fewer grid points in the DVR basis than in the PW basis. Moreover, the size consistency of the DVR basis was tested by plotting the energy per unit cell as a function of lattice constant for eight- and 16-atom systems using DVR grid sizes of 20^3 and 40×20^2, respectively (see figure 2). A similar PW control calculation was also performed using a converged basis set size. It can be seen that the energy curves for each system size are very similar and that the DVR and PW results match each other nearly perfectly.

5. The Car–Parrinello algorithm

5.1. The basic algorithm

In order to obtain the ground-state energy and forces, the KS functional in equation (3.3) must be minimized over the set of single-particle orbitals subject to an orthonormality condition

$$\langle \psi_i | \psi_j \rangle = \delta_{ij}. \tag{5.1}$$

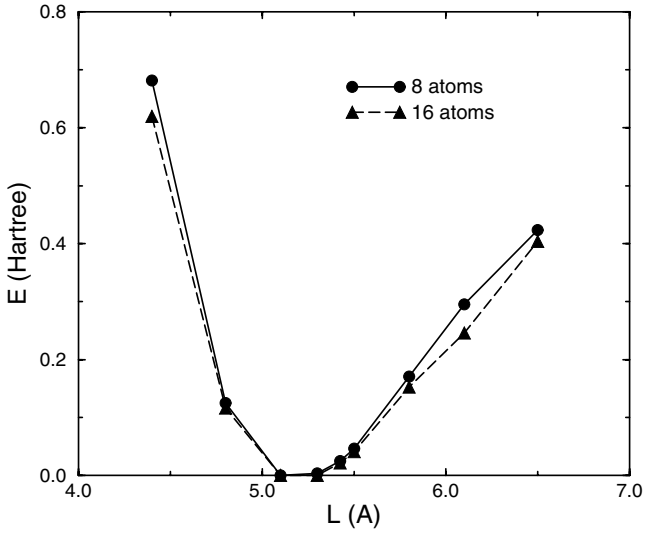

Figure 2. Energy per unit cell as a function of lattice constant (L) using the DVR approach for an eight-Si-atom system (solid curve) and a 16-Si-atom system (dashed curve), respectively. The circles and triangles indicate the results from a converged plane-wave based calculation on the same system.

Table 2. Convergence of atomic forces as measured by equation (4.60) in kcal mol^{-1} Å^{-1} for the periodic eight-Si-atom system (see the text). $|\Delta\bar{F}|$ measures the difference between the force measure and its converged value.

Grid size	$\bar{F}$(DVR)	$\lvert\Delta\bar{F}\rvert$	PW E_{cut}	$\bar{F}$(PW)	$\lvert\Delta\bar{F}\rvert$
16^3	13.32	9.37	4	101.49	97.54
20^3	3.92	0.03	6	12.83	8.88
32^3	3.97	0.02	20	3.89	0.06
48^3	3.95	0	50	3.97	0.02
60^3	3.95	0	80	3.96	0.01

Moreover, in order to combine this minimization with the nuclear dynamics of equation (2.12), it is necessary to carry out the minimization at each nuclear configuration. Thus, if equation (2.12) is integrated in an MD calculation using a numerical integrator, then the minimization would need to be carried out at each step of the MD simulation and the forces computed using the orbitals thus obtained.

Although a step-by-step minimization is an acceptable method for performing AIMD simulations, its use can have a high computational overhead if accurate energy conservation is desired [10]. In 1985, Car and Parrinello (CP) showed that this coupling between nuclear time evolution and electronic minimization could be treated efficiently via an implicit adiabatic dynamics approach [3]. In their scheme, a fictitious dynamics for the electronic orbitals is invented which, given orbitals initially at the minimum for an initial nuclear configuration, allows them to follow the nuclear motion adiabatically, and thus, be automatically at the approximately minimized configuration at each step of the MD evolution.

In order to understand how the CP scheme works, consider a simple model system with two degrees of freedom, x and y, described by a Hamiltonian $H = p_x^2/2m_x + p_y^2/2m_y + V(x, y)$ with $m_x \ll m_y$. Suppose, further, that x and y are maintained at separate temperatures T_x

and T_y, via a thermostat coupling, such that $T_x \ll T_y$. The resulting dynamics under these conditions can be analysed in a rigorous manner [17]. The essence of the analysis is that the y motion can be shown to be driven by a time-averaged force over the x motion. That is, if Δt is a time interval characteristic of the y motion, then a numerical evolution step for y appears as [136]

$$y(\Delta t) = y(0) + \Delta t v_y(0) + \frac{\Delta t}{m_y} \int_0^{\Delta t/2} \mathrm{d}t\, F_y[x_{\mathrm{adb}}(x(0), v_x(0), y(0)); t] \quad (5.2)$$

where $x_{\mathrm{adb}}(x(0), v_x(0), y(0); t)$ denotes the adiabatic evolution of x up to a time $t = \Delta t/2$ starting from initial conditions $x(0)$, $v_x(0)$ at fixed $y = y(0)$. If the system is assumed to be ergodic, then the time average in equation (5.2) can be replaced by a phase space average at fixed y [136],

$$\frac{2}{\Delta t} \int_0^{\Delta t/2} \mathrm{d}t\, F_y[x_{\mathrm{adb}}(x(0), v_x(0), y(0)); t] = \frac{\int \mathrm{d}x\, F_y(x, y) \mathrm{e}^{-\beta_x V(x,y)}}{\int \mathrm{d}x\, \mathrm{e}^{-\beta_x V(x,y)}}. \quad (5.3)$$

The averaged force is derivable from $\langle F_y(x, y)\rangle = (\partial/\partial y)(1/\beta_x) \ln Z_x(y; \beta_x)$ where

$$Z_x(y; \beta_x) = \int \mathrm{d}x\, \mathrm{e}^{-\beta_x V(x,y)} \quad (5.4)$$

from which it follows that the probability distribution of y in configuration space is $P(y) \propto (\int \mathrm{d}x\, \exp[-\beta_x V(x, y)])^{\beta_y/\beta_x}$, where $\beta_x = 1/kT_x$ and $\beta_y = 1/kT_y$. In the limit $T_x \to 0$, $\beta_x \to \infty$, the distribution can be shown to reduce to $P(y) \propto \exp[-\beta_y \min_x V(x, y)]$, where the potential is minimized with respect to x at fixed y. This means that y can be described by an effective Hamiltonian $H_y = p_y^2/2m_y + \min_x V(x, y)$, which is the desired form for AIMD. That is, for $T_x \ll T_y$, only very small fluctuations about $\min_x V(x, y)$ will contribute to the distribution, and y will move on a very good approximation to the correctly minimized surface. As an example illustrating good versus poor adiabatic following for this simple x–y problem, the reader should see figure 10 of [136]. The dynamics of x will be a fictitious *adiabatic* dynamics in this case that only serves to generate the approximately minimized potential at each value of y.

This idea can be realized in AIMD by introducing a fictitious dynamics for the electrons (analogous to x) via a set of orbital 'velocities' $\{\dot\psi_i(\boldsymbol{r})\}$ and a fictitious electronic 'kinetic energy' (not to be confused with the true quantum kinetic energy) given by

$$K_{\mathrm{fict}} = \mu \sum_i \langle \dot\psi_i | \dot\psi_i \rangle \quad (5.5)$$

where μ is a fictitious mass parameter (having units of energy $\times$ time^2) that controls the timescale on which the electrons 'evolve' with the condition that $T_{\mathrm{elec}} \ll T_{\mathrm{ion}}$. The orbitals are incorporated into an extended dynamical system described by a Lagrangian[1] of the form [3]

$$L = \mu \sum_i \langle \dot\psi_i | \dot\psi_i \rangle + \tfrac{1}{2} \sum_{I=1}^{N} M_I \dot{\boldsymbol{R}}_I^2 - E[\{\psi\}, \boldsymbol{R}] + \sum_{i,j} [\Lambda_{ij}(\langle \psi_i | \psi_j \rangle - \delta_{ij})] \quad (5.6)$$

where $E[\{\psi\}, \boldsymbol{R}] = \varepsilon[\{\psi\}, \boldsymbol{R}] + V_{\mathrm{NN}}(\boldsymbol{R})$. Here $V_{\mathrm{NN}}(\boldsymbol{R})$ is now the electrostatic ion–ion repulsion. The matrix Λ_{ij} is a set of Lagrange multipliers introduced in order to ensure that the condition $\langle \psi_i | \psi_j \rangle = \delta_{ij}$ is satisfied dynamically as a constraint. This Lagrangian specifies the true dynamics of the ions and a fictitious adiabatic dynamics for the electrons that

[1] Although the problem could just as well be formulated in terms of an extended Hamiltonian as in the simple x–y model, we prefer to use the Lagrangian formulation as in the original CP paper [3].

generates the instantaneous forces on the ions from the approximately minimized electronic configuration. The equations of motion are obtained from the usual Euler–Lagrange equations:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\delta L}{\delta\dot{\psi}_i^*(\boldsymbol{r})}\right) - \frac{\delta L}{\delta\psi_i^*(\boldsymbol{r})} &= 0\\ \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial\dot{\boldsymbol{R}}_I}\right) - \frac{\partial L}{\partial\boldsymbol{R}_I} &= 0\end{aligned} \tag{5.7}$$

which gives the following coupled dynamical equations of motion:

$$\begin{aligned}M_I\ddot{\boldsymbol{R}}_I &= -\nabla_I E[\{\psi\},\boldsymbol{R}]\\ \mu\ddot{\psi}_i(\boldsymbol{r}) &= -\frac{\delta}{\delta\psi_i^*(\boldsymbol{r})}E[\{\psi\},\boldsymbol{R}] + \sum_j \Lambda_j\psi_j(\boldsymbol{r}).\end{aligned} \tag{5.8}$$

These are known as the CP equations. The electronic equation can also be written in an abstract bra–ket form as

$$\mu|\ddot{\psi}_i\rangle = -\frac{\partial E}{\partial\langle\psi_i|} + \sum_j \Lambda_{ij}|\psi_j\rangle. \tag{5.9}$$

In reality, equation (5.9) formally represents an equation of motion for the expansion coefficients of the orbitals. Thus, for plane wave and DVR basis sets, the electronic equation of motion reads

$$\begin{aligned}\mu\ddot{c}_{i,\boldsymbol{g}} &= -\frac{\partial E}{\partial c_{i,\boldsymbol{g}}^*} + \sum_j \Lambda_{ij}c_{j,\boldsymbol{g}}\\ \mu\ddot{C}_{lmn}^i &= -\frac{\partial E}{\partial C_{lmn}^{i*}} + \sum_j \Lambda_{ij}C_{lmn}^j\end{aligned} \tag{5.10}$$

respectively. For Gaussian basis sets, the equations of motion are complicated somewhat by the position dependence of the basis functions and their nonorthogonality and will be deferred briefly. In abstract form, the coefficient 'force' appearing in the CP equations, $-\partial E/\partial\langle\psi_i|$, can be expressed formally in terms of the action of the KS Hamiltonian on the orbitals as $f_i H_{\mathrm{KS}}|\psi_i\rangle$. Below, an algorithm [137] for integrating the CP equations subject to the orthonormality constraint will be presented.

Beginning with an initially minimized set of Kohn–Sham orbitals $\{|\psi_i(0)\rangle\}$, corresponding to an initial nuclear configuration $\boldsymbol{R}(0)$, and initial velocities $\{\dot{\psi}_i(0)\rangle\}$, $\dot{\boldsymbol{R}}(0)$, the first step is a velocity update,

$$\begin{aligned}|\dot{\psi}_i^{(1)}(0)\rangle &= |\dot{\psi}_i\rangle + \frac{\Delta t}{2\mu}|\varphi_i(0)\rangle \qquad i = 1,\ldots,n_{\mathrm{occ}}\\ \dot{\boldsymbol{R}}_I(\Delta t/2) &= \dot{\boldsymbol{R}}_I(0) + \frac{\Delta t}{2M_I}\boldsymbol{F}_I(0) \qquad I = 1,\ldots,N\end{aligned} \tag{5.11}$$

followed by a position/orbital update,

$$\begin{aligned}|\tilde{\psi}_i\rangle &= |\psi_i(0)\rangle + \Delta t|\dot{\psi}_i^{(1)}\rangle \qquad i = 1,\ldots,n_{\mathrm{occ}}\\ \boldsymbol{R}_I(\Delta t) &= \boldsymbol{R}_I(0) + \Delta t\dot{\boldsymbol{R}}_I(\Delta t/2) \qquad I = 1,\ldots,N\end{aligned} \tag{5.12}$$

where $|\varphi_i(0)\rangle = (\partial E/\partial\langle\psi_i|)|_{t=0}$ is the initial force on the orbital, $|\psi_i\rangle$. At this point, we do not yet have the orbitals at $t = \Delta t$ or orbital velocities at $t = \Delta t/2$ because the constraint force $\Lambda_{ij}|\psi_j\rangle$ needs to be applied to both the orbitals and orbital velocities via the Lagrange multiplier matrix. This is accomplished by enforcing the orthogonality constraint on the orbitals at $t = \Delta t$:

$$\langle\psi_i(\Delta t)|\psi_j(\Delta t)\rangle = \delta_{ij} \tag{5.13}$$

where

$$|\psi_i(\Delta t)\rangle = |\tilde{\psi}_i\rangle + \sum_j X_{ij}|\psi_j(0)\rangle \tag{5.14}$$

where $X_{ij} = (\Delta t^2/2\mu)\Lambda_{ij}$. The multipliers, X_{ij}, are determined such that equation (5.13) is satisfied. Substituting equation (5.14) into (5.13) yields a matrix equation for the Lagrange multipliers:

$$XX^\dagger + XB + B^\dagger X^\dagger + A = \mathrm{I} \tag{5.15}$$

where $A_{ij} = \langle\tilde{\psi}_i|\tilde{\psi}_j\rangle$ and $B_{ij} = \langle\psi_i(0)|\tilde{\psi}_j\rangle$. Noting that $A = \mathrm{I} + \mathcal{O}(\Delta t^2)$ and $B = \mathrm{I} + \mathcal{O}(\Delta t)$, the matrix equation can be solved iteratively via

$$X_{n+1} = \tfrac{1}{2}[\mathrm{I} - A + X_n(\mathrm{I} - B) + (I - B^\dagger)X_n^\dagger - X_n^2] \tag{5.16}$$

starting from an initial guess

$$X_0 = \tfrac{1}{2}(\mathrm{I} - A). \tag{5.17}$$

Once the matrix X_{ij} is obtained, the orbitals are updated using equation (5.14) and an orbital velocity update

$$|\dot{\psi}_i^{(2)}\rangle = |\dot{\psi}_i^{(1)}\rangle + \frac{1}{\Delta t}\sum_j X_{ij}|\psi_j(0)\rangle \tag{5.18}$$

is performed.

At this point, the new orbital and nuclear forces, $|\varphi_i(\Delta t)\rangle$ and $\boldsymbol{F}_I(\Delta t)$, are calculated, and a velocity update of the form

$$\begin{aligned} |\dot{\psi}_i^{(3)}\rangle &= |\dot{\psi}_i^{(2)}\rangle + \frac{\Delta t}{2\mu}|\varphi_i(\Delta t)\rangle \qquad i = 1, \ldots, n_{\mathrm{occ}} \\ \dot{\boldsymbol{R}}_I(\Delta t) &= \dot{\boldsymbol{R}}_I(\Delta t/2) + \frac{\Delta t}{2M_I}\boldsymbol{F}_I(\Delta t) \end{aligned} \tag{5.19}$$

is performed. Again, we do not have the final orbital velocities until an appropriate constraint force is applied. For the velocities, the appropriate force is the first time derivative of the orthogonality constraint:

$$\langle\psi_i(\Delta t)|\dot{\psi}_j(\Delta t)\rangle + \langle\dot{\psi}_i(\Delta t)|\psi_j(\Delta t)\rangle = 0 \tag{5.20}$$

where

$$|\dot{\psi}_i(\Delta t)\rangle = |\dot{\psi}_i^{(3)}\rangle + \sum_j Y_{ij}|\psi_i(\Delta t)\rangle \tag{5.21}$$

and Y_{ij} are a new set of Lagrange multipliers for enforcing the condition equation (5.20). Substituting equation (5.21) into (5.20) gives a simple solution for Y_{ij}:

$$Y = -\tfrac{1}{2}(C + C^\dagger) \tag{5.22}$$

where $C_{ij} = \langle\psi_i(\Delta t)|\dot{\psi}_i^{(3)}\rangle$. Given the matrix, Y_{ij}, the final orbital velocities are obtained via equation (5.21).

Under certain circumstances, the orthonormality constraint can become dependent on atomic positions, in which case the basic CP algorithm acquires another level of complexity. The position dependence can come about, for example, when Gaussian basis sets are used. It can also occur when using the so-called ultrasoft pseudopotential scheme [119, 138, 139], in which the standard orthogonality constraint is replaced by a more general constraint of the form

$$\langle\psi_i|\hat{B}(\boldsymbol{R})|\psi_j\rangle = \delta_{ij} \tag{5.23}$$

resulting from a relaxation of the norm-conservation condition. Here, $\hat{B}$ is a position-dependent operator required in the pseudopotential formulation. For details of this scheme, the interested reader is referred to [119, 138, 139]. Defining the general overlap matrix as $S_{ij}(\boldsymbol{R}) = \langle\psi_i|\hat{B}(\boldsymbol{R})|\psi_j\rangle$, which reduces to the ordinary overlap when $\hat{B} = I$, it is clear that whenever S is position dependent, the CP equations need to be modified to read [138–140]

$$\begin{aligned}\mu|\ddot{\psi}_i\rangle &= -\frac{\partial E}{\partial\langle\psi_i|} + \sum_{k,j}\Lambda_{kj}\frac{\partial S_{kj}(\boldsymbol{R})}{\partial\langle\psi_i|} = -\frac{\partial E}{\partial\langle\psi_i|} + \sum_j \Lambda_{ij}\hat{B}(\boldsymbol{R})|\psi_j\rangle \\ M_I\ddot{\boldsymbol{R}}_I &= -\frac{\partial E}{\partial \boldsymbol{R}_I} + \sum_{i,j}\Lambda_{ij}\frac{\partial S_{ij}(\boldsymbol{R})}{\partial \boldsymbol{R}_I} = -\frac{\partial E}{\partial \boldsymbol{R}_I} + \sum_{i,j}\Lambda_{ij}\langle\psi_i|\nabla_I\hat{B}(\boldsymbol{R})|\psi_j\rangle\end{aligned} \tag{5.24}$$

which includes a contribution to the nuclear equation of motion from the constraint. Because the constraint is now coupled to both the electronic and nuclear equations of motion, it is necessary to iterate the constraint procedure through the nuclear update [139, 140], which can have a high computational overhead. One way to overcome this problem is to employ nonorthogonal orbitals as described in the next section.

5.2. Nonorthogonal orbitals

As alluded to above, the reformulation of the electronic structure problem in terms of a set of nonorthogonal orbitals has a number of advantages in AIMD; in particular, it can simplify the problem of a position-dependent overlap matrix. It can also aid in the control of adiabaticity as will be discussed in section 5.3. The standard orthogonal orbitals $|\psi_i\rangle$ may be transformed to a set of nonorthogonal orbitals $|\phi_i\rangle$ via a transformation of the form

$$|\psi_i\rangle = \sum_j |\phi_j\rangle T_{ji} \tag{5.25}$$

where the matrix T is defined to be

$$T = O^{-1/2} \tag{5.26}$$

where O is the overlap matrix with respect to the nonorthogonal orbitals:

$$O_{ij} = \langle\phi_i|\hat{B}(\boldsymbol{R})|\phi_j\rangle. \tag{5.27}$$

Note that if $\hat{\mathrm{B}} = \mathrm{I}$, O is just the standard overlap of the nonorthogonal orbitals. It is easily verified that the this transformation preserves the generalized orthogonality of the original orbitals.

For use in the CP equations of motion, we begin with an extended Lagrangian of the form [140]

$$L = \mu\sum_i\langle\dot{\phi}_i|\dot{\phi}_i\rangle + \tfrac{1}{2}\sum_I M_I\dot{\boldsymbol{R}}_I^2 - f\sum_{i,j}M_{ji}\langle\phi_i|H_{\mathrm{KS}}|\phi_j\rangle + \sum_\alpha\lambda_\alpha\sigma_\alpha[\{\phi\}] \tag{5.28}$$

where $M = O^{-1}$ and uniform occupation number, $f_i = f$, have been assumed. In equation (5.28), the fictitious kinetic energy is expressed directly in terms of the nonorthogonal orbitals, and the energy functional is expressed in terms of the nonorthogonal orbitals via the transformation in equation (5.25). The last term in equation (5.28) involves an arbitrary set of constraints, $\sigma_\alpha[\{\phi\}]$, $\alpha = 1, \ldots, N_c$, enforced by a set of Lagrange multipliers, $\{\lambda_\alpha\}$. The introduction of a constraint into the CP dynamics is done as a means of preventing the orbitals from becoming linearly dependent as the dynamics proceeds. The precise form of the constraint is not particularly important and, therefore, can be chosen to be simpler than the generalized orthogonality constraint obeyed by the original orbitals. This method is known as

the constrained nonorthogonal orbital (CNO) approach [140]. Two possibilities are a simple orthogonality constraint,

$$\langle \phi_i | \phi_j \rangle = \delta_{ij} \tag{5.29}$$

which involves $n_{\rm occ} \times n_{\rm occ}$ constraints, or a simple norm constraint,

$$\langle \phi_i | \phi_i \rangle = 1 \tag{5.30}$$

which involves just $n_{\rm occ}$ constraints. Finally, it is easily verified that the energy derivatives required for the CP equations of motion are

$$\begin{aligned} \frac{\partial E}{\partial \langle \phi_i |} &= f \sum_j \left(\frac{\partial E}{\partial \langle \psi_i |} - \sum_k \hat{B} |\psi_k\rangle \langle \psi_k | \hat{H}_{\rm KS} | \psi_j \rangle \right) T_{ji} \\ \frac{\partial E}{\partial \boldsymbol{R}_I} &= f \sum_i \left(\langle \psi_i | \nabla_I \hat{H}_{\rm KS} | \psi_i \rangle - \sum_j \langle \psi_j | \nabla_I \hat{B} | \psi_j \rangle \langle \psi_j | \hat{H}_{\rm KS} | \psi_i \rangle \right) \end{aligned} \tag{5.31}$$

expressed in terms of matrix elements and derivatives involving the original orbitals. The use of the CNO method with a simple constraint such as the norm constraint condition can yield a non-negligible saving in computational overhead in both the ultrasoft pseudopotential and standard (norm-conserving) pseudopotential ($\hat{B} = I$) schemes [140].

5.3. Adiabaticity control through isokinetic constraints

The CP technique relies heavily on the assumption that an adiabatic separation between the fictitious electron dynamics and the nuclear dynamics can be maintained. In general, this will only be true if the separation between the ground and first excited electronic surfaces, i.e. the bandgap, is large compared to kT for all nuclear configurations. While this is generally true for insulators and semiconductors, it is not true for metals and can be problematic in numerous chemical reactions where the two surfaces approach each other at a transition state. In fact, for metals, the CP dynamics is not correct because the motion does not occur on the ground-state surface (see section 10), although it can, nevertheless, yield some useful information. A proper treatment will be discussed in section 10. In such cases, the fictitious thermal energy in the electronic subsystem can lead to excitations that destroy adiabaticity and lead to a rapid exchange of energy between the nuclear and electronic subsystems. This will cause the nuclei to cool and the electrons to heat, and the CP dynamics will cease to be meaningful.

One way to ameliorate this problem is to employ one of the widely used thermostatting methods on the electronic subsystem, such as the Nosé–Hoover [141] or Nosé–Hoover chain [142] methods, and several such approaches have been reported [137, 143]. However, these may not be robust enough in certain systems or may allow fluctuations large enough that adiabaticity is lost despite the action of the thermostat. Here, we describe an alternative, highly robust, approach based on the Gaussian isokinetic ensemble method. The adaptation of this method for CP dynamics was recently described by Minary *et al* [144]. The isokinetic ensemble method employs a nonholonomic constraint to keep the fictitious electronic kinetic energy in the CP Lagrangian fixed:

$$\mu \sum_i \langle \dot{\psi}_i | \dot{\psi}_i \rangle = K_{\rm e}. \tag{5.32}$$

In order to impose this constraint, an additional Lagrange multiplier, α, is introduced into the CP equations. Thus, in terms of orthogonal orbitals, the new electronic equation of motion reads

$$\mu |\ddot{\psi}_i\rangle = |\varphi_i\rangle + \sum_j \Lambda_{ij} |\psi_j\rangle - \alpha |\dot{\psi}_i\rangle \tag{5.33}$$

where $|\varphi_i\rangle = \partial E/\partial\langle\psi_i|$. Applying Gauss's principle of least constraint, analytical expressions for the Lagrange multipliers Λ_{ij} and α can be determined analytically by differentiating the orthogonality constraint twice and the isokinetic constraint once yielding

$$\begin{aligned}&\langle\psi_i|\ddot{\psi}_j\rangle + 2\langle\dot{\psi}_i|\dot{\psi}_j\rangle + \langle\ddot{\psi}_i|\psi_j\rangle = 0\\&\sum_i[\langle\dot{\psi}_i|\ddot{\psi}_i\rangle + \langle\ddot{\psi}_i|\dot{\psi}_i\rangle] = 0.\end{aligned} \tag{5.34}$$

Then, using the equation of motion to substitute in for $|\ddot{\psi}_i\rangle$ and its complex conjugate for $\langle\ddot{\psi}_i|$ and solving for the two multipliers, one obtains

$$\begin{aligned}\lambda_{ij} &= \frac{f_i + f_j}{2}\langle\psi_i|\hat{H}_{\mathrm{KS}}|\psi_j\rangle - \mu\langle\dot{\psi}_i|\dot{\psi}_j\rangle\\\alpha &= \frac{1}{2K_{\mathrm{e}}}\sum_i\big[\langle\dot{\psi}_i|\hat{H}_{\mathrm{KS}}|\psi_i\rangle + \langle\psi_i|\hat{H}_{\mathrm{KS}}|\dot{\psi}_i\rangle\big].\end{aligned} \tag{5.35}$$

Remarkably, each of these expressions is what would be obtained for each multiplier in the absence of the other constraint, showing that the two constraints are completely uncoupled. Substituting equations (5.35) into the electronic equation of motion yields

$$\begin{aligned}\mu|\ddot{\psi}_i\rangle = &-f_i\hat{H}_{\mathrm{KS}}|\psi_i\rangle + \frac{f_i + f_j}{2}\sum_j|\psi_i\rangle\langle\psi_i|\hat{H}_{\mathrm{KS}}|\psi_i\rangle\\&- \mu\sum_j|\psi_j\rangle\langle\dot{\psi}_j|\dot{\psi}_i\rangle + \frac{1}{K_{\mathrm{e}}}|\dot{\psi}_i\rangle\sum_j f_j\big[\langle\dot{\psi}_j|\hat{H}_{\mathrm{KS}}|\psi_j\rangle + \langle\psi_j|\hat{H}_{\mathrm{KS}}|\dot{\psi}_j\rangle\big].\end{aligned} \tag{5.36}$$

In principle, equation (5.36) could be solved using a Liouville operator based approach as discussed in [137]. Alternatively, the analytical expression for α could be used while retaining a numerical approach for Λ_{ij}; however, as was shown in [144], this requires an iterative procedure. Therefore, the simplest approach is to combine the isokinetic method with the CNO approach outlined in section 5.2. This scheme would be described by an equation of motion of the form

$$\mu|\ddot{\phi}_i\rangle = |\varphi_i^{\mathrm{CNO}}\rangle + \sum_\alpha\lambda_\alpha\frac{\partial\sigma_\alpha}{\partial\langle\phi_i|} + \frac{f}{2K_{\mathrm{e}}}|\dot{\phi}_i\rangle\sum_j\big[\langle\dot{\phi}_j|\hat{H}_{\mathrm{KS}}|\phi_j\rangle + \langle\phi_j|\hat{H}_{\mathrm{KS}}|\dot{\phi}_j\rangle\big] \tag{5.37}$$

where $|\varphi_i^{(\mathrm{CNO})}\rangle$ is the CNO force in equation (5.31). An algorithm for integrating equation (5.37) was recently presented by Minary *et al* [144].

6. Calculating observables

Up to now, we have discussed a wide variety of simulation techniques including basis sets, orbital choices and adiabaticity control methods. All of this methodology would, of course, be useless if one could not compute experimentally measurable observables. In this regard, AIMD simulations have some distinct advantages over force field based MD in that the former permit direct access to the electronic structure and, hence, any observable that can be derived directly from it. Thus, greatly widens the range of observables that can be computed from MD simulations.

In MD calculations, observables are computed by performing averages of appropriate functions, $O(\boldsymbol{P}, \boldsymbol{R})$, of the momenta and coordinates of the particles in the system. The procedure relies on the ergodic hypothesis which states that given an infinite amount of time, a system will visit all of its accessible phase space so that ensemble averages of $O(\boldsymbol{P}, \boldsymbol{R})$ can be directly related to time averages of the MD trajectory:

$$\langle O(\boldsymbol{P}, \boldsymbol{R})\rangle = \lim_{\mathcal{T}\to\infty}\frac{1}{\mathcal{T}}\int_0^{\mathcal{T}}\mathrm{d}t\, O(\boldsymbol{P}(t), \boldsymbol{R}(t)); \tag{6.1}$$

equation (6.1) will, therefore, yield an equilibrium average for the system. The ensemble average in equation (6.1) could refer to any pertinent ensemble. For example, a microcanonical (NVE) ensemble average would be given by

$$\langle O(\boldsymbol{P},\boldsymbol{R})\rangle = \frac{1}{N!h^{3N}\Omega(N,V,E)}\int \mathrm{d}^N\boldsymbol{P}\,\mathrm{d}^N\boldsymbol{R}\,O(\boldsymbol{P},\boldsymbol{R})\delta(H_\mathrm{N}(\boldsymbol{P},\boldsymbol{R})-E) \tag{6.2}$$

where $H_\mathrm{N}(\boldsymbol{P},\boldsymbol{R})$ is the classical nuclear Hamiltonian of equation (2.11) and $\Omega(N,V,E)$ is the microcanonical ensemble partition function. An average in the canonical (NVT) ensemble is given by

$$\langle O(\boldsymbol{P},\boldsymbol{R})\rangle = \frac{1}{N!h^{3N}Q(N,V,T)}\int \mathrm{d}^N\boldsymbol{P}\,\mathrm{d}^N\boldsymbol{R}\,O(\boldsymbol{P},\boldsymbol{R})\mathrm{e}^{-\beta H_\mathrm{N}(\boldsymbol{P},\boldsymbol{R})}. \tag{6.3}$$

As a concrete examples, note that a radial distribution function (RDF) is given as an average:

$$\begin{aligned} g(r) &= \frac{1}{4\pi r^2\rho N Q(N,V,T)}\int \mathrm{d}^N\boldsymbol{P}\,\mathrm{d}^N\boldsymbol{R}\sum_{I\neq J}\delta(|\boldsymbol{R}_I-\boldsymbol{R}_J|-r)\mathrm{e}^{-\beta H_\mathrm{N}(\boldsymbol{P},\boldsymbol{R})} \\ &= \frac{1}{4\pi\rho r^2}\left\langle\frac{1}{N}\sum_{I\neq J}\delta(|\boldsymbol{R}_I-\boldsymbol{R}_J|-r)\right\rangle. \end{aligned} \tag{6.4}$$

Similarly, the elastic neutron scattering structure factor can be computed by Fourier transforming the RDF or by directly performing an ensemble (or trajectory) average over $S(\boldsymbol{k}) = (1/N)|\sum_{I=1}^{N}\exp(\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{R}_I)|^2$.

Dynamical properties such as spectra and transport coefficients can be obtained within classical linear response theory from time correlation functions. The time correlation function between two observables, $A(\boldsymbol{P},\boldsymbol{R})$ and $B(\boldsymbol{P},\boldsymbol{R})$, is given by

$$\langle A(0)B(t)\rangle = \frac{1}{Q(N,V,T)}\int \mathrm{d}^N\boldsymbol{P}\,\mathrm{d}^N\boldsymbol{R}\,A(\boldsymbol{P},\boldsymbol{R})B(\boldsymbol{P}_t(\boldsymbol{P},\boldsymbol{R}),\boldsymbol{R}_t(\boldsymbol{P},\boldsymbol{R}))\mathrm{e}^{-\beta H_\mathrm{N}(\boldsymbol{P},\boldsymbol{R})} \tag{6.5}$$

where $(\boldsymbol{P}_t(\boldsymbol{P},\boldsymbol{R}),\boldsymbol{R}_t(\boldsymbol{P},\boldsymbol{R}))$ designates the phase space trajectory obtained from the initial condition $(\boldsymbol{P},\boldsymbol{R})$). For $A = B$, equation (6.5) becomes an autocorrelation function. For example, the diffusion coefficient can be computed from the velocity autocorrelation function:

$$D = \frac{1}{3}\int_0^\infty \mathrm{d}t\,\frac{1}{N}\sum_{I=1}^{N}\langle\boldsymbol{V}_I(0)\cdot\boldsymbol{V}_I(t)\rangle. \tag{6.6}$$

The velocity autocorrelation function can also be used to obtain a frequency spectrum for the system known as the *power spectrum* by Fourier transformation:

$$I(\omega) = \int_{-\infty}^{\infty}\mathrm{d}t\,\mathrm{e}^{\mathrm{i}\omega t}\,\frac{1}{N}\sum_{i=1}^{N}\langle\boldsymbol{V}_I(0)\cdot\boldsymbol{V}_I(t)\rangle. \tag{6.7}$$

Although the power spectrum itself is not a directly measurable quantity, the frequencies at which the peaks in the function occur can be compared to other spectroscopic measurements. While this can be a useful comparison, it is obviously preferable to compute, for example, the infrared (IR) or Raman spectrum, which can can be directly measured. For such quantities, having access to the electronic structure is a significant advantage of the AIMD method. Below, we discuss the calculation of several types of spectrum based on the use of the electronic structure.

6.1. Infrared spectrum

In linear response theory, the IR absorption coefficient, $\alpha(\omega)$, is given by the Fourier transform of the electric dipole moment correlation function

$$\alpha(\omega) = \frac{4\pi\omega\tanh(\beta\hbar\omega/2)}{3\hbar n(\omega)cV}\int_{-\infty}^{\infty} \mathrm{d}t\, \mathrm{e}^{-\mathrm{i}\omega t}\left\langle \frac{1}{2}\sum_{k=x,y,z}[\hat{M}_k(0), \hat{M}_k(t)]_+\right\rangle \tag{6.8}$$

where $n(\omega)$ is the index of refraction of the medium, V is the volume, c is the speed of light, $\hat{M}_k(t)$ is the kth component of the electric dipole moment operator and $[\hat{M}_k(0), \hat{M}_k(t)]_+$ is the anticommutator between the operators, $[A, B]_+ = AB + BA$. Of the two terms in the anticommutator, $\langle\hat{M}_k(t)\hat{M}_k(0)\rangle$ corresponds to an absorption process while $\langle\hat{M}_k(0)\hat{M}_k(t)\rangle$ corresponds to an emission process. In the approximation of classical nuclei, the dipole moment operator is replaced by the classical dipole moment function. However, in order to retain some quantum information, the factor $\tanh(\beta\hbar\omega/2)$ is retained so that the IR absorption is expressed as

$$\alpha(\omega) = \frac{4\pi\omega\tanh(\beta\hbar\omega/2)}{3\hbar n(\omega)cV}\int_{-\infty}^{\infty} \mathrm{d}t\, \mathrm{e}^{-\mathrm{i}\omega t}\left\langle \sum_{k=x,y,z}\hat{M}_k(0)\hat{M}_k(t)\right\rangle. \tag{6.9}$$

Note that more sophisticated quantum corrections, for example, Egelstaff type corrections [145], can also be applied. The total dipole moment can be decomposed into ionic and electronic contributions according to

$$\boldsymbol{M} = \boldsymbol{M}^{(\mathrm{ion})} + \boldsymbol{M}^{(\mathrm{elec})}. \tag{6.10}$$

Since the ions are treated as classical point particles, the ionic contribution can be computed straightforwardly from the particle positions. The electronic contribution is more subtle. For a non-periodic or cluster system, the expression

$$\boldsymbol{M}^{(\mathrm{elec})} = -e\int \mathrm{d}\boldsymbol{x}_1\cdots\mathrm{d}\boldsymbol{x}_{N_e}\,\phi_0^*(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{N_e})\left[\sum_{i=1}^{N_e}\boldsymbol{r}_i\right]\phi_0(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{N_e}) = -e\int \mathrm{d}\boldsymbol{r}\, n_0(\boldsymbol{r})\boldsymbol{r} \tag{6.11}$$

where ϕ_0 is the exact ground-state wavefunction, $n_0(\boldsymbol{r})$ is the ground-state density and $\boldsymbol{r}_i$ is the ith electron position operator, can be evaluated easily because the dipole moment operator is a one-body operator. For periodic systems, however, equation (6.11) is not translationally invariant. The proper generalization of the dipole moment expression is based on the so-called Berry phase approach [146–149] and takes the form

$$M_k^{(\mathrm{elec})} = -e\,\mathrm{Im}\,\ln\int \mathrm{d}\boldsymbol{x}_1\cdots\mathrm{d}\boldsymbol{x}_{N_e}\,\phi_0^*(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{N_e})\left[\mathrm{e}^{2\pi\mathrm{i}\sum_{i=1}^{N_e} r_{i,k}/L_k}\right]\phi_0(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{N_e}) \tag{6.12}$$

where $M_k^{(\mathrm{elec})}$ is the kth component of the electronic contribution to the dipole moment, $r_{i,k}$ is the kth component of the ith electron position operator and L_k is the length of the supercell in the kth direction (assuming an orthorhombic cell—see [150, 151] for generalizations to arbitrary cell shapes), which is assumed to be cubic in this case. Clearly, equation (6.12) possesses the correct translational symmetry. However, it introduces an additional complication because the operator $\exp(2\pi\mathrm{i}\sum_i r_{i,k}/L_k)$ is now a many-body operator [148] so that the expectation value in equation (6.12) cannot be expressed simply in terms of the electron density. As was shown in [146, 147, 152], in order to compute the electronic component of the dipole moment within KS theory, it is necessary to compute the matrix elements

$$R_{ij,k} = \langle\psi_i|\mathrm{e}^{2\pi\mathrm{i}r_k/L_k}|\psi_j\rangle \tag{6.13}$$

using the KS orbitals (or their periodic parts when $\boldsymbol{k}$-points other than the Γ-point are used) and compute the dipole moment contribution from

$$M_k^{(\mathrm{elec})} = -\frac{eL_k}{\pi}\,\mathrm{Im}\,\ln\det(R_{ij,k}). \tag{6.14}$$

It should be noted that the Berry phase approach is also the starting point for methods to find unitary transformations among the orbitals that lead to a maximally set of orbitals known as *Wannier functions* [150–152], which have been shown to be of great utility in studying other electronic properties of a system, for example, local dipole moments [24].

6.2. Other spectra

Recently, a general formalism within DFT was introduced to compute the response of a system to a small applied external perturbation [153], which is often what is measured in experiments. The resulting variational DFT perturbation theory, which is in the same spirit as, though conceptually different from, more standard formulations [154–156], leads to a general scheme for computing the second derivative of the energy with respect to an applied field, and, therefore, incorporates observables such as anharmonic Raman spectra [66] and nuclear magnetic resonance (NMR) chemical shifts [157].

In the case of Raman spectroscopy, it is necessary to compute the full polarizability tensor [66, 158]

$$\alpha_{\mu\nu} = -\frac{\partial \boldsymbol{P}_\mu}{\partial \boldsymbol{E}_\nu} = \frac{\partial^2 E}{\partial \boldsymbol{E}_\mu \boldsymbol{E}_\nu} \qquad \mu, \nu = x, y, z \tag{6.15}$$

where $\boldsymbol{P}$ is the induced polarization vector due to an externally applied electric field, $\boldsymbol{E}$, and E is the total energy. Within linear response theory, the Raman scattering cross section can then be related to the autocorrelation functions of the polarizability tensor. For cubic systems, the full tensor can be expressed as [66]

$$\alpha_{\mu\nu}(t) = \lambda(t)\delta_{\mu\nu} + \beta_{\mu\nu}(t) \tag{6.16}$$

where $\beta_{\mu\nu}(t)$ is traceless. This subdivision will lead to isotropic and anisotropic spectra given by [66]

$$\begin{aligned} I_{\mathrm{iso}}(\omega) &= \frac{N}{2\pi}\int \mathrm{d}t\, \mathrm{e}^{-\mathrm{i}\omega t}\langle\lambda(0)\lambda(t)\rangle \\ I_{\mathrm{aniso}}(\omega) &= \frac{N}{2\pi}\int \mathrm{d}t\, \mathrm{e}^{-\mathrm{i}\omega t}\frac{1}{10}\langle \mathrm{Tr}[\beta(0)\beta(t)]\rangle. \end{aligned} \tag{6.17}$$

Again, the Berry phase formalism can be used to compute the polarization of the system [66].

Finally, it is important to mention that the DFT perturbation theory has also been employed to derive a new approach to the calculation of NMR chemical shifts in periodic systems [157]. NMR spectra are among the most important tools used in chemistry to characterize a chemical environment. Indeed, a variety of novel approaches have been introduced for computing the chemical shifts [27, 157, 159, 160]. Unfortunately, the complexities of this methodology are such that they will not be dealt with here in any amount of detail, but the interested reader is referred to the above mentioned literature. However, in order to give a feel for the problem, we note that the chemical shift tensor, like the polarizability tensor, is expressible as the derivative of an induced effect due to an external field. In particular, it is the derivative of the induced local magnetic field due to an externally applied magnetic field, $\boldsymbol{B}$:

$$\sigma_{\mu\nu}(\boldsymbol{r}) = \frac{\partial \boldsymbol{B}_\mu^{(\mathrm{ind})}(\boldsymbol{r})}{\partial \boldsymbol{B}_\nu}. \tag{6.18}$$

The induced field is determined by the total electronic current $\boldsymbol{j}(\boldsymbol{r})$ via

$$\boldsymbol{B}^{(\mathrm{ind})}(\boldsymbol{r}) = \frac{\mu_0}{4\pi}\int \mathrm{d}\boldsymbol{r}'\,\frac{(\boldsymbol{r}-\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}\times\boldsymbol{j}(\boldsymbol{r}) \tag{6.19}$$

where μ_0 is the permeability of free space. For a Hamiltonian based formalism, the applied magnetic field, $\boldsymbol{B}$, is represented in terms of a vector potential, $\boldsymbol{A}(\boldsymbol{r})$ with $\boldsymbol{B} = \nabla\times\boldsymbol{A}(\boldsymbol{r})$. Thus, a gauge choice must be made for $\boldsymbol{A}(\boldsymbol{r})$, and a typical form is [157]

$$\boldsymbol{A}(\boldsymbol{r}) = -\tfrac{1}{2}(\boldsymbol{r}-\boldsymbol{R})\times\boldsymbol{B} \tag{6.20}$$

where $\boldsymbol{R}$ is known as the gauge origin. Thus, it can be seen that the position operator problem in periodic systems is prevalent here as well. In order to solve the problem, Mauri *et al* proposed modulating $\boldsymbol{B}$ by a periodic function [27, 161–165], whereas Sebastiani and Parrinello have applied the Berry phase approach [157]. Another technical difficulty that arises in a plane-wave basis stems from the fact that the chemical shifts are very sensitive to the shape of the wavefunction in the core region, which is pseudized. This problem has been addressed by Gregor *et al* [159], who have developed a set of additive constant corrections that accurately reproduce the all-electron magnetic shieldings.

7. Structure and dynamics in liquids and solutions

7.1. Structure and dynamics of liquid water

Because of its obvious special importance, liquid water at 300 K has received considerable attention in the AIMD literature [18–30]. Obtaining accurate RDFs compared to experimental measurements [166–170] has proved to be challenging because of both system and basis set size effects in the AIMD simulations [24, 30] and because of the difficulty of extracting these quantities directly from the neutron and x-ray diffraction data [171]. However, the most recent studies [30] have shown that reasonable agreement is obtained with the latest experimental data [169, 170] using a system of 64 water molecules with a plane-wave cut-off of 80 Ryd. In this calculation, Troullier–Martins type atomic pseudopotentials were used to represent core electrons [118] and the exchange functional of Becke [103] and the correlation functional of Lee and Parr [104] (the so-called BLYP functional) were employed. A run length of 11 ps with a time step of 0.17 fs at the Γ-point of the Brillouin zone was carried out and the RDFs extracted from the trajectory. The tendency toward overstructuring in the OH and HH distribution functions is attributed to the neglect of nuclear quantum effects (see section 9). Another recent study [29] showed that the exchange correlation functional of Perdew *et al* [106] also yields an OO RDF in good agreement with experiment for a run of length 7.5 ps.

In addition to structural properties of water, AIMD simulations have been performed to compute a variety of other properties, including the IR spectrum [21] and the average dipole moment [24], both directly connected to the electronic structure of the system, using the techniques alluded to in section 6. In both cases, good agreement with experiment [172–174] was obtained using system sizes of 32 and 64 water molecules, respectively, a plane-wave cut-off of 70 Ryd and the BLYP functional, and run lengths of 10–12 ps. Finally, the problem of autoionization of water was addressed by Geissler *et al* [28] in a novel study that combined the transition path sampling method [175, 176] with AIMD. It was shown that a metastable charge-separated state requires the breaking of a hydrogen bond along the hydrogen-bonded 'wire' or 'chain' of water molecules connecting a nascent contact H_3O^+/OH^- ion pair formed upon dissociation.

7.2. Structure and dynamics of liquid ammonia

Ammonia is an important weakly hydrogen-bonded liquid that is employed as a solvent in many common organic reactions and in solutions with metals. Its structure has recently been

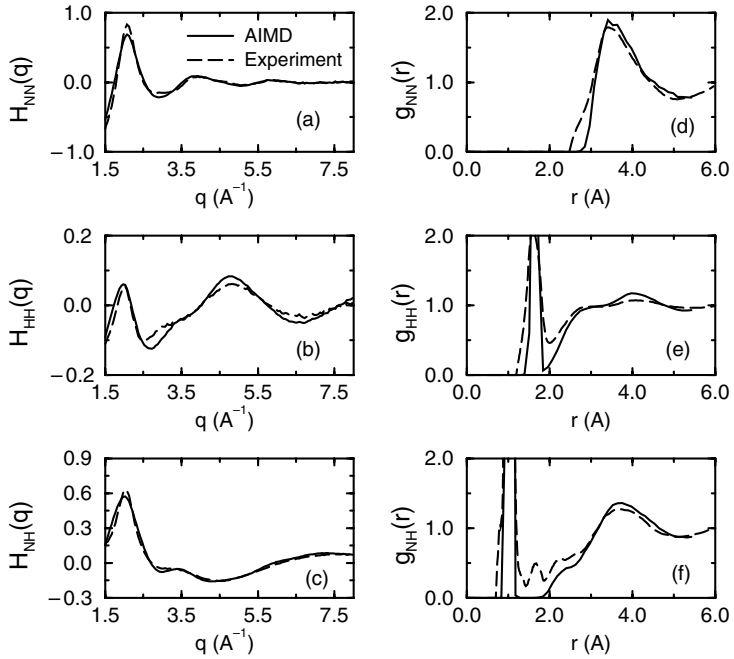

Figure 3. The neutron scattering partial structure factors and RDFs for liquid ammonia at 260 K. Panels (a)–(c) compare, respectively, the computed NN, HH and NH partial structure factors (solid curves) [31, 33] to those from experiment (dashed curves) [177]. Panels (d)–(f) show the same comparison for the RDFs.

determined experimentally by neutron diffraction [177] so that experimental partial structure factors and RDFs are now available. As studies of chemical processes in ammonia solvent are a prime application for AIMD techniques, it is important to validate the approach by studying the properties of the neat liquid and comparing to available experimental data.

To this end, AIMD simulations based on the CP equations of motion have been carried out on a sample of 32 ammonia molecules in a box of length 11.27 Å with periodic boundary conditions at temperatures of 260 and 273 K [31, 33]. Exchange and correlation were treated using the BLYP GGA functional [103, 104] and the KS orbitals were expanded in a plane-wave basis up to a cut-off of 70 Ryd. Core electrons were treated using the pseudopotentials of Troullier and Martins [118]. The system was allowed to equilibrate for 2.2 ps, and production runs of 6.0 ps were carried out using a time step of 0.12 fs.

Figure 3 shows the computed neutron scattering partial structure factors, $H_{\mathrm{NN}}(q)$, $H_{\mathrm{HH}}(q)$ and $H_{\mathrm{NH}}(q)$, together with the experimental results [177]. As can be seen, very good agreement with experiment is obtained. Also shown in figure 3 are the computed RDFs together with those extracted from the experimental data [177]. Again, reasonably good agreement is obtained. In addition, the self-diffusion constant was determined from the calculation to be 1.1×10^{-4} cm^2 s^{-1} [33], which compares favourably with the experimental value of 1.0×10^{-4} cm^2 s^{-1} [178].

7.3. Structure of liquid methanol at 300 K

Another important hydrogen-bonded liquid is methanol (CH_3OH). Like ammonia, methanol is also used as a solvent in many common organic reactions. It is also a industrially important

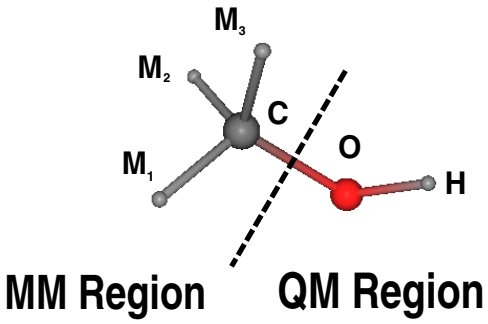

Figure 4. Illustration of the subdivision of a methanol molecule into a QM region and an MM region. The figure shows that the division occurs by cutting the CO bond. Methyl hydrogens are denoted M_i, $i = 1, 2, 3$.

liquid because of its role in emerging fuel-cell technologies. The structure of liquid methanol has also been determined recently by neutron diffraction [179, 180], again making partial structure factors and RDFs readily available for comparison with AIMD calculations. We have recently carried out studies of proton transport in liquid methanol [34], and as part of these studies, we carried out a simulation of the pure liquid in order to compare the structural properties with those determined experimentally.

The AIMD simulation protocol employed 32 methanol molecules in a periodic box of 12.93 Å. Exchange and correlation were, again, treated using the BLYP GGA functional. Studies employing two different modelling schemes were carried out. First, an ultrasoft pseudopotential approach [119, 138, 139] with a plane-wave cut-off of 25 Ryd was employed for a run length of 20 ps. In a second scheme, a hybrid DFT/force field approach (so-called quantum mechanical/molecular mechanical or QM/MM scheme—see section 8.5) was developed [35] in which the OH group was treated at the QM level via the BLYP functional [103, 104] and the CH_3 group was treated at the MM level via the AMBER force field [181] (see figure 4). In this scheme, the CO bond bridges the QM and MM regions in each methanol molecule via a monovalent carbon pseudopotential [182–184], and the electrons interact with the MM atoms with a pseudopotential of the form $V_{\mathrm{pseud}}(r) = A\exp(-\gamma r)$ with $A = 81.5763$ kcal mol^{-1} and $\gamma = 0.5292$ Å. For this case, a run length of 10 ps was carried out [35].

Figure 5 shows the computed and experimentally determined partial structure factors. The collective partial structure factors $H_{\mathrm{XX}}(q)$ and $H_{\mathrm{XH}}(q)$ are defined to be

$$H_{\mathrm{XX}}(q) = 0.042H_{\mathrm{CC}}(q) + 0.073H_{\mathrm{CO}}(q) + 0.253H_{\mathrm{CM}}(q) + 0.032H_{\mathrm{OO}}(q) + 0.220H_{\mathrm{OM}}(q) + 0.380H_{\mathrm{MM}}(q) \quad (7.1)$$

$$H_{\mathrm{XH}}(q) = 0.205H_{\mathrm{CH}}(q) + 0.179H_{\mathrm{OH}}(q) + 0.616H_{\mathrm{MH}}(q) + 0.0019H_{\mathrm{CM}}(q) + 0.0055H_{\mathrm{MM}}(q) + 0.0016H_{\mathrm{OM}}(q) \quad (7.2)$$

where M denotes a methyl hydrogen. Again, it can be seen that good agreement is obtained, although, for this system, the experimental data are somewhat noisier. Also shown in figure 5 are the OO, OH and CC RDFs together with those reduced from the experimental data. It can be seen that the QM/MM model gives somewhat better agreement with experiment for the OO and OH RDFs compared to the ultrasoft pseudopotential method, although, as was discussed in [34], there is a considerable spread in reported values of the location of the OO peak.

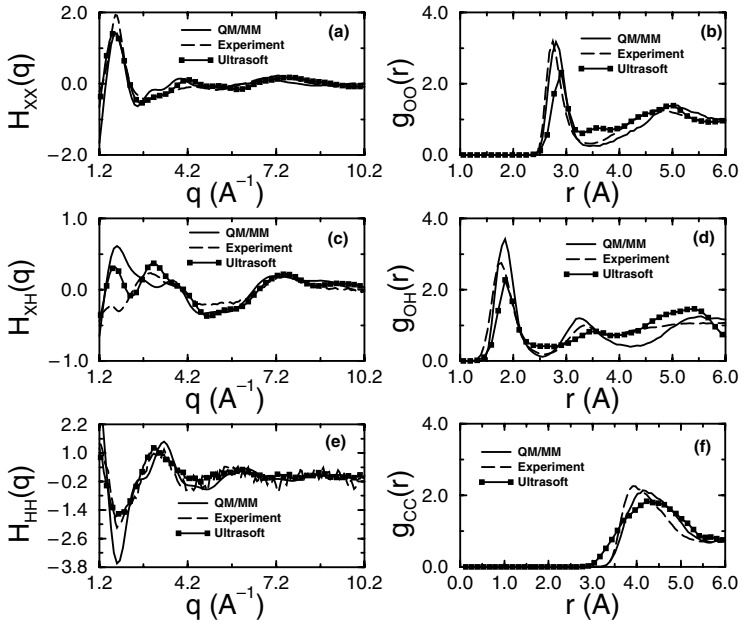

Figure 5. The neutron scattering partial structure factors and RDFs for liquid methanol at 300 K. Panels (a), (c), (e) compare, respectively, the computed XX, XH and HH partial structure factors (solid curves) as defined in equations (7.1) and (7.2) [34, 35] to those from experiment (dashed curves) [179]. The solid line with filled squares shows the partial structure factors from [34]. Panels (b), (d), (f) show the same comparison for the OO, OH and CC RDFs, respectively.

7.4. Structure and dynamics of high and low concentration KOD solutions

Understanding the behaviour of charged defects in aqueous environments created by the addition of protons (H^+) or hydroxide ions (OH^-) is of fundamental importance in the chemistry of acids and bases, in biological processes such as proton transport across lipid membranes and in industrially and technologically important applications such as fuel cell operation, saponification and industrial catalysis. Aqueous solutions containing hydronium and hydroxide ions are known to be excellent proton conductors; however, the microscopic details of the transport process remain controversial. Part of the controversy is due to the difficulty in obtaining definitive experimental evidence pointing to a particular mechanism. Thus, high level AIMD simulations can play a unique role in uncovering these mechanistic details while generating observables that can be compared directly with experiment. In this section, we will report AIMD studies of the IR spectrum of low (1.5 M) and high (13 M) concentration KOD solutions and then interpret these results in section 9.3, where we will discuss AIMD studies aimed at elucidating these mechanisms.

In the present study [42], the 1.5 M solution consisted of 32 water molecules and one dissociated KOD in a 10.25 Å cubic periodic box, while the 13 M solution consisted of 27 water molecules and eight dissociated KODs in a 10.15 Å cubic periodic box. KOD is chosen over KOH in order to reduce the importance of nuclear quantum effects associated with the light hydrogen atoms. The KS orbitals were expanded in a plane-wave basis up to a cut-off of 70 Ryd, and exchange and correlation were treated using the BLYP functional [103, 104]. Core electrons were represented using the Troullier–Martins pseudopotential [118] with a semi-core pseudopotential employed for potassium. Simulations of length 10 ps were carried

out using a time step of 5 au using a fictitious electron mass of 600 au. This simulation length includes 2 ps of equilibration. During production, the dipole moment was computed and the methodology of section 6 was used to compute the IR spectrum. Because the simulation length is relatively short, it is necessary to employ maximum entropy techniques [185] to extract the spectrum. In the present study, a 260-point Levinson algorithm [185] was employed. In order to check the sensitivity of the spectrum to the order of the algorithm, the spectrum was also computed using 220, 240 and 280 points and was found not to be sensitive to the order.

The IR spectra obtained are shown in figure 6 together with the measured spectrum of a 14 M KOH solution [186]. The figure shows that all of the features of the experimental spectrum at 14 M are reproduced. The shift in the frequencies is due to the use of deuterium rather than hydrogen in the present simulations, a fact that may also account for the increased sharpness of the peaks compared to experiment (although other influences such as finite-size effects and the short timescale may also contribute). An important difference between the high and low concentration spectra can be seen: the broad stretching peak in the low concentration spectrum splits into a distinct peak at approximately 2200 cm^{-1} and a shoulder at approximately 2400 cm^{-1}. These two features are assigned to the water and hydroxide OH stretching modes, respectively. In addition, the small peak at 1950 cm^{-1} in the 1.5 M solution is dramatically enhanced in the 13 M solution. This peak is assigned to the stretch of the OH bond of waters in the first solvation shell of hydroxide donated in hydrogen bonds to the hydroxide oxygen. The pronounced peaks in the 13 M spectrum are a consequence of the high coordination of the hydroxide ion in the solution [44, 45, 48, 187]. In figure 7, the O*O and O*H RDFs, where O* indicates the hydroxide oxygen, together with the integrated coordination numbers at both concentrations, are shown. The figure shows that, on average, the hydroxide is coordinated by approximately four water molecules. The specific complexes and role of these complexes in charged defect transport will be discussed in section 9.3. Figure 7 also shows the OO and OH RDFs and integrated coordination numbers of water molecules in both the 1.5 and 13 M KOD solutions. The most striking feature of these RDFs is the fact that, in the 13 M solution, there is a shift in the peak locations and a dramatic change in the average coordination. Specifically, the ordinary fourfold coordination is not present at 13 M but, rather, water molecules tend to be threefold coordinated on average. This change in the water structure will be seen to have an important effect on the transport mechanism of charged defects in basic solutions (see section 9.3).

8. Other applications

8.1. Solid-state proton conduction

Solid-state proton conductors are systems of considerable importance because of their potential uses in solid-state batteries, sensors and other electrochemical devices. For this reason, there is a growing interest in such systems. However, unlike in electrolyte solutions, where proton conduction occurs readily, solid-state proton conduction is a relatively rare process. Hence, proton conducting solids are of particular scientific interest because of their novelty. Indeed, a particular class of materials, proton conducting molecular crystals, have received considerable attention because of their wide range of structural and dynamical characteristics. Many such crystals can undergo temperature and/or pressure induced structural phase transitions, and, in some cases, thermal decomposition may occur with a concomitant release of large amount of energy. One material that has been studied extensively is the ammonium salt crystal, ammonium perchlorate (NH_4ClO_4).

Ammonium perchlorate (AP) is one of the showcase examples of a thermally labile material that can undergo a decomposition process at high temperature, leading to its use

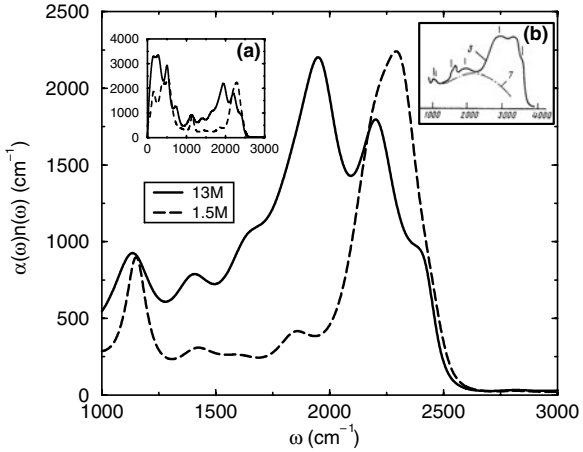

Figure 6. Computed IR absorption spectrum in aqueous KOD solution at concentrations of 13 M (solid curve) and 1.5 M (dashed curve) (cf equation (6.9)) for the frequency range 1000–3000 cm^{-1}. Inset (a) shows the spectrum over the full frequency range. Inset (b) shows the experimental spectrum for 14 M KOH solution from [42].

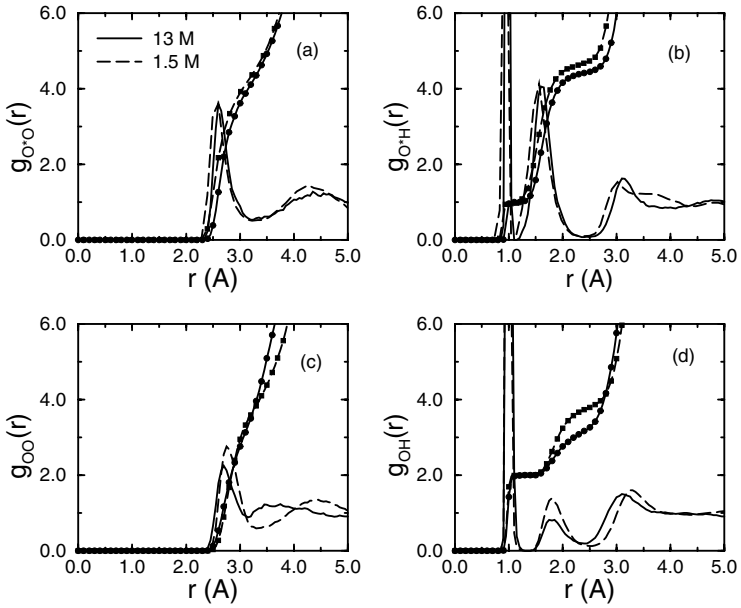

Figure 7. Panels (a)–(b)—RDFs with respect to the hydroxide oxygen O* for 13 M KOD solution (solid curves) and 1.5 M KOD solution (dashed curves) at 300 K. Panels (a) and (b) show the O*O and O*H RDFs, respectively. The integrated coordination numbers are shown in solid curves with filled circles for 13 M KOD solution and in dashed curves with filled squares for 1.5 M KOD solution. Panels (c)–(d)—Oxygen–oxygen (panel (c)) and oxygen–hydrogen (panel (d)) RDFs for bulk water in 13 M KOD solution (solid curves) and 1.5 M KOD solution (dashed curves) at 300 K. Also shown are the integrated coordination curves using the same convention as in panels (a) and (b).

in applications such as rocket propulsion. The crystal structure of AP depends on temperature: at temperatures below 513 K it possesses an orthorhombic unit cell, while at temperatures

above 513 K it undergoes a structural phase transition to a simple cubic structure. Particular attention has been devoted to the electrical conductivity of AP because of the suggestion that a charge-transfer step plays a fundamental role in the thermal decomposition process [188, 189].

Additional experiments on AP crystals [190] indicate that the electrical conductivity of AP increases markedly when it is exposed to an ammonia-rich atmosphere. Based on these findings, it was suggested that the increased conductivity is due to an anomalous mechanism in which the mobile NH_4^+ ions can transfer protons to the neutral ammonia molecules at lattice vacancies or at interstitial sites. The newly formed neutral ammonia molecules at lattice sites subsequently become recipients of excess protons from NH_4^+ ions now at interstitial sites. This mechanism is expected to dominate, for example, over the formation of Schottky or Frenkel defects [188].

In order to explore this proposed mechanism, we carried out AIMD simulations of AP crystal in the low temperature (orthorhombic) [191] and high temperature (cubic) [192] phases [52] and extracted the IR spectra using the methods of section 6. In both cases, run lengths of approximately 15 ps were carried out using the BLYP functional [103, 104] and a plane-wave basis with a cut-off of 70 Ryd in a supercell consisting of two unit cells (box lengths given by $L_x = 9.20$ Å, $L_y = 11.60$ Å, $L_z = 7.45$ Å, orthorhombic and $L_x = 15.26$ Å, $L_y = L_z = 7.63$ Å) at temperature of 300 and 530 K, respectively. In addition, we performed simulations in which neutral ammonia molecules were placed at interstitial sites up to a mole fraction of 0.5 (this system is denoted the AP·$(NH_3)_{0.5}$ system) and extracted the IR spectra from run lengths of approximately 27 ps for supercells consisting of two ($L_x = 15.26$ Å, $L_y = L_z = 7.63$ Å) and four ($L_x = L_y = 15.26$ Å, $L_z = 7.63$ Å) unit cells at a temperature of 530 K. Figure 8 shows the comparative IR spectra. Again, the spectra are obtained using the Levinson maximum entropy algorithm [185]. For the AP·$(NH_3)_{0.5}$ case, practically no difference is observed between the two- and four-unit-cell simulations (hence, only the latter is shown). The spectrum for the orthorhombic phase is generally in good agreement with the measured spectrum [193]. The most striking feature of the spectra is that the NH stretch band broadens in going from the orthorhombic to the cubic phases and extends into a lower frequency region in the AP·$(NH_3)_{0.5}$ system. Visual inspection of the trajectories in the AP·$(NH_3)_{0.5}$ case shows that this red shifting in the NH stretch band is due to the formation of $N_2H_7^+$ complexes in which excess protons are bound between two ammonia moieties, forming a relatively strong hydrogen bond. In these complexes, frequent proton transfer is observed. Indeed, recent AIMD studies of the $N_2H_7^+$ complex in the gas phase [33] confirm that proton transfer between the NH_3 and NH_4^+ moieties is possible, although the proton transfer free energy barrier is somewhat higher in the crystal ($\sim$1.8 kcal mol^{-1}). The computed IR spectra constitute clear experimentally verifiable evidence for the formation of such hydrogen-bonded complexes and, therefore, of the anomalous proton transfer mechanism. That the NH_3 and NH_4^+ units are observed to exhibit 'nearly free' rotation in all phases (in agreement with experimental observation [194]) suggests that in the AP·$(NH_3)_{0.5}$ system, rapid rotation of newly formed neutral NH_3 units allows the latter to pick up protons by rapidly rotating their lone pairs toward other NH_4^+ units in lattice and interstitial sites.

8.2. Ziegler–Natta catalysis

Ziegler–Natta polymerization of α-olefins is one of the most widely used industrial catalytic processes currently in use. In a typical preparation, a $MgCl_2(110)$ surface is used as a support for the deposition of $TiCl_4$. The mechanistic details of the polymerization process, which were until recently not precisely known, have been elucidated in a series of elegant AIMD calculations employing the BLYP functional [103, 104] and a plane-wave basis with a cut-off

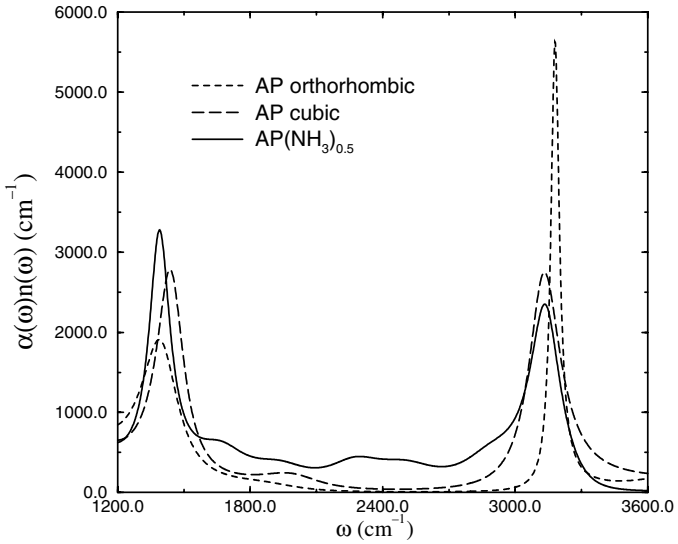

Figure 8. Computed IR absorption spectrum for an AP crystal in the orthorhombic phase (dashed curve), cubic phase (long dashed curve) and the cubic AP·$(NH_3)_{0.5}$ system (solid curve).

of 40 Ryd and nonlinear core corrections [195] for titanium of the formation of polyethylene on the surface by Parrinello and co-workers [76–78]. These studies suggested that fivefold coordinated Ti sites allow for relatively more facile insertion of ethylene than sixfold sites as had been previously assumed [196]. These fivefold sites were also shown to have the proper stereoselectivity needed to catalyse formation of polypropylene [77]. Stereoselectivity is one of the most important features of Ziegler–Natta catalysis. Finally, a series of studies of different $MgCl_2$ surface cuts was performed in order to explore how the morphology of the surface affects the mechanism of the polymerization process [78].

8.3. Influence of knots on polymer strength

An important question in polymer chemistry concerns the influence of knots on the strength of the polymer chain. Generally, any long polymer chain, whether it be polyethylene or DNA, is thought to contain some degree of knotting, usually in the form of simple trefoil or overhand knots. The presence of such knots is suspected to affect the strength of the chain, causing it to weaken and allow breaking at the entrance to the knot. In a recent series of AIMD based studies [84–87], the influence of knots and the mechanistic details of the chain rupturing process was explored using the BLYP functional [103, 104] and a plane-wave basis with a cut-off of 60 Ryd and explicit treatment of electron spin. These studies revealed that knots in a polyethylene chain do, indeed, weaken the chain and that rupturing occurs preferentially at the entrance of the knot [84, 85]. Moreover, it was shown [86, 87] that the radicals formed after the chain has broken are able to recombine to form cyclic alkanes and to undergo disproportionation processes with nearby chain segments. Finally, the influence of the environment, namely neighbouring chains, was studied [87], and was shown to make little difference to the fundamental findings obtained in the single-molecule simulations.

8.4. Molten silicates

Molten silicates are the precursors to industrially and technological important materials such as ceramics and nuclear waste confinement glasses. They also occur naturally in the form of geologic magmas. Despite their importance, the microscopic characteristics of silicate melts is not well understood because of the difficulty of studying the molten state experimentally. Recently, we carried out a series of AIMD simulations aimed at characterizing the structure of a calcium aluminosilicate (CAS) $CaO–Al_2O_3–SiO_2$ melt in comparison to molten SiO_2 [74]. The composition of the CAS system was taken to be $22SiO_2$, $4Al_2O_3$ and 7CaO because of its similarity to nuclear waste confinement materials. The system was studied using the BLYP functional [103, 104] using a plane-wave basis set with a cut-off of 70 Ryd. The introduction of the network modifier cation Ca^{2+} into the tetrahedral silica network has a dramatic effect on the physical properties of the system, including, for example, a significant reduction in the viscosity. This is due to a disruption of the network by formation of so-called *non-bridging* oxygens (NBOs), i.e. oxygens which do not bridge two tetrahedral SiO_4 or AlO_4^- units. In principle, the effect of the modifier cations can be compensated by the AlO_4^- tetrahedra, so that the number of NBOs should be given accurately by stoichiometry. Our calculations showed, however, that for a CAS melt, the number of NBOs exceeds the stoichiometric prediction as has been observed experimentally for CAS glasses [197, 198]. In addition, the calculations predict a strong violation of the aluminium avoidance principle, an empirical rule which states that in aluminosilicates, Al–O–Al linkages are not present [199]. Our finding is in accordance, however, with the experimental observation that the aluminium avoidance principle can be violated when Ca is the network modifier [200–202]. Finally, defect structures, in particular fivefold coordinated Al and oxygen tri-clusters (oxygens bonded to three network forming atoms), and their possible role in the creation of NBOs, were identified. Currently, we are performing quench simulations to obtain CAS glass structures and computing the NMR chemical shifts of these structures [203].

8.5. Systems of biological importance

Despite their high computational overhead, AIMD calculations are beginning to have an impact on problems of biological importance. Clearly, it is not possible to treat entire biological macromolecules in solution at a fully *ab initio* level, as these systems contain typically 10^4–10^5 atoms. However, with the identification of specific issues that may be explored via a small fragment [204] or model system [50] or through emerging techniques which combine *ab initio* level treatments for small, chemically active regions of a large molecule and an empirical force field for the remainder of the system [182, 184, 205–215] (so-called quantum mechanical/molecular mechanical or QM/MM calculations), important questions can be answered.

Recently, AIMD calculations were employed to address an important problem in the design of inhibitors for the HIV-1 protease [88] and the structure of the uncomplexed active site [216]. The HIV-1 protease plays a vital role in the replication cycle of the AIDS virus by cleaving a polyprotein encoded by the *gag* and *pol* genes into the structural proteins and enzymes of the virus. It is a relatively small homodimeric enzyme consisting of two covalently bonded 99-residue strands with an active site composed of two catalytic aspartic acid residues and has become an important target for AIDS therapies. The carboxylate oxygens assist in the attack of a water molecule on the scissile peptide bond that causes the required cleavage. Inhibition of this enzyme interrupts the replication cycle at a crucial stage, leading only to immature and noninfectious viral particles. An important question in the design of inhibitors for this enzyme

concerns the protonation state of the catalytic oxygens alone and in complexes with particular classes of drugs, as these serve as possible anchor points for hydrogen bonding of an inhibitor to the active site. Indeed, a design feature that is often employed in commercially available drugs as well as novel fullerene-derivative inhibitors [217–219] is an OH group that can form such hydrogen bonds. Recent ^{13}C NMR chemical shift measurements of a complex of the HIV-1 protease with pepstatin A has been interpreted as indicating a monoprotonated state for the active site because the signal appears to arise from chemically nonequivalent sites. However, the chemical inequivalence can also arise from interactions with nonsymmetric neighbouring groups. A combination of *ab initio* geometry optimizations and AIMD simulations over which the ^{13}C NMR chemical shift was averaged strongly suggest that the interpretation of a monoprotonated active site is not correct and that the actual protonation state is diprotonated, i.e. the Asp residue from each strand is actually monoprotonated. This finding not only has important implications for understanding the catalytic mechanism of the protease but also provides important information for the design of effective inhibitor compounds.

In other studies, hybrid QM/MM techniques employing DFT based AIMD with empirical force fields have been used to study catalytic mechanisms of various enzymes and to explore the design of possible biomimetics [90, 91] and structure, dynamics and binding mechanisms in myoglobin [92, 93]. Moreover, various schemes have been introduced for placing QM/MM type calculations on a rigorous footing [184, 215], and it is expected that as computing power increases, such methods will play a vital role in addressing numerous important questions of biological relevance.

9. The path integral Born–Oppenheimer approximation and *ab initio* path integral molecular dynamics

Up to this point, we have treated the nuclear motion within the framework of classical mechanics. Although this approximation is often acceptable in many situations, there will be numerous cases, particularly involving light nuclei such as hydrogen, where this approximation breaks down. In such cases, it is necessary to account for nuclear quantum effects at finite temperature. One of the most widely used methods for accomplishing this in large systems is within the Feynman path integral formulation of quantum statistical mechanics [12, 13]. This formulation allows the quantum equilibrium properties of a system to be computed within an AIMD scheme [14–16]. The procedure is relatively straightforward when spin statistics can be neglected. For bosonic exchange, stable MD based schemes have been proposed based on the calculation of the permanent of a matrix [220]. For fermionic exchange, although similar schemes exist based on the determinant, the inherent sign problem still remains an extremely challenging problem. In addition, the problem of computing quantum dynamical properties within the path integral formulation is also an immensely challenging problem, and, to date, no completely satisfactory procedure exists for effectively dealing with the rapidly oscillating phase factors that arise. However, effective approximate strategies have been proposed based on the use of the so-called 'path centroid' [17, 51, 221, 222] which can be employed when quantum effects are not strongly dominant or are not strongly anharmonic, the use of filter functions [223, 224], analytic continuation methods [225] or hierarchical Monte Carlo (MC) schemes [226].

The Born–Oppenheimer approximation described in section 2 can be recast in statistical mechanical language using the canonical ensemble, for which the partition function is given by

$$Q(N, V, T) = \mathrm{Tr}\exp[-\beta H] \tag{9.1}$$

where H is the Hamiltonian of equation (2.1). The trace in equation (9.1) can be performed in the coordinate basis, which leads to an imaginary time path integral expression of the form

$$Q(N,V,T) = \oint' \mathcal{D}^N \boldsymbol{R}(\tau) \oint' \mathcal{D}^{N_e} \boldsymbol{r}(\tau) \exp\left\{-\frac{1}{\hbar}\int_0^{\beta\hbar} \mathrm{d}\tau \left[T_N(\dot{\boldsymbol{R}}(\tau)) + T_e(\dot{\boldsymbol{r}}(\tau)) + V_{ee}(\boldsymbol{r}(\tau)) + V_{NN}(\boldsymbol{R}(\tau)) + V_{eN}(\boldsymbol{r}(\tau), \boldsymbol{R}(\tau))\right]\right\} \tag{9.2}$$

where the $3N \times 3N_e$-dimensional functional integral is taken over all cyclic paths in imaginary time, $\tau \in [0, \beta\hbar]$, satisfying

$$\boldsymbol{r}_i(0) = \boldsymbol{r}_i(\beta\hbar) \qquad \boldsymbol{R}_I(0) = \boldsymbol{R}_I(\beta\hbar) \qquad i = 1, \ldots, N_e,\ I = 1, \ldots, N \tag{9.3}$$

and all permutations of these paths as required by Bose–Einstein or Fermi–Dirac statistics (as indicated by the primes on the integrals) [10]. If the above path integral is treated using the influence functional approach [227], then equation (9.2) is written as

$$Q(N,V,T) = \oint' \mathcal{D}^N \boldsymbol{R}(\tau) \exp\left\{-\frac{1}{\hbar}\int_0^{\beta\hbar} \left[T_N(\dot{\boldsymbol{R}}(\tau)) + V_{NN}(\boldsymbol{R}(\tau))\right]\right\} F[\boldsymbol{R}(\tau)] \tag{9.4}$$

$$F[\boldsymbol{R}(\tau)] = \oint' \mathcal{D}^N \boldsymbol{R}(\tau) \exp\left\{-\frac{1}{\hbar}\int_0^{\beta\hbar} \left[T_e(\dot{\boldsymbol{r}}(\tau)) + V_{ee}(\boldsymbol{r}(\tau)) + V_{eN}(\boldsymbol{r}(\tau), \boldsymbol{R}(\tau))\right]\right\} \tag{9.5}$$

where $F[\boldsymbol{R}(\tau)]$ is known as the *influence functional*. Indeed, it is clear that $F[\boldsymbol{R}(\tau)]$ is a partition function for the electronic subsystem along a given nuclear path $\boldsymbol{R}(\tau)$. At a fixed nuclear configuration, $\boldsymbol{R}$, this partition function could be computed from the electronic eigenvalues, $\varepsilon_n(\boldsymbol{R})$, and would be related to the negative exponential of the free energy:

$$F(\boldsymbol{R}) = \sum_n \mathrm{e}^{-\beta\varepsilon_n(\boldsymbol{R})} = \mathrm{e}^{-\beta A(\boldsymbol{R})} \tag{9.6}$$

where $A(\boldsymbol{R})$ is the free energy at nuclear configuration $\boldsymbol{R}$. In the Born–Oppenheimer approximation, adiabaticity must be assumed along a nuclear path, i.e. at each τ, the electronic eigenvalue problem is solved for the specific configuration, $\boldsymbol{R}(\tau)$, at each imaginary time point, τ. This means that the influence functional, $F[\boldsymbol{R}(\tau)]$, and hence the free energy, must be local in τ. This leads to a convenient expression for the path integral:

$$Q(N,V,T) = \oint' \mathcal{D}^N \boldsymbol{R}(\tau) \exp\left\{-\frac{1}{\hbar}\int_0^{\beta\hbar} \left[T_N(\dot{\boldsymbol{R}}(\tau)) + V_{NN}(\boldsymbol{R}(\tau)) + A(\boldsymbol{R}(\tau))\right]\right\} \tag{9.7}$$

known as the free energy Born–Oppenheimer path integral approximation introduced by Cao and Berne [228].

On the other hand, if we simply considered the nuclear eigenvalue problem in equation (2.6), we would arrive at a path integral expression in which we perform a separate path integral on each electronic surface, $\varepsilon_n(\boldsymbol{R})$, and then sum over the surfaces, i.e.

$$Q(N,V,T) = \sum_n \oint' \mathcal{D}^N \boldsymbol{R}(\tau) \exp\left\{-\frac{1}{\hbar}\int_0^{\beta\hbar} \left[T_N(\dot{\boldsymbol{R}}(\tau)) + V_{NN}(\boldsymbol{R}(\tau)) + \varepsilon_n(\boldsymbol{R}(\tau))\right]\right\}. \tag{9.8}$$

Differences between these two path integral expressions are discussed in detail by Cao and Berne [228]. It is important to note that when only the electronic ground state is important, the two expressions are equivalent, since then

$$F[\boldsymbol{R}(\tau)] = \exp\left[-\frac{1}{\hbar}\int_0^{\beta\hbar} \mathrm{d}\tau\, \varepsilon_0(\boldsymbol{R}(\tau))\right] \tag{9.9}$$

and the path integral expression reduces to

$$Q(N,V,T) = \oint' \mathcal{D}^N \boldsymbol{R}(\tau) \exp\left\{-\frac{1}{\hbar}\int_0^{\beta\hbar} \left[T_N(\dot{\boldsymbol{R}}(\tau)) + V_{NN}(\boldsymbol{R}(\tau)) + \varepsilon_0(\boldsymbol{R}(\tau))\right]\right\} \tag{9.10}$$

which is the form on which we shall focus here.

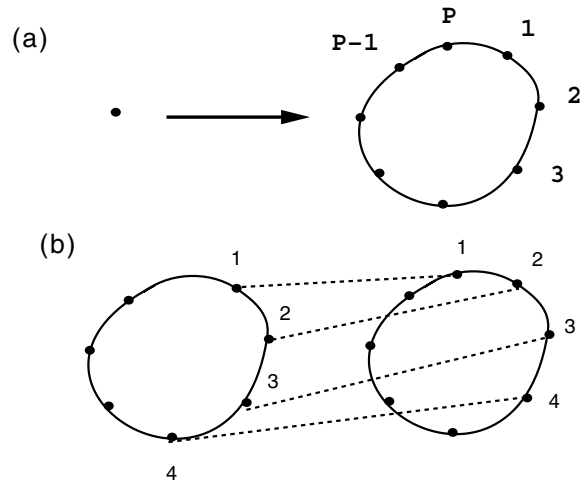

Figure 9. Illustration of the path integral representation of the quantum canonical partition functions. The representation of a quantum particle as a cyclic polymer chain in the discretized form of the path integral is shown in (a). The interaction between two quantum particles is shown in (b), where the illustration indicates that points on each path with the same imaginary time index *only* interact with each other.

In what follows, we shall neglect the spin statistics of the ions and only consider Boltzmann statistics. This approximation is well justified in many systems at ordinary temperatures. In this case, the Born–Oppenheimer path integral form in equation (9.10) can be written as a discrete path integral:

$$Q_P(N,V,T) = \left[\prod_{I=1}^{N}\left(\frac{M_I P}{2\pi\beta\hbar^2}\right)^{3P/2}\int \mathrm{d}\boldsymbol{R}_I^{(1)}\cdots\mathrm{d}\boldsymbol{R}_I^{(P)}\right] \times \exp\left\{-\beta\left[\sum_{s=1}^{P}\left(\sum_{I=1}^{N}\frac{1}{2}M_I\omega_P^2\left(\boldsymbol{R}_I^{(s+1)}-\boldsymbol{R}_I^{(s)}\right)^2+\frac{1}{P}E_0(\boldsymbol{R}_1^{(s)},\ldots,\boldsymbol{R}_N^{(s)})\right)\right]\right\} \tag{9.11}$$

where the trace condition, equation (9.3), requires that $\boldsymbol{R}_I^{(P+1)} = \boldsymbol{R}_I^{(1)}$. The true quantum partition function $Q(N,V,T) = \lim_{P\to\infty} Q_P(N,V,T)$ by virtue of the Trotter theorem.

Equation (9.11) is isomorphic to, for finite P, a system of N classical ring polymer chains with P points (see figure 9(a)) coupled through a nearest-neighbour harmonic interaction and interacting through a potential $(1/P)E_0(\boldsymbol{R}_1^{(s)},\ldots,\boldsymbol{R}_N^{(s)}) = \varepsilon_0(\boldsymbol{r}_1^{(s)},\ldots,\boldsymbol{R}_N^{(s)} + V_{\mathrm{NN}}(\boldsymbol{R}_1^{(s)},\ldots,\boldsymbol{R}_N^{(s)})$. The nature of the interaction is not equivalent to that which would be expected for a classical system, however, since only points on polymer chains with the same Trotter index, s, interact (see figure 9(b)). The partition function in equation (9.11) and its associated equilibrium averages can be evaluated via MD. However, as has been shown by Hall and Berne [229] and Tuckerman *et al* [230], a standard MD approach suffers from severe ergodicity problems. The origin of these problems lies in the presence of the nearest-neighbour harmonic coupling, the strength of which increases with P, which can be quite large compared to the E_0 interaction which is attenuated by a factor of $1/P$. This gives rise to a strong separation of timescales that causes MD trajectories to sample regions of phase space away from the invariant tori very slowly. Moreover, the harmonic interaction couples modes with a wide range of frequencies. The highest of these frequencies dictates the maximum time

step that can be used and causes the low frequency modes to be sampled inadequately. A solution to these problems was proposed by Tuckerman *et al* [230] who introduced a linear variable transformation from Cartesian coordinates to a set of mode coordinates and suggested coupling each mode degree of freedom to a Nosé–Hoover chain thermostat [142]. Two types of transformation were proposed. The first is a 'staging' transformation [16, 230] so named because it is derived from the staging MC algorithm [231] and takes the form

$$\begin{aligned} \boldsymbol{u}_I^{(1)} &= \boldsymbol{R}_I^{(1)} \\ \boldsymbol{u}_I^{(s)} &= \boldsymbol{R}_I^{(s)} - \boldsymbol{R}_I^{(s)^*} \qquad s = 2, \ldots, P \end{aligned} \tag{9.12}$$

where

$$\boldsymbol{R}_I^{(s)^*} = \frac{(s-1)\boldsymbol{R}_I^{(s+1)} + \boldsymbol{R}_I^{(1)}}{s}. \tag{9.13}$$

The inverse of this transformation is given by

$$\begin{aligned} \boldsymbol{R}_I^{(1)} &= \boldsymbol{u}_I^{(1)} \\ \boldsymbol{R}_I^{(s)} &= \boldsymbol{u}_I^{(1)} + \sum_{l=s}^{P} \frac{s-1}{l-1}\boldsymbol{u}_I^{(l)} \qquad l = 2, \ldots, P \end{aligned} \tag{9.14}$$

which can also be expressed recursively as

$$\begin{aligned} \boldsymbol{R}_I^{(1)} &= \boldsymbol{u}_I^{(1)} \\ \boldsymbol{R}_I^{(s)} &= \boldsymbol{u}_I^{(s)} + \frac{s-1}{s}\boldsymbol{R}_I^{(s+1)} + \frac{1}{s}\boldsymbol{R}_I^{(1)} \qquad s = 2, \ldots, P \end{aligned} \tag{9.15}$$

where, because of the condition $\boldsymbol{R}_I^{(P+1)} = \boldsymbol{R}_I^{(1)}$, the $s = P$ term is used to start the recursion. Substitution of this transformation into equation (9.11) gives

$$\begin{aligned} Q_P(N, V, T) = &\left[\prod_{I=1}^{N}\left(\frac{M_I P}{2\pi\beta\hbar^2}\right)^{3P/2} \int \mathrm{d}\boldsymbol{u}_I^{(1)} \cdots \mathrm{d}\boldsymbol{u}_I^{(P)}\right] \\ &\times \exp\left\{-\beta\left[\sum_{s=1}^{P}\left(\sum_{I=1}^{N}\frac{1}{2}M_I^{(s)}\omega_P^2\left(\boldsymbol{u}_I^{(s)}\right)^2 + \frac{1}{P}E_0(\boldsymbol{R}_1^{(s)}(\boldsymbol{u}), \ldots, \boldsymbol{R}_N^{(s)}(\boldsymbol{u}))\right)\right]\right\} \end{aligned} \tag{9.16}$$

where the masses are given by

$$\begin{aligned} M_I^{(1)} &= 0 \\ M_I^{(s)} &= \frac{s}{s-1}M_I \qquad s = 2, \ldots, P \end{aligned} \tag{9.17}$$

and where $\boldsymbol{R}_I^{(s)}(\boldsymbol{u})$ indicates that E_0 becomes a function of the staging variables via the transformation. A second possible transformation scheme is based on a simple normal mode decomposition of the path. Expanding the path in a Fourier series according to

$$\boldsymbol{R}_I^{(s)} = \sum_{k=1}^{P} \boldsymbol{a}_I^{(k)} \mathrm{e}^{2\pi\mathrm{i}(s-1)(k-1)/P} \tag{9.18}$$

where $\boldsymbol{a}_I^{(k)}$ are the expansion coefficients, a transformation may be introduced of the form

$$\begin{aligned} \boldsymbol{u}_I^{(1)} &= \boldsymbol{a}_I^{(1)} \\ \boldsymbol{u}_I^{(P)} &= \boldsymbol{a}_I^{(P+2)/2} \\ \boldsymbol{u}_I^{(2s-2)} &= \mathrm{Re}(\boldsymbol{a}_I^{(s)}) \\ \boldsymbol{u}_I^{(2s-1)} &= \mathrm{Im}(\boldsymbol{a}_I^{(s)}). \end{aligned} \tag{9.19}$$

Substitution of this transformation into equation (9.11) yields an expression similar to equation (9.16) with the masses given by

$$\begin{aligned} M_I^{(1)} &= 0 \\ M_I^{(s)} &= \lambda_s M_I \qquad s = 2, \ldots, P \end{aligned} \tag{9.20}$$

where

$$\lambda_{(2s-1)} = \lambda_{(2s-2)} = 2P\left\{1 - \cos\left[\frac{2\pi(s-1)}{P}\right]\right\} \tag{9.21}$$

are the normal mode frequencies.

Given the staging and normal mode transformations, a CP extended Lagrangian can be written down for AIPIs [16]:

$$L = \sum_{s=1}^{P}\left\{\frac{\mu}{P}\sum_i \langle\dot{\psi}_i^{(s)}|\dot{\psi}_i^{(s)}\rangle + \frac{1}{2}\sum_{I=1}^{N}\tilde{M}_I^{(s)}(\dot{\boldsymbol{u}}_I^{(s)})^2 - \frac{1}{2}\sum_{I=1}^{N}M_I^{(s)}\omega_P^2(\boldsymbol{u}_I^{(s)})^2 - E\big[\{\psi\}^{(s)}, \{\boldsymbol{R}(\boldsymbol{u})\}\big] + \sum_{i,j}\big[\Lambda_{ij}^{(s)}(\langle\psi_i^{(s)}|\psi_j^{(s)}\rangle - \delta_{ij})\big]\right\}. \tag{9.22}$$

At this point, a few comments are in order. First, note that both orbital velocities, $\{\dot{\psi}_i^{(s)}\}$, and mode velocities, $\{\dot{\boldsymbol{u}}_I^{(s)}\}$ have been introduced. This indicates that both the orbital and mode dynamics generated by equation (9.22) are fictitious. The former is used to generate the instantaneous forces and the latter to sample path configurations. Note, further, that the kinetic masses introduced for the modes, $\{\tilde{M}_I^{(s)}\}$, which are given by $\tilde{M}_I^{(1)} = M_1$, $\tilde{M}_I^{(s)} = M_I^{(s)}$ for $s = 2, \ldots, P$, have been chosen so that all staging or normal modes move on the same timescale. Since one is free to choose any set of kinetic masses for the fictitious dynamics, this particular choice ensures that slow and fast modes are sampled with equal efficiency. Finally, note that $\{\psi_i\}^{(s)}$ indicates that a complete set of KS orbitals is needed for each imaginary time slice s. That is, each time slice requires a separate electronic structure calculation. While this certainly makes the already costly computations P times more expensive, it should be recognized that these are completely independent calculations that can be handled easily in parallel. As stated earlier, the equations of motion derived from equation (9.22) need to be coupled to a canonical dynamics or thermostatting method. In addition, it is absolutely essential that each Cartesian component of each mode variable be coupled to its own thermostat [16, 230]. Detailed discussions of the equations of motion and integration algorithms for them can be found in [16].

As a final comment, it is well known that path integral computations are best performed in some type of mode variable (e.g. staging or normal modes) because, in these variables, paths are sampled more efficiently. For this reason, in the framework of AIPI, it is preferable to use MD over MC as a means of sampling path configurations. The reason for this is that, in MC, generally only a small subset of mode coordinates can be sampled in each trial move, which, nevertheless, changes the entire path configuration. It is, then, necessary to reminimize all P sets of KS orbitals in order to recompute the new potential energy. Since the electronic structure strongly dominates the computational time in AIPI calculations, this will lead to an efficient approach. By contrast, MD generates, in each step, moves of all the modes on all the cyclic polymer chains, which leads to an increase in efficiency, particularly when combined with the fictitious CP dynamics for the electrons.

9.1. Proton transfer in malonaldehyde

As discused in section 1, one of the main advantages of the AIMD technique is that it allows the study of chemical bond-breaking and forming events, for which reliable empirical potential

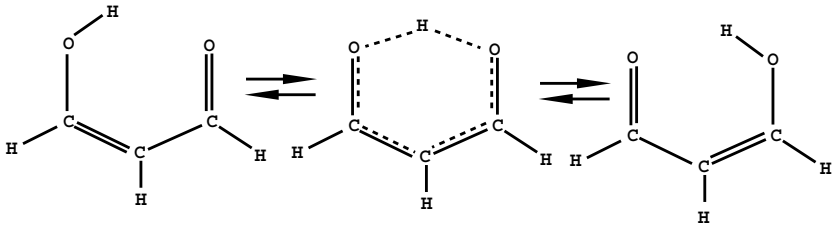

Figure 10. Illustration of the intramolecular proton transfer process in malonaldehyde. The C_s minimum with the proton on the left is transformed to a C_{2v} transition state and to an equivalent C_2 minimum with the proton localized on the other oxygen.

Table 3. Optimized C_s ground-state structure of malonaldehyde (see figure 10). The notation 70/15/pbc indicates a plane-wave cut-off of 70 Ryd, a box length of 15 Bohr, and periodic boundary conditions, with similar indications for the other columns (cbc indicates cluster boundary conditions).

Quantity	70/15/pbc	70/20/pbc	100/15/pbc	100/15/cbc	[234]	[234]	Expt [237]
r(O−H)	1.01	1.01	1.00	1.00	0.994	0.956	0.969
r(O···H)	1.73	1.72	1.73	1.73	1.694	1.908	1.68
r(O···O)	2.64	2.62	2.62	2.62	2.589	2.703	2.553
r(O−C)	1.34	1.34	1.34	1.34	1.328	1.326	1.320
r(O=C)	1.26	1.26	1.26	1.25	1.248	1.217	1.234
r(C=C)	1.37	1.37	1.37	1.37	1.362	1.353	1.348
r(C−C)	1.44	1.44	1.44	1.44	1.439	1.459	1.454
r(C−H_1)	1.08	1.08	1.08	1.08	1.083	1.071	1.089
r(C−H_2)	1.08	1.08	1.08	1.08	1.077	1.068	1.091
r(C−H_3)	1.10	1.10	1.10	1.10	1.098	1.088	1.094
∠(O−H···O)	147.4	147.4	147.0	146.4	—	139.0	147.6
∠(O−C=C)	123.9	124.0	123.8	123.9	124.5	126.1	124.5
∠(C=C−C)	120.8	120.4	120.4	120.5	119.5	121.3	119.4
∠(C−C=O)	124.0	123.7	123.8	123.5	123.5	123.8	123.0
∠(H−C=C)	122.9	122.8	123.0	123.0	122.5	121.4	122.3
∠(C=C−H)	119.7	119.6	119.8	119.7	120.0	119.4	128.1
∠(C−C−H)	117.1	117.4	117.2	117.4	117.6	116.7	117.6

models generally do not exist. As an illustration of the AIMD and AIPI methods, we investigate the role of nuclear quantum effects on a very common chemical reaction, the proton transfer reaction. Here, a simple example, proton transfer through the internal hydrogen bond in malonaldehyde ($C_3H_4O_2$), is explored [232]. The process is illustrated in figure 10. As the figure makes clear, the transfer of the proton between the two oxygens gives rise to a change in the chemical bonding pattern around the ring. At zero temperature, the barrier to proton transfer, as computed by high level *ab initio* calculations, ranges from 3.1 to 4.6 kcal mol^{-1}.

Our AIMD and AIPI simulations are based on the use of the Becke GGA exchange [103] and Perdew–Zunger [233] correlation functionals. The malonaldehyde molecule was placed in an 8.0 Å periodic box, and a plane-wave cut-off of 70 Ryd was employed. Core electrons were treated using the Troullier–Martins pseudopotentials [118]. This level of electronic structure theory gives a good description of the geometry of the molecule both in the C_s minimum and C_{2v} transition state configurations compared to quantum chemical calculations [234–236] and to experiment [237] as tables 3 and 4 make clear. The tables also show that raising the cut-off to 100 Ryd has little effect on the geometry. Moreover, the present DFT scheme

Table 4. Optimized C_{2v} transition-state structure of malonaldehyde (see figure 10). The notation is the same as in table 3.

Quantity	70/15/pbc	100/15/pbc	100/15/cbc	[235]	[235]
r(O···H)	1.20	1.20	1.21	1.197	1.203
r(O···O)	2.36	2.36	2.36	2.355	2.361
r(O−C)	1.30	1.30	1.30	1.277	1.285
r(C−C)	1.40	1.40	1.41	1.400	1.396
r(C−H)	1.09	1.09	1.10	—	—
∠(O···H···O)	158.0	157.8	157.5	—	—
∠(O−C−C)	121.3	121.0	120.9	—	—
∠(C−C−C)	116.4	116.5	116.4	—	—

gives a zero-temperature proton transfer barrier of 3.5 kcal mol^{-1}, in good agreement with both experimental estimates [238] and quantum chemical calculations [234–236]. A similar calculation at 100 Ryd gives nearly the same barrier.

The aim of the study is to evaluate the importance of nuclear quantum effects on the proton transfer free energy profile at 300 K using different levels of approximation to the nuclear motion. For path integral calculations, a discretization of $P = 16$ imaginary time slices was employed. In this study, three different types of calculation were performed:

(1) all nuclei treated as classical point particles;

(2) all nuclei treated as quantum particles;

(3) quantization of *only* the transferring proton (called *classical skeleton* calculations).

In each case, thermodynamic integration in conjunction with the bluemoon ensemble approach [239, 240] is used to obtain the proton transfer free energy profile at 300 K.

Figure 11 shows the free energy profiles thus obtained for each of the three simulation types. The reaction coordinate on the x-axis is the asymmetric stretch $\delta = d_{O_1H} - d_{O_2H}$, i.e. the difference between the distances of each oxygen to the shared proton. In the quantum profiles, the centroid of the reaction coordinate, $\delta^{(c)} = (1/P)\sum_{i=1}^{P} \delta_i$, is used. One of the striking features of these profiles is that quantum tunnelling of the proton is seen to be important even at 300 K. The profiles also show that inclusion of only thermal fluctuations, via the classical nucleus simulation, gives rise to very little difference between the free energy and zero-temperature barriers. When the transferring proton only is quantized, the barrier is considerably lower (approximately 2.1 kcal mol^{-1}). Finally, when all nuclei are properly quantized, the barrier is further lowered to approximately 1.6 kcal mol^{-1}. This implies that there is a nontrivial quantum effect due to the heavy-atom skeleton. Failure to include this quantum effect leads to an overestimation of the free energy barrier of 31%. (and, hence, underestimation of the proton transfer rate (in transition state theory) by a factor of roughly two. Other recent AIPI calculations of the free energy profile [241–243] suggest that there could be some sensitivity to the choice of exchange–correlation functional. We note, moreover, that our classical skeleton approximation is a comparatively mild one in comparison to approximations that are commonly made in the modelling of such processes. The latter usually completely disregard the structure of the heavy-atom skeleton or attempt to reduce the dimensionality of the problem to a few relevant degrees of freedom. Therefore, the classical skeleton approximation most likely leads to a lower bound estimate of the amount by which more severe approximations would tend to overestimate the free energy barrier and underestimate the rate.

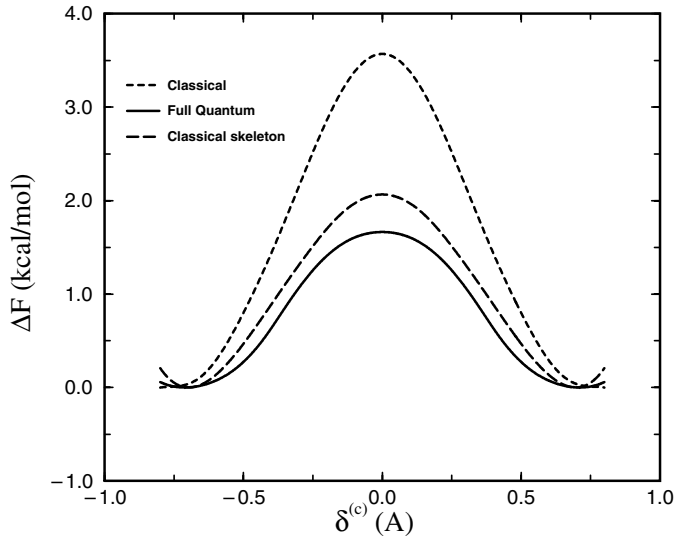

Figure 11. Free energy profiles for intramolecular proton transfer in malonaldehyde at 300 K. The fully classical profile is shown with the dashed curve, the fully quantum profile is shown in the solid curve and the profile in which only the transferring proton is quantized is shown with the long dashed curve.

9.2. Proton transport in water

Aqueous proton transport is a fundamentally important process in the chemistry of acids and bases and in many biologically important systems. In water, protonic defects (hydronium, H_3O^+, and hydroxide, OH^-, ions) have an anomalously high mobility that cannot be explained by an ordinary hydrodynamic diffusion picture. In fact, the commonly accepted mobility mechanism is the so-called 'structural diffusion' or 'Grotthuss' mechanism, in which solvation structures diffuse through the hydrogen-bond network via sequential proton transfer reactions. However, the microscopic details of the Grotthuss mechanism for different situations, were, until recently, largely unelucidated. Here, we describe AIMD and AIPI simulations [44–46, 48, 244] that have led to a clear picture of the structural diffusion mechanism of the hydronium ion in water, a picture that has since been shown to be consistent with all available experimental data [245]. In addition, one of the key controversial issues, concerning the dominant solvation structures, is resolved. Briefly, one school of thought, put forth by Eigen [246], considers the dominant solvation structure to be that of a H_3O^+ core surrounded by three water molecules consisting of a $H_9O_4^+$ cation. The other school of thought, due to Zundel [247], favours a picture in which the dominant structure consists of a protonated water dimer or $H_5O_2^+$ cation, in which the proton is equally shared between two water molecules.

The simulation protocol consists of 31 water molecules and one hydronium ion in a 10 Å periodic box. Exchange and correlation are, again, treated using the BLYP functional, and a plane-wave basis set truncated at a cut-off of 70 Ryd was employed. Core electrons were treated using the Troullier–Martins pseudopotentials. For path integral simulations, a discretization of eight imaginary time slices was employed. AIMD and AIPI trajectories of length 20 ps using a time step of 0.17 fs were generated.

Figure 12 shows schematically the structural diffusion mechanism that is uncovered in these simulations. Similar 'cartoons' of the proton transfer process can also be found in recent

articles by Agmon [245, 248] (see, for example, figure 1 in [248]) as well as by Tuckerman *et al* [44, 48]. As the figure shows, the process involves the breaking of a hydrogen bond between the first and second solvation shell members of H_3O^+, i.e., a second solvation shell fluctuation [16, 249]. Following this hydrogen-bond breaking event, a first solvation shell water is left in a state in which its coordination is three instead of the usual average value of four for water. In this state, it is coordinated like hydronium rather than like water, and it is, therefore, 'prepared' to become a properly solvated hydronium via proton transfer. When its coordination number changes, the oxygen–oxygen distance between the hydronium and the undercoordinated water shrinks by approximately 0.1 Å, and the proton moves to the middle of the bond, forming an intermediate $H_5O_2^+$ cation state. The proton can then either return to the original hydronium or continue to cross the hydrogen bond to the new oxygen site. If the latter occurs, then there is a new $H_9O_4^+$ cation formed with a new hydronium core. Thus, the solvation structure has migrated through the hydrogen bond network via the proton transfer step. There may also be several hops before resolvation of the undercoordinated first solvation shell water occurs [44]. The rate-limiting process is the hydrogen bond-breaking event, which requires approximately 1.5 ps to occur. This number is in good agreement with the experimentally determined rate of structural diffusion from NMR measurements [250]. In addition, the activation enthalpy, approximately 3 kcal mol^{-1}, can be explained by this mechanism, which requires approximately 2.5 kcal mol^{-1} to break the hydrogen bond and another 0.5 kcal mol^{-1} to shrink the oxygen–oxygen distance after the hydrogen bond is broken [251].

Inclusion of nuclear quantum effects via the path integral can resolve the controversy of the solvation structures [46, 47]. In particular, if we plot the probability distribution of the oxygen–oxygen distance, R_{OO}, and the proton transfer coordinate, δ, for the hydrogen bond in which proton transfer is 'most likely to occur' (defined as the hydrogen bond with the smallest value of δ, equivalent to the 'special bond' in [44, 45]), which is shown in the top panel of figure 13, we see that the probability that the solvation complex is $H_9O_4^+$ or $H_5O_2^+$ or any complex in between these two ideal, limiting structures is approximately the same. This is also confirmed by studying the free energy profile along the coordinate δ also shown in figure 13 (bottom panel). The fact that there is a broad flat minimum in this free energy confirms the notion that there is no single dominant solvation structure. Rather, the defect is best described as a 'fluxional' defect, that can take on, with the same probability, the characteristics of the $H_9O_4^+$ or $H_5O_2^+$ cations and all structures in between these. Interestingly, a purely classical treatment predicts that the $H_9O_4^+$ is considerably more stable than the $H_5O_2^+$ cation by approximately 0.6 kcal mol^{-1}. This proton transfer barrier is completely washed out by nuclear zero-point motion, leading to the fluxional defect picture proposed in [46, 47].

9.3. Hydroxide transport in water

Like the hydronium ion, the hydroxide ion is known to have an anomalously high mobility in water. For at least a century, it has been thought that the hydroxide ion can be viewed as a water molecule with a missing proton (or a 'proton hole') and, therefore, that the mechanism of hydroxide transport could be deduced from the hydronium transport mechanism as a kind of 'mirror image' in which all of the hydrogen bond polarities are reversed [252]. In this proton hole picture, the analogue of the Eigen cation would be an anionic complex in which the hydroxide is coordinated by three water molecules (one for each lone pair on the oxygen), i.e. an $H_7O_4^-$ structure, while the analogue of the Zundel cation would be an anionic complex in which the excess proton is equally shared between two OH^- moieties, i.e. an $H_3O_2^-$ structure. In this picture, the $H_3O_2^-$ structure is the most stable complex, which offers a possible explanation for the lack of a pronounced peak at the hydroxide OH stretch frequency in the IR spectrum

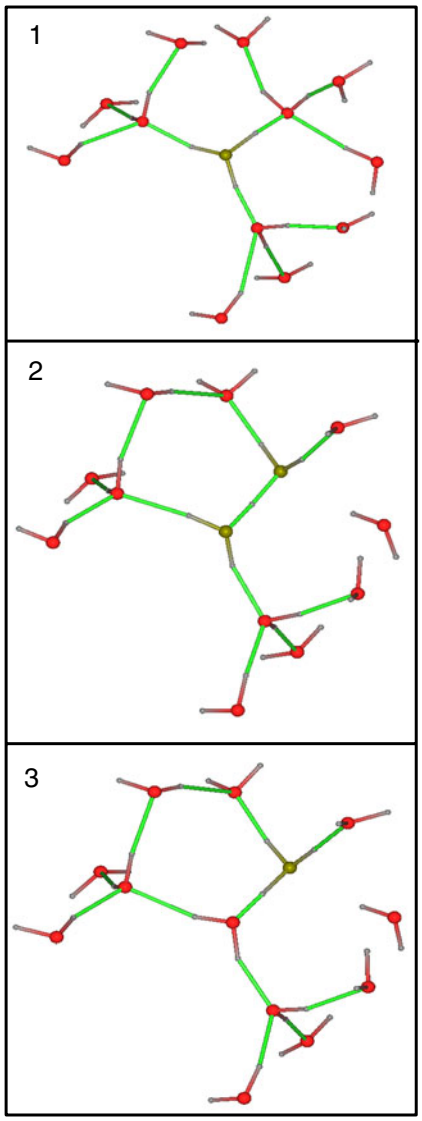

Figure 12. Illustration of the Grotthuss proton transfer mechanism in acidic solution. Panel 1 shows the hydronium with its first and second solvation shells taken directly from the simulation study of [44, 45]. Oxygens are shown as red spheres, hydrogens as grey spheres and hydrogen bonds are shown as green lines. The yellow oxygen is the centre of the hydronium. Panel 2 shows the breaking of a hydrogen bond between first and second solvation shell members, leaving a first solvation shell water undercoordinated. This event is followed by the migration of the proton to the centre of the hydrogen bond, forming an intermediate Zundel cation complex. Panel 3 shows the proton transfer to the undercoordinated first solvation shell water, illustrating the process of structural diffusion.

of concentrated KOH and NaOH solutions [186]: the symmetry of the $H_3O_2^-$ complex is such that the two OH moieties point in opposite directions, hence, the changes in their respective dipole moments upon excitation tend to cancel out [253].

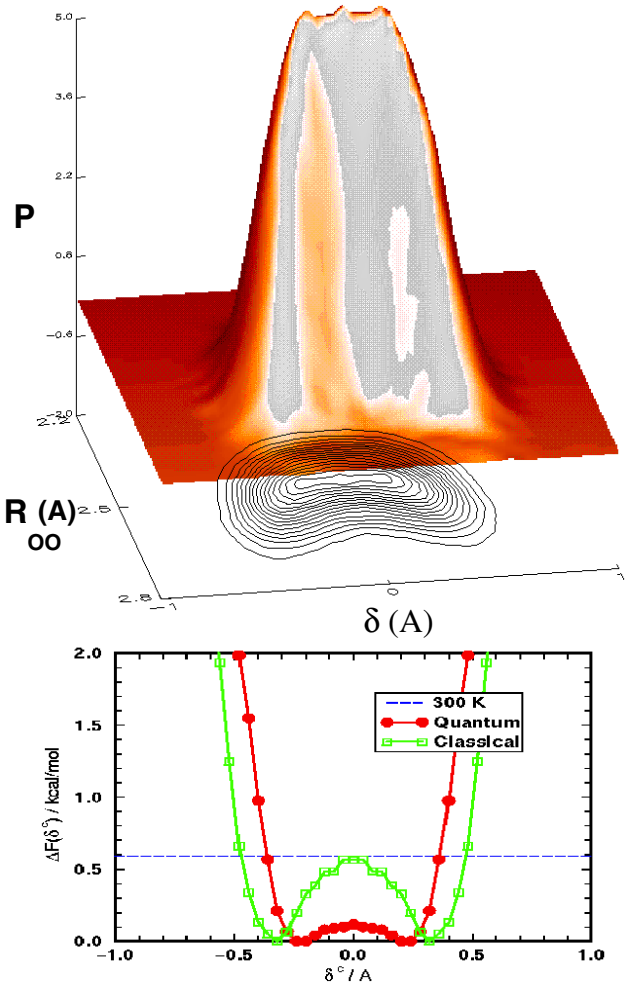

Figure 13. Top: the quantum mechanical two-dimensional probability distribution function $P(R_{OO}, \delta)$ for an excess proton in water at 300 K. Bottom: free energy profiles along the reaction coordinate, δ. The curve with filled circles shows the fully quantum free energy profile, while the curve with hollow squares shows the profile obtained if all nuclei are treated classically. The dashed line is the thermal energy kT in kcal mol^{-1}.

However, the high coordination number of the hydroxide ion in solution as revealed in figure 7 and as estimated in numerous experimental studies (see [187] and references therein) has called into question the validity of the proton hole picture. Here, we employ AIPI simulations of a hydroxide ion in a bath of 32 water molecules to identify the dominant solvation complexes of the hydroxide ion in water and to propose a new transport mechanism. AIPI simulations using the same protocol as that used for the hydronium in water were carried out generating trajectories of length approximately 30 ps. In order to identify the dominant solvation complexes, we examine the O*O and O*H RDFs, where O* denotes the hydroxide oxygen, in two regions of the proton transfer coordinate, δ, in particular, $\delta < 0.1$ Å and $\delta > 0.5$ Å. Again, δ denotes the smallest asymmetric stretch coordinate over all hydrogen bonds in the system. The RDFs are shown in figure 14, together with the integrated coordination

numbers. The figure clearly shows that for $\delta > 0.5$ Å the O*O and O*H coordination numbers of the hydroxide oxygen are four and five, respectively. These numbers indicate a fourfold coordinated ($H_9O_5^-$) complex, in which the hydroxide ion accepts four hydrogen bonds from neighbouring waters. (The additional hydrogen in the O*H coordination number is the hydroxide ion's own covalently bonded hydrogen.) At $\delta < 0.1$ Å, the coordination numbers are reduced by one, indicating that as the proton is transferred, the dominant solvation complex is a threefold coordinated ($H_7O_4^-$) complex, in which the $H_9O_5^-$ complex loses a first solvation shell hydrogen bond, leaving the hydroxide oxygen accepting just three hydrogen bonds. Importantly, however, there is a second element that appears in the first solvation shell for $\delta < 0.1$ Å. This can be gleaned by plotting the H′O RDF, where H′ is the hydroxide hydrogen. This is plotted in figure 14 as a function of δ and shows that as $\delta \to 0$, a definite peak at $r = 2.0$ Å emerges, indicating that a weak hydrogen bond between the hydroxide hydrogen and a nearby water is formed. The role of this hydrogen bond in the proton-transfer process is clear: before the hydroxide ion can accept the proton from a first solvation shell water, it must be coordinated like a water molecule. In the $H_7O_4^-$ complex state, it accepts three hydrogen bonds. The role of the H′O hydrogen bond is to provide a fourth member of the solvation shell to which the hydroxide ion donates a hydrogen bond. After completion of the proton transfer process, the newly formed water molecule will then accept two hydrogen bonds and donate two, which is the correct coordination structure for a water molecule. Given these results, the complete proposed transport mechanism is illustrated in figure 15. The isosurface in the figure is the so-called electron localization function (ELF) [254], which indicates regions of high spatial probability for finding electrons. The ELF shows that the three lone pairs on the hydroxide ion tend to be highly delocalized, forming a continuous ring perpendicular to the O*H′ bond axis. The lack of any distinct attractive basins in the ELF help to explain why the hydroxide ion can have such a high coordination number in water.

A study of the free energy profile along the reaction coordinate, δ (not shown), indicates that, in the present case, a small barrier of approximately 0.3 kcal mol^{-1} is obtained. When this is combined with a cost of approximately 1.2 kcal mol^{-1} to convert the $H_9O_5^-$ complex to the $H_7O_4^-$ complex [255] and approximately 1.5 kcal mol^{-1} to relax the $H_7O_4^-$ structure [251] after a first solvation shell hydrogen bond breaking, a total activation enthalpy of roughly 3 kcal mol^{-1} is obtained which agrees well with the experimental value [253]. This value also indicates that the $H_3O_2^-$ complex is not the most stable complex in solution but emerges in the present picture as a clear low-barrier transition state, which is in contrast to the proton hole mechanism, in which $H_3O_2^-$ is assumed to be the most stable complex in solution.

Finally, it is worth noting that high coordination numbers for anionic defects in other hydrogen-bonded liquids have been obtained using AIMD simulations. Recently, Liu and Tuckerman [33] studied the solvation structure of the amide ion in ammonia using the BLYP functional [103, 104] and a plane-wave basis set with a cut-off of 70 Ryd. It was found that the solvation shell has a similar planar structure to OH^- in water with a total coordination number of approximately 7.5, of which roughly 40% is due to hydrogen-bonding. Moreover, the solvation shell is highly dynamic like that of OH^- in water and in contrast to H_3O^+ in water or NH_4^+ in ammonia [33]. It is expected that through studies of other hydrogen-bonded liquids, reproducible patterns such as these may emerge.

10. Beyond the ground state

Extending the AIMD technique beyond the ground state is a challenging problem. While it is straightforward to treat excited states in wavefunction based methods such as Hartree–Fock theory, DFT based approaches have not yet become routine, although there are many promising

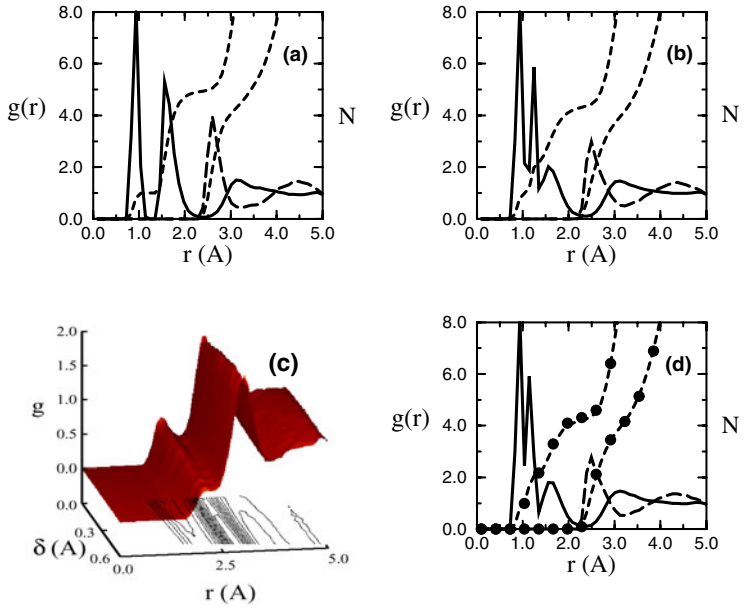

Figure 14. RDFs, $g(r)$, and running coordination numbers, N, with respect to the hydroxide ion and a first solvation shell water at different stages in the proton transfer process. (a) RDFs, g_{O^*O} (solid curve) and g_{O^*H} (long dashed curve), and coordination numbers, N_{O^*O} and N_{O^*H} (dashed curves), with respect to the OH^-, O*, for values of the proton transfer coordinate, $|\delta| > 0.5$ Å. (b) The same as (a) except for $|\delta| < 0.1$ Å. (c) Two-dimensional generalization of a hydrogen–oxygen RDF $g_{H'O}(r, \delta)$ with respect to the OH^- hydrogen, H′, as a function of r and δ. (d) RDFs and coordination numbers with respect to $\bar{O}$ for $|\delta| < 0.1$ Å. Here, $\bar{O}$ belongs to the first solvation shell water that transfers its proton to the hydroxide. The line scheme is the same as in (a) and (b). The filled circles are the coordination numbers from (b).

developments along these lines, and it is likely that such methods will soon become a standard feature of state-of-the-art packages (see section 12). Electronic excitations are clearly important in systems with many closely spaced electronic levels such as metals with collective excitations and systems at temperatures higher than the Fermi temperature. They are also important in many classes of reactions where transitions to a single excited state surface can occur.

The case of many closely spaced electronic levels can be treated by assuming a thermal population of electronic levels and treating the problem within a free energy formulation [256]. In a noninteracting electron theory, described by a single-particle Hamiltonian of the form

$$h = -\tfrac{1}{2}\nabla^2 + V(\boldsymbol{r}, \boldsymbol{R}) \tag{10.1}$$

such a formulation is achieved within the grand canonical ensemble (GCE). In the GCE, the temperature, T, volume, V, and chemical potential, μ, are the thermodynamic control variables, and the partition function, including the Pauli exclusion principle, is given by

$$\mathcal{Q}(\mu, V, T) = \left\{ \prod_i [1 + \exp(-\beta(\varepsilon_i - \mu)] \right\}^2 \tag{10.2}$$

where ε_i are the single-particle energy eigenvalues, $\beta = 1/kT$ and the power of two is due to the spin multiplicity. The free energy in the GCE is, then, given by

$$\mathcal{F}(\mu, V, T) = -PV = -kT \ln \mathcal{Q}(\mu, V, T) = -2kT \sum_i \ln\{1 + [\exp(-\beta(\varepsilon_i - \mu)]\} \tag{10.3}$$

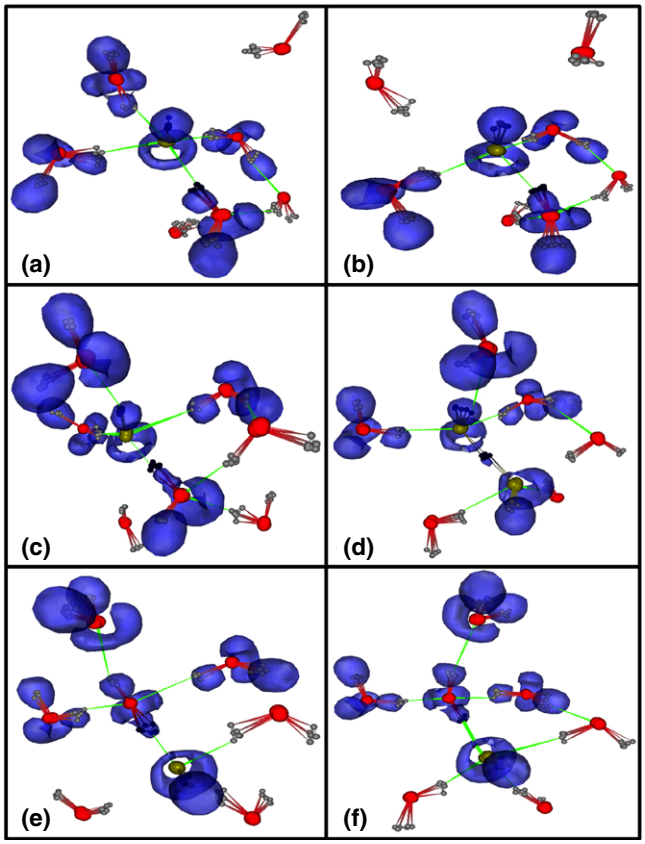

Figure 15. Representative configurations showing the proton transfer mechanism. Each panel shows quantum mechanical particle density snapshots of the OH^- together with relevant first and second solvation shell water molecules. Oxygen and hydrogen atoms are shown as red and grey spheres, respectively. The defect oxygen, O*, and shared proton in the most active H-bond are shaded yellow and black, respectively. H-bonds are schematically shown as green lines. In blue are shown the $\eta = 0.93$ isosurfaces of the ELF [254] of the path centroid configuration for the OH^- and only those water molecules which are H-bonded to the OH^-. (a) The OH^- is in its inactive $H_9O_5^-$ structure with fourfold planar coordination around O*. The ELF shows that the *three* lone pairs around O* are in a delocalized ring structure, thus supporting the hypercoordination of the OH^- (see text). An additional water molecule, which participates in the activation process ((c), (d)), appears in the upper right corner. (b) A first solvation shell H-bond breaking event occurs which transforms the $H_9O_5^-$ structure into an approximately tetrahedral $H_7O_4^-$ structure. (c) A weak H-bond between the OH^- hydrogen, H′, and the bulk water molecule from panels (a) and (b) is formed. As shown, when this H-bond forms, one of the shared protons between O* and a coordinating water begins to transfer. (d) As the proton is transferred, the environments around each OH moiety become similar (see, also, figure 14(d)). (Note that a complete solvation shell for the water that transfers its proton to O* is not shown. However, the coordination number around this water is, nevertheless, 4.0—see figure 14(d)). The ELF shows two distorted rings around the yellow oxygens, emphasizing the symmetry of this configuration. (e) After the proton transfer, the original OH^- has been transformed into a properly solvated water molecule with two lone pairs, whereas the newly formed OH^- (now shown in yellow) possesses the ring-shaped ELF. (f) The process is completed by the acceptance of a fourth H-bond from the newly formed O*, which 'closes the gate' by forming a new, inactive $H_9O_5^-$ complex with a fully intact ELF ring as in panel (a).

where P is the pressure. In the GCE, the average particle number is given by

$$N = -\left(\frac{\partial \mathcal{F}}{\partial \mu}\right)_{V,T}. \tag{10.4}$$

Therefore, it is possible to perform a Legendre transform from the GCE to the canonical ensemble, where N is fixed, and obtain the Helmholtz free energy by

$$A(N, V, T) = \mathcal{F}(\mu(N), V, T) + \mu(N)N \tag{10.5}$$

where $\mu(N)$ is obtained by equation (10.4) for μ as a function of N. In this way, the total number of electrons can be fixed.

Since KS theory is a theory for noninteracting electrons in an effective potential, $V_{\rm KS}(\boldsymbol{r}, \boldsymbol{R})$, the free energy formulation can be generalized to include electron interactions via the KS formalism. At finite temperature, the KS Hamiltonian is given by

$$H_{\rm KS} = -\frac{1}{2}\nabla^2 + V_{\rm ext}(\boldsymbol{r}, \boldsymbol{R}) + V_{\rm H}(\boldsymbol{r}) + \frac{\delta A_{\rm xc}}{\delta n(\boldsymbol{r})} + U(\boldsymbol{R}) \tag{10.6}$$

where $A_{\rm xc}[n]$ is the finite temperature exchange–correlation functional and $V_{\rm H}(\boldsymbol{r}) = \int \mathrm{d}\boldsymbol{r}'\, n(\boldsymbol{r}')/|\boldsymbol{r} - \boldsymbol{r}'|$. However, caution is needed because the sum of the KS eigenvalues does not yield the total energy of the system [95, 96]. It can be shown that the correct formulation of the Helmholtz free energy functional is

$$\begin{aligned} A(N, V, T) = \mathcal{F}(\mu(N), V, T) + \mu(N) \int \mathrm{d}\boldsymbol{r}\, n(\boldsymbol{r}) + U(\boldsymbol{R}) - \tfrac{1}{2} \int \mathrm{d}\boldsymbol{r}\, V_{\rm H}(\boldsymbol{r}) n(\boldsymbol{r}) \\ + A_{\rm xc}[n] - \int \mathrm{d}\boldsymbol{r}\, \frac{\delta A_{\rm xc}[n]}{\delta n(\boldsymbol{r})} n(\boldsymbol{r}) \end{aligned} \tag{10.7}$$

where the density is computed from

$$\begin{aligned} n(\boldsymbol{r}) &= \sum_i f_i(T) |\psi_i(\boldsymbol{r})|^2 \\ f_i(T) &= \frac{f_i(0)}{1 + \exp[-\beta(\varepsilon_i^{\rm KS} - \mu(N))]}. \end{aligned} \tag{10.8}$$

Here $\varepsilon_i^{\rm KS}$ are the KS eigenvalues and $f_i(T)$ are the usual Fermi–Dirac occupation numbers in the GCE. This free energy functional can, then, be used in an AIMD simulation with forces derived from $\boldsymbol{F}_I = -\nabla_I A(N, V, T)$. Note that the free energy formulation reduces to the correct ground-state KS formulation when $T \to 0$. It is important to point out, however, that the self-consistent density corresponding to the exact free energy functional of Mermin [257] is not an extremum of equation (10.7) but rather a saddle point. Therefore, AIMD simulations cannot be easily performed using an extended Lagrangian based approach but can employ a direct diagonalization approach as a viable alternative.

The case of electronic transitions to a single excited state surface can, in principle, be treated using non-adiabatic self-consistent formulations [258, 259] based on time-dependent density functional theory (TDDFT) [260, 261]. The latter has been used within a linear-response formalism to compute vertical excitation spectra with some degree of success [262]. The computational overhead of such an approach is rather formidable, however, so that only small systems have been treated with this approach thus far [259]. However, it is possible to develop approximate analytical expressions for excited singlet states by introducing, for these surfaces, approximate excited state exchange–correlation functionals [263] and combining the approach with a surface-hopping algorithm [264, 265]. An appropriate CP type scheme for this approach was recently introduced by Doltsinis and Marx [266] and was applied to the *cis–trans* photoisomerization of the Schiff base, formaldimine.

11. Linear scaling methods

In order to complete our survey of the AIMD methodology, a few words on the subject of linear scaling methods is in order. A more detailed review of this subject can be found in the recent article by Bowler *et al* [267]. One of the most severe limitations to the size of system that can be treated by DFT based AIMD methods is the fact that KS DFT scales as N^2M where N is the number of electronic states and M is the number of basis functions used to represent the KS orbitals. The reason for this can be understood by considering the computation of the overlap matrix $\langle\psi_i|\psi_j\rangle$, which requires $N(N+1)/2$ scalar products each of length M. However, physics and chemistry both tell us that the properties of matter are local. Thus, we do not expect the electron distribution in a covalent bond of an adsorbate on a surface to be strongly influenced by distant atoms in the bulk. If the orbitals are truly local, then we would expect that a formulation of DFT that scales linearly with the number of atoms should be possible.

Perhaps the most natural approach is to use DFT in its original Hohenberg–Kohn formulation, which involves only the density [102]. Such an approach [268] manifestly scales linearly with the number of basis functions used to expand the density. The main difficulty with this approach is that the kinetic energy functional becomes an additional unknown that needs to be approximated [268–271]. Moreover, the option to use nonlocal pseudopotentials in place of core electrons is no longer available [272]. For this reason, such 'density only' or 'orbital free' methods have largely been limited in their scope of applications.

First, if the orbitals are well localized, then expanding them in a *localized* basis set such as a Gaussian [100], DVR [101] or wavelet [273] basis sets means that the overlap matrix $\langle\psi_i|\psi_j\rangle$ only scales as N^2 since not all of the basis functions will be needed for each orbital. Moreover, for localized orbitals, overlaps between orbitals in distant parts of a large system will be zero so the overlap matrix should be relatively sparse. This means that the number of nonzero elements will not increase as the system size increases so that the calculation of the matrix really scales linearly with the size of the system. A more quantitative statement of this fact can be achieved via the one-particle density matrix given by

$$\rho(\boldsymbol{r},\boldsymbol{r}') = \sum_i f_i \psi_i(\boldsymbol{r})\psi_i^*(\boldsymbol{r}') \tag{11.1}$$

where $\rho(\boldsymbol{r},\boldsymbol{r}) = n(\boldsymbol{r})$. The physical behaviour of the density matrix is such that as $|\boldsymbol{r}-\boldsymbol{r}'| \to \infty$, $\rho(\boldsymbol{r},\boldsymbol{r}')$ tends to zero exponentially for insulators and algebraically for metals. Therefore, linear scaling can be achieved by introducing the approximation that $\rho(\boldsymbol{r},\boldsymbol{r}') = 0$ for $|\boldsymbol{r}-\boldsymbol{r}'| > r_{\text{cut}}$ for some cut-off radius, r_{cut}. This condition enforces locality up to a distance r_{cut} and, therefore, leads to linear scaling [274]. Based on these physical considerations, various novel schemes have been introduced employing orbitals [275, 276] or density matrix [274] formulations that achieve linear scaling and have been used in tight-binding MD calculations [276]. Recently, a CP scheme for density-matrix propagation that purports to scaling linearly was introduced [277]. For orbital formulations, the challenge in CP evolution is that the fictitious dynamics causes the orbitals to mix as the system evolves so that initially well localized orbitals tend to become rapidly delocalized. Thus, in order to achieve linear scaling within an orbital based CP approach using DFT with KS orbitals, it is necessary to introduce a device by which the orbitals could be forced to remain maximally localized as the system evolves. Here, quantum field theoretic approaches might be of utility.

12. Available software packages

Applications of the type presented here would not be possible without powerful software packages containing efficient parallel implementations of the AIMD technology. Some of

the simulations described herein were performed with the PINY_MD package developed by the present author and his collaborators [278]. [278] not only contains a description of the capabilities of the PINY_MD package (which includes both AIMD and force field modules with a path integral option) but it also gives a description of parallelization algorithms. PINY_MD is available as open-source code under the Common Public License and can be obtained from homepages.nyu.edu/m̃t33/PINY_MD/PINY.html. Other simulations were carried out with the CPMD package [279], which is also freely available from www.cpmd.org. We note that other powerful codes are available, including CASTEP (Molecular Simulations, Inc., www.tcm.phy.cam.ac.uk/casstep), CP-PAW (P E Blöchl, Clausthal University of Technology, www.pt.tu-clausthal.de/p̃aw/index.html), NWChem (Pacific Northwest National Laboratory, www.emsl.pnl.gov:2080/docs/nwchem/nwchem.html), ABINIT (X Gonze *et al*, www.abinit.org), VASP (cms.mpi.univie.ac.at/vasp/) and CAMPOS (www.fysik.dtu.dk/CAMPOS). The reader may also wish to make note of a new Web site (www.fsatom.org) established specifically for the purpose of addressing free software issues, such as standardization of input/output and interfacing and providing links to other open-source initiatives in the molecular modelling field.

13. Conclusion

A broad survey of the state of the art in AIMD, the marriage of finite temperature MD with 'on the fly' electronic structure calculations, has been presented. Our discussion began with a derivation of the approach starting from the Born–Oppenheimer approximation. Following this, representations of electronic structure were considered and basis set expansions were discussed. The extended Lagrangian or CP algorithm for performing AIMD was presented in detail and issues of adiabaticity control were considered. In addition, the calculation of observables and, in particular, observables that depend on knowledge of the electronic structure were discussed. It was also shown how nuclear quantum effects can be included via the Feynman path integral and the AIPI approach. A number of example applications were presented in order to demonstrate the power of the approach. Finally, issues related to excited states and scaling of the method with system size were discussed. A brief list of available packages for AIMD was also presented.

The basic goal of this survey was to give the reader a basic description of the current capabilities and limitations of AIMD and to give the reader a broad literature in which to look for more details. It is our hope that the present discussion will serve the non-expert as a means of entering or becoming familiar with this rapidly growing and evolving field.

Acknowledgments

The author gratefully acknowledges his colleagues throughout the world for many longstanding and fruitful collaborations, in particular, Michele Parrinello, ETH Zürich and Swiss Center for Scientific Computing, Dominik Marx, Ruhr-Universität, Bochum, Jürg Hutter, University of Zürich, Magali Benoit, Université Montpellier, and Glenn Martyna, IBM, Yorktown Heights, NY. The author also acknowledges members of his research group at New York University, Yi Liu, Zhongwei Zhu, Lula Rosso, Peter Minary and Joseph A Morrone, for their hard work and dedication. Much of the original work presented herein was supported by NSF CHE-9875824, NSF CHE-0121375, the Research Corporation RI0218, an NYU Whitehead Fellowship in Biomedical and Biological Sciences and the Camille and Henry Dreyfus Foundation, Inc. TC-02-012.

References

[1] Rick S W and Stuart S J 2002 *Rev. Comput. Chem.* **18**
[2] Warshel A and Weiss R M 1980 *J. Am. Chem. Soc.* **102** 6218
[3] Car R and Parrinello M 1985 *Phys. Rev. Lett.* **55** 2471
[4] Remler D K and Madden P A 1990 *Mol. Phys.* **70** 921
[5] Payne M C, Teter M P, Allan D C, Arias T A and Joannopoulos J D 1992 *Rev. Mod. Phys.* **64** 1045
[6] Galli G and Parrinello M 1993 *Computer Simulation in Chemical Physics (NATO ASI Series C vol 397)* p 261
[7] Tuckerman M E, Ungar P J, von Rosenvinge T and Klein M L 1996 *J. Phys. Chem.* **100** 12878
[8] Gillan M J 1997 *Contemp. Phys.* **38** 115
[9] Parrinello M 1997 *Solid State Commun.* **102** 107
[10] Marx D and Hutter J 2000 *Modern Methods and Algorithms of Quantum Chemistry (NIC Series (vol 1))* ed J Grotendorst (Forschungszentrum Juelich)
[11] Car R 2002 *Quantum. Struct. Act. Rel.* **21** 97
[12] Feynman R P and Hibbs A R 1965 *Quantum Mechanics and Path Integrals* (New York: McGraw-Hill)
[13] Feynman R 1972 *Statistical Mechanics* (Reading, MA: Benjamin)
[14] Marx D and Parrinello M 1994 *Z. Phys.* B **95** 143
[15] Marx D and Parrinello M 1996 *J. Chem. Phys.* **104** 4077
[16] Tuckerman M, Marx D, Klein M L and Parrinello M 1996 *J. Chem. Phys.* **104** 5579
[17] Marx D, Tuckerman M E and Martyna G J 1999 *Comput. Phys. Commun.* **118** 166
[18] Laasonen K, Sprik M, Parrinello M and Car R 1993 *J. Chem. Phys.* **99** 9080
[19] Fois E S, Sprik M and Parrinello M 1994 *Chem. Phys. Lett.* **223** 411
[20] Sprik M, Hutter J and Parrinello M 1996 *J. Chem. Phys.* **105** 1142
[21] Silvestrelli P L, Bernasconi M and Parrinello M 1997 *Chem. Phys. Lett.* **277** 478
[22] Trout B L and Parrinello M 1998 *Chem. Phys. Lett.* **288** 343
[23] Site L D, Alavi A and Lynden-Bell R M 1999 *Mol. Phys.* **96** 1683
[24] Silvestrelli P L and Parrinello M 1999 *Phys. Rev. Lett.* **82** 3308
[25] Sprik M 2000 *Chem. Phys.* **258** 139
[26] Krack M and Parrinello M 2000 *Phys. Chem. Chem. Phys.* **2** 2105
[27] Pfrommer B G, Mauri F and Louie S G 2001 *J. Am. Chem. Soc.* **122** 123
[28] Geissler P L, Dellago C, Chandler D, Hutter J and Parrinello M 2001 *Science* **291** 2121
[29] Schwegler E, Galli G, Gygi F and Hood R Q 2001 *Phys. Rev. Lett.* **87** 265501
[30] Izvekov S and Voth G A 2002 *J. Chem. Phys.* **116** 10372
[31] Diraison M, Martyna G J and Tuckerman M E 1999 *J. Chem. Phys.* **111** 1096
[32] Tsuchida E, Kanada Y and Tsukada M 1999 *Chem. Phys. Lett.* **311** 236
[33] Liu Y and Tuckerman M E 2001 *J. Phys. Chem.* B **105** 6598
[34] Morrone J A and Tuckerman M E 2002 *J. Chem. Phys.* **117** 4403
[35] Morrone J A and Tuckerman M E *Chem. Phys. Lett.* submitted
[36] Laasonen K and Klein M L 1994 *J. Am. Chem. Soc.* **116** 11620
[37] Laasonen K and Klein M L 1996 *Mol. Phys.* **88** 135
[38] Sprik M 1996 *J. Phys.: Condens. Matter* **8** 9405
[39] Laasonen K and Klein M L 1997 *J. Phys. Chem.* A **101** 98
[40] Meijer E J and Sprik M 1998 *J. Am. Chem. Soc.* **120** 6345
[41] Kim D and Klein M L 2000 *J. Phys. Chem.* B **104** 10074
[42] Zhu Z and Tuckerman M E 2002 *J. Phys. Chem.* B **106** 8009
[43] Chen B, Park J M, Ivanov I, Tabacchi G, Klein M L and Parrinello M 2002 *J. Am. Chem. Soc.* **124** 8534
[44] Tuckerman M E, Laasonen K, Sprik M and Parrinello M 1995 *J. Phys. Chem.* **99** 5749
[45] Tuckerman M E, Laasonen K, Sprik M and Parrinello M 1995 *J. Chem. Phys.* **103** 150
[46] Marx D, Tuckerman M E, Hutter J and Parrinello M 1999 *Nature* **367** 601
[47] Marx D, Tuckerman M E and Parrinello M 2000 *J. Phys.: Condens. Matter* **12** A153
[48] Tuckerman M E, Marx D and Parrinello M 2002 *Nature* **417** 925
[49] Sagnella D E, Laasonen K and Klein M L 1996 *Biophys. J.* **71** 1172
[50] Mei H S, Tuckerman M E, Sagnella D E and Klein M L 1998 *J. Phys. Chem.* B **102** 10446
[51] Pavese M, Berard D R and Voth G A 1999 *Chem. Phys. Lett.* **300** 93
[52] Rosso L and Tuckerman M E *J. Am. Chem. Soc.* submitted
[53] Laasonen K, Parrinello M, Car R, Lee C and Vanderbilt Y 1993 *Chem. Phys. Lett.* **207** 208
[54] Laasonen K and Klein M. L 1994 *J. Phys. Chem.* **98** 10079
[55] Tuckerman M E, Marx D, Klein M L and Parrinello M 1997 *Science* **275** 817
[56] Arstila H, Laasonen K and Laaksonen A 1998 *J. Chem. Phys.* **108** 1031

[57] Geissler P L, Dellago C, Chandler D, Hutter J and Parrinello M 2000 *Chem. Phys. Lett.* **321** 225
[58] Lee C Y, Vanderbilt D, Laasonen K, Car R and Parrinello M 1992 *Phys. Rev. Lett.* **69** 462
[59] Lee C Y, Vanderbilt D, Laasonen K, Car R and Parrinello M 1993 *Phys. Rev.* B **47** 4863
[60] Benoit M, Bernasconi M and Parrinello M 1996 *Phys. Rev. Lett.* **76** 2934
[61] Bernasconi M, Benoit M, Parrinello M, Chiarotti G L, Focher P and Tosatti E 1996 *Phys. Scr.* A **166** 98
[62] Bernasconi M, Silvestrelli P L and Parrinello M 1998 *Phys. Rev. Lett.* **81** 1235
[63] Benoit M, Marx D and Parrinello M 1998 *Nature* **392** 258
[64] Benoit M, Marx D and Parrinello M 1999 *Solid State Ion.* **125** 23
[65] Liu Z F, Siu C K and Tse J S 1999 *Chem. Phys. Lett.* **309** 335
[66] Putrino A and Parrinello M 2002 *Phys. Rev. Lett.* **88** 176401
[67] Sarnthein J, Pasquarello A and Car R 1997 *Science* **275** 1925
[68] Pasquarello A and Car R 1997 *Phys. Rev. Lett.* **79** 1766
[69] Boero M, Pasquarello A, Sarnthein J and Car R 1997 *Phys. Rev. Lett.* **78** 887
[70] Pasquarello A, Sarnthein J and Car R 1998 *Phys. Rev.* B **57** 14133
[71] Pasquarello A and Car R 1998 *Phys. Rev. Lett.* **80** 5145
[72] Massobrio C, Pasquarello A and Car R 1999 *J. Am. Chem. Soc.* **121** 2943
[73] Mauri F, Pasquarello A, Pfrommer B G, Yoon Y G and Louie S G 2000 *Phys. Rev.* B **62** R4786
[74] Benoit M, Ispas S and Tuckerman M E 2001 *Phys. Rev.* B **64** 224205
[75] Pickard C J and Mauri F 2002 *Phys. Rev. Lett.* **88** 086403
[76] Boero M, Parrinello M and Terakura K 1998 *J. Am. Chem. Soc.* **120** 2746
[77] Boero M, Parrinello M, Hueffer S and Weiss H 2000 *J. Am. Chem. Soc.* **122** 501
[78] Boero M, Parrinello M, Weiss H and Hueffer S 2001 *J. Phys. Chem.* A **105** 5096
[79] Haas K C, Schneider W F, Curioni A and Andreoni W 1998 *Science* **282** 265
[80] Stampfl C and Scheffler M 2000 *Surf. Sci.* **435** 119
[81] Haas K C, Schneider W F, Curioni A and Andreoni W 2000 *J. Phys. Chem.* B **104** 5527
[82] Stampfl C, Ganduglia-Pirovano M V, Reuter K and Scheffler M 2002 *Surf. Sci.* **500** 368
[83] Kroes G J, Gross A, Baerends E J, Scheffler M and McCormack D A 2002 *Acc. Chem. Res.* **35** 193
[84] Saitta M and Klein M L 1999 *Nature* **399** 46
[85] Saitta M and Klein M L 1999 *J. Chem. Phys.* **111** 9434
[86] Saitta M and Klein M L 1999 *J. Am. Chem. Soc.* **121** 11827
[87] Saitta M and Klein M L 2001 *J. Phys. Chem.* B **105** 6495
[88] Piana S, Sebastiani D, Carloni P and Parrinello M 2001 *J. Am. Chem. Soc.* **123** 8730
[89] Hutter J, Carloni P and Parrinello M 1996 *J. Am. Chem. Soc.* **118** 871
[90] Roethlisberger U and Carloni P 1999 *Int. J. Quantum. Chem.* **73** 209
[91] Andreoni W, Curioni A and Mordasini T 2001 *IBM J. Res. Dev.* **45** 397
[92] Rovira C and Parrinello M 2000 *Int. J. Quantum. Chem.* **80** 1172
[93] Rovira C, Schulze B, Eichinger M, Evanseck J D and Parrinello M 2001 *Biophys. J.* **81** 435
[94] Kohn W and Sham L J 1965 *Phys. Rev.* **140** A1133
[95] Parr R G and Yang W 1989 *Density Functional Theory of Atoms and Molecules* (Oxford: Oxford University Press)
[96] Dreizler R M and Gross E K U 1990 *Density Functional Theory* (Berlin: Springer)
[97] Liu Z H, Carter L E and Carter E A 1995 *J. Phys. Chem.* **99** 4355
[98] Martino B D, Celino M and Rosato V 1999 *Comput. Phys. Commun.* **120** 255
[99] Friesner R A 1985 *Chem. Phys. Lett.* **116** 39
[100] Lippert G, Hutter J and Parrinello M 1997 *Mol. Phys.* **92** 477
[101] Liu Y and Tuckerman M E *Phys. Rev. Lett.* submitted
[102] Hohenberg P and Kohn W 1964 *Phys. Rev.* B **136** 864
[103] Becke A D 1988 *Phys. Rev.* A **38** 3098
[104] Lee W Y C and Parr R C 1988 *Phys. Rev.* B **37** 785
[105] Perdew J P and Wang Y 1992 *Phys. Rev.* B **45** 13244
[106] Perdew J P, Burke K and Ernzerhof M 1996 *Phys. Rev. Lett.* **77** 3865
[107] Handy N C and Cohen A J 2001 *Mol. Phys.* **99** 403
[108] Cohen A J and Handy N C 2001 *Mol. Phys.* **99** 607
[109] Handy N C and Cohen A J 2002 *J. Chem. Phys.* **116** 5411
[110] Wu Q and Yang W 2002 *J. Chem. Phys.* **116** 515
[111] Becke A D 1992 *J. Chem. Phys.* **96** 2155
[112] Becke A D and Roussel M R 1989 *Phys. Rev.* A **39** 3761
[113] Becke A D 2000 *J. Chem. Phys.* **112** 4020

[114] Proynov E, Chermette H and Salahub D R 2000 *J. Chem. Phys.* **113** 10013
[115] Ernzerhof M, Maximoff S N and Scuseria G E 2002 *J. Chem. Phys.* **116** 3980
[116] White J A and Bird D M 1994 *Phys. Rev.* B **50** 4954
[117] Bachelet G, Hamann D and Schluter M 1982 *Phys. Rev.* B **26** 4199
[118] Troullier N and Martins J L 1991 *Phys. Rev.* B **43** 1993
[119] Vanderbilt D 1990 *Phys. Rev.* B **41** 7892
[120] Bloechl P E 1994 *Phys. Rev.* B **50** 17953
[121] Kleinman L and Bylander D M 1982 *Phys. Rev. Lett.* **48** 1425
[122] Gonze X, Kaeckell P and Scheffler M 1990 *Phys. Rev.* B **41** 12264
[123] Gonze X, Stumpf R and Scheffler M 1991 *Phys. Rev.* B **44** 1991
[124] Tuckerman M E and Martyna G J to be submitted
[125] Hockney R W 1993 *Phys. Rev.* B **48** 2081
[126] Barnett R N and Landmann U 1978 *Methods Comput. Phys.* **9** 136
[127] Martyna G and Tuckerman M 1999 *J. Chem. Phys.* **110** 2810
[128] Minary P, Tuckerman M E, Pihakari K A and Martyna G J 2002 *J. Chem. Phys.* **116** 5351
[129] Tuckerman M E, Minary P, Pihakari K A and Martyna G J 2002 *Computational Methods for Macromolecules: Challenges and Applications* ed T Schlick and H H Gan (Berlin: Springer)
[130] Mortensen J J and Parrinello M 2000 *J. Phys. Chem.* B **104** 2901
[131] Light J C, Hamilton I P and Lill J V 1985 *J. Chem. Phys.* **82** 1400
[132] Muckerman J T 1990 *Chem. Phys. Lett.* **173** 200
[133] Colbert D T and Miller W H 1992 *J. Chem. Phys.* **96** 1982
[134] Littlejohn R G, Cargo M, Carrington T, Mitchell K A and Poirier B 2002 *J. Chem. Phys.* **116** 8691
[135] Guerin S and Jauslin H R 1999 *Comput. Phys. Commun.* **121/122** 496
[136] Rosso L, Minary P, Zhu Z and Tuckerman M E 2002 *J. Chem. Phys.* **116** 4389
[137] Tuckerman M E and Parrinello M 1994 *J. Chem. Phys.* **101** 1301
[138] Laasonen K, Car R, Lee C and Vanderbilt D 1991 *Phys. Rev.* B **43** 6796
[139] Laasonen K, Pasquarello A, Car R, Lee C and Vanderbilt D 1993 *Phys. Rev.* B **47** 10142
[140] Hutter J, Tuckerman M E and Parrinello M 1995 *J. Chem. Phys.* **102** 859
[141] Hoover W 1985 *Phys. Rev.* A **31** 1695
[142] Martyna G, Klein M and Tuckerman M 1992 *J. Chem. Phys.* **97** 2635
[143] Blochl P and Parrinello M 1991 *Phys. Rev.* B **45** 9413
[144] Minary P, Martyna G J and Tuckerman M E *J. Chem. Phys.* submitted
[145] Egelstaff P A 1962 *Adv. Phys.* **11** 203
[146] King-Smith R D and Vanderbilt D 1993 *Phys. Rev.* B **47** 1651
[147] Resta R 1994 *Rev. Mod. Phys.* **66** 899
[148] Resta R 1998 *Phys. Rev. Lett.* **80** 1800
[149] Resta R 2002 *J. Phys.: Condens. Matter* **14** R625
[150] Silvestrelli P L 1999 *Phys. Rev.* B **59** 9703
[151] Berghold G, Mundy C J, Romero A H, Hutter J and Parrinello M 2000 *Phys. Rev.* B **61** 10040
[152] Marzari N and Vanderbilt D 1997 *Phys. Rev.* B **56** 12847
[153] Putrino A, Sebastiani D and Parrinello M 2000 *J. Chem. Phys.* **113** 7102
[154] Baroni S, Gianozzi P and Testa A 1985 *Phys. Rev. Lett.* **58** 1861
[155] Gonze X and Vigneron J P 1989 *Phys. Rev.* B **39** 13120
[156] Gonze X 1995 *Phys. Rev.* A **52** 1096
[157] Sebastiani D and Parrinello M 2001 *J. Phys. Chem.* A **105** 1951
[158] Berne B J and Pecora R 1976 *Dynamic Light Scattering* (New York: Wiley)
[159] Gregor T, Mauri F and Car R 1999 *J. Chem. Phys.* **111** 1815
[160] Pickard C J and Mauri F 2001 *Phys. Rev.* B **63** 245101
[161] Mauri F and Louie S 1996 *Phys. Rev. Lett.* **76** 4246
[162] Mauri F, Pfrommer B and Louie S 1996 *Phys. Rev. Lett.* **77** 5300
[163] Mauri F, Pfrommer B and Louie S 1997 *Phys. Rev. Lett.* **79** 2340
[164] Yoon Y, Pfrommer B, Mauri F and Louie S 1998 *Phys. Rev. Lett.* **80** 3388
[165] Mauri F, Pfrommer B and Louie S 1999 *Phys. Rev.* B **60** 2941
[166] Narten A H and Levy H A 1971 *J. Chem. Phys.* **55** 2263
[167] Soper A K and Phillips M G 1986 *Chem. Phys.* **107** 47
[168] Soper A K, Bruni F and Ricci M A 1997 *J. Chem. Phys.* **106** 247
[169] Sorenson J M, Hura G, Glaeser R M and Head-Gordon T 2000 *J. Chem. Phys.* **113** 9149
[170] Soper A K 2000 *Chem. Phys.* **258** 121

[171] Soper A K 1996 *Chem. Phys.* **202** 295
[172] Asfar M N and Hasted J B 1977 *J. Opt. Soc. Am.* **67** 902
[173] Coulson C A and Eisenberg D 1966 *Proc. R. Soc.* A **291** 445
[174] Batista E R, Xantheas S S and Jonsson H 1998 *J. Chem. Phys.* **109** 4546
[175] Dellago C, Bolhuis P G, Csajka F S and Chandler D 1998 *J. Chem. Phys.* **108** 1964
[176] Dellago C, Bolhuis P G and Chandler D 1999 *J. Chem. Phys.* **110** 6617
[177] Ricci M A, Nardone M, Ricci F P, Andreani C and Soper A K 1995 *J. Chem. Phys.* **102** 7650
[178] O'Reilly D E, Peterson E M and Scheie C E 1973 *J. Chem. Phys.* **58** 4072
[179] Yamaguchi T, Hidaka K and Soper A K 1999 *Mol. Phys.* **96** 1159
[180] Yamaguchi T, Hidaka K and Soper A K 1999 *Mol. Phys.* **97** 603
[181] Cornell W D, Cieplak P, Bayly C I, Gould I R, Merz K M Jr, Ferguson D M, Spellmeyer D C, Fox T, Caldwell J W and Kollman P A 1995 *J. Am. Chem. Soc.* **117** 5179
[182] Zhang Y, Lee T and Yang W 1999 *J. Chem. Phys.* **110** 46
[183] Rothlisberger U private communication
[184] Laio A, VandeVondele J and Rothlisberger U 2002 *J. Chem. Phys.* **116** 6941
[185] Nailong W 1997 *The Maximum Entropy Method* (Berlin: Springer)
[186] Librovich N B and Maiorov V D 1982 *Russ. J. Phys. Chem.* **56** 380
[187] Buchner R, Hefter G, May P M and Sipos P 1999 *J. Phys. Chem.* B **103** 11186
[188] Glasser L 1975 *Chem. Rev.* **75** 21
[189] Colomban P and Novak A 1992 *Proton Conductors: Solids, Membranes and Gels–Materials and Devices* ed P Colomban (Cambridge: Cambridge University Press)
[190] Wise H 1967 *J. Phys. Chem.* **71** 2843
[191] Venkatesan K 1957 *Proc. Indian Acad. Sci.* A **46** 134
[192] Braekken H and Harang L 1930 *Z. Kristallogr.* **75** 538
[193] Waddington T C 1958 *J. Chem. Soc.* **Dec.** 4340
[194] Corn R M and Strauss H L 1983 *J. Chem. Phys.* **79** 2641
[195] Louie S G, Froyen S and Cohen M L 1982 *Phys. Rev.* B **26** 1738
[196] Arlman E and Cossee P 1964 *J. Catal.* **3** 99
[197] Stebbins J F and Xu Z 1997 *Nature* **390** 60
[198] Farnan I 1997 *Nature* **390** 14
[199] Myers E R, Heine V and Dove M T 1998 *Phys. Chem. Mineral.* **25** 457
[200] Klinowski J, Carr W W, Tarling S E and Barnes P 1987 *Nature* **330** 56
[201] Lee S K and Stebbins J F 1999 *Am. Mineral.* **84** 937
[202] Himmel B, Weigelt J, Gerber T and Nofz M 1991 *J. Non-Cryst. Solids* **136** 27
[203] Benoit M, Tuckerman M E and Mauri F to be submitted
[204] Piana S, Sebastiani D, Carloni P and Parrinello M 2001 *J. Am. Chem. Soc.* **123** 8730
[205] Warshel A and Levitt M 1976 *J. Mol. Biol.* **103** 227
[206] Alagona G, Desmeules P, Ghio C and Kollman P A 1984 *J. Am. Chem. Soc.* **106** 3623
[207] Field M J, Bash P A and Karplus M 1990 *J. Comput. Chem.* **11** 700
[208] Gao J and Xia X 1992 *Science* **258** 631
[209] Aqvist J and Warshel A 1993 *Chem. Rev.* **93** 2523
[210] Monard G, Loos M, Thery V, Baka K and Rivail J L 1996 *Int. J. Quantum. Chem.* **58** 153
[211] Eurenius K P, Chatfield D C, Brooks B R and Hodoscek M 1996 *Int. J. Quantum. Chem.* **60** 1189
[212] Lyne P D, Hodoscek M and Karplus M 1999 *J. Phys. Chem.* A **103** 3462
[213] Eichinger M, Tavan P, Hutter J and Parrinello M 1999 *J. Chem. Phys.* **110** 10452
[214] Zhang Y, Lee T and Yang W 2000 *J. Chem. Phys.* **112** 3483
[215] Yarne D A, Tuckerman M E and Martyna G J 2001 *J. Chem. Phys.* **115** 3531
[216] Piana S and Carloni P 2000 *Proteins* **39** 26
[217] Friedman S H, Decamp D L, Sijbesma R P, Srdanov G, Wudl F and Kenyon G L 1993 *J. Am. Chem. Soc.* **115** 6506
[218] Sijbesma R, Srdanov G, Wudl F, Castoro J A, Wilkins C, Friedman S H, Decamp D L and Kenyon G L 1993 *J. Am. Chem. Soc.* **116** 6510
[219] Zhu Z, Schuster D I and Tuckerman M E 2002 *Biochemistry* at press
[220] Miura S and Okazaki S 2000 *J. Chem. Phys.* **112** 10116
[221] Cao J and Voth G 1993 *J. Chem. Phys.* **99** 10070
[222] Cao J and Voth G 1994 *J. Chem. Phys.* **100** 5106
[223] Doll J D and Beck T L 1988 *J. Chem. Phys.* **89** 5753
[224] Doll J D, Freeman D L and Gillan M J 1988 *Chem. Phys. Lett.* **143** 277

[225] Krilov G, Sim E and Berne B J 2001 *J. Chem. Phys.* **114** 1075
[226] Mak C H and Egger R 1999 *J. Chem. Phys.* **110** 12
[227] Feynman R P and Vernon F L 1963 *Ann. Phys., NY* **24** 118
[228] Cao J and Berne B J 1993 *J. Chem. Phys.* **99** 2902
[229] Hall R W and Berne B J 1984 *J. Chem. Phys.* **81** 860
[230] Tuckerman M, Berne B, Martyna G and Klein M 1993 *J. Chem. Phys.* **99** 2796
[231] Pollock E L and Ceperley D M 1984 *Phys. Rev.* B **30** 2555
[232] Tuckerman M E and Marx D 2001 *Phys. Rev. Lett.* **86** 4946
[233] Perdew J P and Zunger A 1981 *Phys. Rev.* B **23** 5048
[234] Chiavassa T, Roubin P, Pizzala L, Verlaque P, Allouche A and Marinelli F 1992 *J. Phys. Chem.* **96**
[235] Luth K and Scheiner S 1994 *J. Phys. Chem.* **98**
[236] Sirois S, Proynov E I, Nguyen D T and Salahub D R 1997 *J. Chem. Phys.* **107**
[237] Baughcum S L, Duerst R W, Rowe W F, Smith Z and Wilson E B 1981 *J. Am. Chem. Soc.* **103**
[238] Baughcum S L, Smith Z, Wilson E B and Duerst R W 1984 *J. Am. Chem. Soc.* **106**
[239] Carter E A, Ciccotti G, Hynes J T and Kapral R 1989 *Chem. Phys. Lett.* **156** 472
[240] Sprik M and Ciccotti G 1998 *J. Chem. Phys.* **109** 7737
[241] Iftimie R, Salahub D, Wei D Q and Schofield J 2000 *J. Chem. Phys.* **113** 4852
[242] Iftimie R and Schofield J 2001 *J. Chem. Phys.* **114** 6763
[243] Iftimie R and Schofield J 2001 *J. Chem. Phys.* **115** 5891
[244] Tuckerman M E, Laasonen K, Sprik M and Parrinello M 1994 *J. Phys.: Condens. Matter* **6** A93
[245] Agmon N 1995 *Chem. Phys. Lett.* **244** 456
[246] Eigen M 1964 *Angew. Chem. Int. Edn* **3**
[247] Zundel G 1976 *The Hydrogen Bond, Recent Developments in Theory and Experiments* ed P Schuster, G Zundel and C Sandorfy (Amsterdam: North-Holland) pp 687–766
[248] Agmon N 1999 *Israel J. Chem.* **39** 493
[249] Martyna G, Tuckerman M and Klein M 1992 *J. Chem. Phys.* **97** 2635
[250] Meiboom S 1961 *J. Chem. Phys.* **34** 375
[251] Agmon N 1997 *J. Mol. Liq.* **73** 513
[252] Danneel H 1905 *Z. Elektrochem.* **16**
[253] Agmon N 2000 *Chem. Phys. Lett.* **319** 247
[254] Becke A D and Edgecombe K E 1990 *J. Chem. Phys.* **92** 5397
[255] Novoa J J, Mota F, del Valle C P and Planas M 1997 *J. Phys. Chem.* A **101** 7842
[256] Alavi A, Kohanoff J, Parrinello M and Frenkel D 1994 *Phys. Rev. Lett.* **73** 2599
[257] Mermin N D 1965 *Phys. Rev.* **137** A1441
[258] Theilhaber J 1992 *Phys. Rev.* B **46** 12990
[259] Saalmann U and Schmidt R 1996 *Z. Phys.* D **38** 153
[260] Runge E and Gross E K U 1984 *Phys. Rev. Lett.* **52** 997
[261] Gross E K U and Kohn W 1990 *Adv. Quantum. Chem.* **21** 255
[262] Parac M and Grimme S 2002 *J. Phys. Chem.* A **106** 6844
[263] Frank I, Hutter J, Marx D and Parrinello M 1998 *J. Chem. Phys.* **108** 4060
[264] Tully J C and Preston R K 1971 *J. Chem. Phys.* **55** 562
[265] Tully J C 1990 *J. Chem. Phys.* **93** 1061
[266] Doltsinis N L and Marx D 2002 *Phys. Rev. Lett.* **88** 166402
[267] Bowler D R, Miyazaki T and Gillan M J 2002 *J. Phys.: Condens. Matter* **14** 2781
[268] Pearson M, Smargiassi E and Madden P A 1993 *J. Phys.: Condens. Matter* **5** 3221
[269] Foley M and Madden P A 1996 *Phys. Rev.* B **53** 10589
[270] Wang Y A, Govind N and Carter E A 1998 *Phys. Rev.* B **58** 13465
[271] Wang Y A, Govind N and Carter E A 2001 *Phys. Rev.* B **64** 129901
[272] Watson S, Jesson B J, Carter E A and Madden P A 1998 *Europhys. Lett.* **41** 37
[273] Cho K, Arias T A, Joannopoulos J and Lam P K 1993 *Phys. Rev. Lett.* **71** 1808
[274] Li X P, Nunes R W and Vanderbilt D 1993 *Phys. Rev.* B **48** 14646
[275] Galli G and Parrinello M 1992 *Phys. Rev. Lett.* **69** 3547
[276] Galli G and Mauri F 1994 *Phys. Rev.* B **50** 4316
[277] Schlegel H B, Millam J M, Iyengar S S, Voth G A, Daniels A D, Scuseria G E and Frisch M J 2001 *J. Chem. Phys.* **114** 9758
[278] Tuckerman M E, Yarne D A, Samuelson S O, Hughes A L and Martyna G J 2000 *Comput. Phys. Commun.* **128** 333
[279] Hutter J, Alavi A, Deutsch T, Bernasconi M, Goedecker S, Marx D, Parrinello M and Tuckerman M E 1995–99 *CPMD Version 3.3* Max-Planck Institut fuer Festkoerperforschung and IBM Zurich Research Laboratory