

Linear-Scaling First-Principles Calculations of Entire Polar Semiconductor Nanorods

Philip Avraam

Thesis submitted for the degree of
Doctor of Philosophy of the University of London
and the
Diploma of Imperial College

The Thomas Young Centre
Bessemer Building
Imperial College
London

July 2011

Abstract

This thesis describes a study of polar semiconductor nanorods using first-principles quantum-mechanical computer simulations.

Semiconductor nanostructures in solution are a very exciting class of material due to our growing ability to manipulate their shapes, sizes and the superstructures they assemble in to, to produce a wide range of technologically useful properties.

Nanocrystals of binary semiconductor, such as those of ZnO, have been observed to exhibit very large dipole moments which affect their internal electronic structure (and therefore their optical properties, for example) as well as their interactions with the surrounding environment, which can affect the kinetics of self-assembly and the stability of the structures formed.

A detailed understanding of the factors contributing to this large polarity in nanocrystals has proven elusive for two main reasons: (1) the multitude of factors that are involved, ranging from the effects of surface chemistry, to the non-centrosymmetric nature of the underlying crystal, to quantum confinement, to long-range electrostatics, to interactions with the solvent and considerations of thermodynamic stability; and (2) the limitations of current experimental technique, which don't allow us the level of control over, or knowledge of the state of our system, that is necessary to be able to disaggregate these factors.

The main advantage of computer simulation is the level of control over and knowledge of our 'experimental' conditions that it allows, thus making it an ideal tool for addressing this problem. Recent developments in linear-scaling density-functional theory, combined with improvements in computational power, have for the first time brought accurate quantum-mechanical methods in to the realm of applicability to nanocrystals of realistic size.

This thesis focuses on how a nanorod's polarity and electronic structure are affected by changes in the surface terminating species, by surface relaxations, nanorod size, semiconductor type, applied electric fields, and interactions with neighbouring nanorods.

I would like to thank my supervisor Peter Haynes, and my second supervisors Nick Hine and Paul Tangney for their wisdom and guidance throughout the course of this degree.

Contents

List of Figures	xi
List of Tables	xii
Introduction	1
1 Semiconductor nanocrystals: structure, surfaces and charge distribution	4
1.1 Colloidal nanocrystals	4
1.2 The quantum mechanics of semiconductor nanocrystals	7
1.2.1 From atoms to molecules to nanocrystals	8
1.2.2 From bulk crystals to nanocrystals	12
1.3 Charge distribution in nanocrystals	13
1.3.1 A review of the evidence	13
1.3.2 Theory: from bulk surfaces to slabs to thin-films to nanocrystals	15
1.3.3 Nanocrystal superstructures and self-assembly	19
2 Introduction to ab initio computational materials science	21
2.1 Many-body quantum mechanics	22
2.1.1 The many-body Schrödinger equation	22
2.1.2 The Born-Oppenheimer approximation	24
2.1.3 The variational principle	26
2.2 Density functional theory	27
2.2.1 Hohenberg-Kohn Theorems	27
2.2.2 Levy's Constrained Search Formulation	29

2.2.3	The Kohn-Sham approach	30
2.3	Periodic supercells	36
2.3.1	Bloch's theorem	37
2.3.2	k -point sampling	38
2.3.3	The plane-wave basis set	39
2.4	Pseudopotentials	40
2.4.1	Non-linear core corrections	43
2.5	Summary	44
3	Linear scaling	46
3.1	The density matrix	46
3.2	Constraints on the density matrix	48
3.2.1	Normalisation	48
3.2.2	Idempotency	48
3.3	Locality in quantum systems	49
3.4	Non-orthogonal orbitals and the density kernel	50
3.5	Fractional occupancies and idempotency	52
3.6	Imposing idempotency	53
3.7	The <i>psinc</i> basis set	54
3.8	FFT box technique	55
3.9	Finding the ground-state energy	58
3.9.1	Functionals to be minimised	58
3.9.2	Minimisation strategy	60
3.10	Isolated supercells: the cutoff-Coulomb interaction	61
3.11	Summary	63
4	Nanorod charge distribution and electronic structure	64
4.1	The effect of surface termination	65
4.1.1	Results and discussion	67
4.2	The effect of structural relaxation	76

4.3	The effect of nanorod size	81
4.4	The effect of semiconductor type (GaAs, GaN, AlN)	85
4.5	The effect of applied electric fields	88
4.5.1	Introduction	88
4.5.2	Electronic and ionic polarisability	90
4.5.3	Polarisability: a comparison between nanorods and bulk . . .	92
4.5.4	The effect of surface termination	96
4.5.5	The effect of size	96
4.6	The effect of pairwise interactions between nanorods	99
4.6.1	Introduction	99
4.6.2	Results and discussion	100
4.7	Summary and conclusion	105
A	Simulation details and preliminary calculations	110
B	Considerations of nanorod stability and environment	113
C	Description of relaxed geometries	118
	References	121

List of Figures

1.1	Model of a stoichiometric, wurtzite CdSe nanocrystal obtained from the analysis of high resolution transmission electron microscopy images [1]. (figure is from Ref. [2]).	6
1.2	Energy level diagram from Ref. [3], showing the hybridisation of atomic orbitals, the formation of bonding and anti-bonding orbitals, and the emergence of bands and surface dangling-bond states.	10
1.3	Characteristic density of states plots for dimensions three to zero. As the dimension is lowered, the density of states displays sharper features.	12
1.4	Dependence of the 2D-slab's formation energy on slab thickness. There are three identifiable regimes. Figure from Ref. [4]	18
2.1	A schematic illustration of a pseudopotential and pseudo-wavefunction	41
3.1	One dimensional psinc function with $M=9$. The values are unity on the grid point on which it is centred and zero on all other grid points within the cell (which is $M=9$ grid points wide). From Ref. [5]	56
4.1	An unrelaxed GaAs nanorod with Ga, As, and H atoms colored red, blue, and green, respectively. The Ga and As terminated ends are indicated.	66
4.2	Dipole moment d_z as a function of H atom coverage of the polar surfaces for a nanorod with fully H-terminated lateral surfaces.	68

4.3	Charge density profile along z , with x and y directions averaged out .	69
4.4	(a) Charge density and (b) electrostatic potential as a function of position z along the long axes of the H/H, H/B, and H/P nanorods. Both quantities are integrated over the plane perpendicular to the long axis and smoothed on the length scale of a wurtzite unit cell. . .	70
4.5	Density of states for: (a) H/H (top) and B/H (bottom); (b) H/B and B/B; (c) H/P and B/P. The Fermi level is at 0eV. Nanorods with bare lateral surfaces have relatively many states near the Fermi-level.	71
4.6	Local densities of states for three nanorods: (a) H/H; (b) H/B; (c) B/H; (d) one of the nanorods of figure 4.2 with 67% hydrogen coverage on polar surfaces. The global Fermi level is plotted in each case. Red (blue) curves correspond to slabs close to the polar surfaces terminated by Ga (As).	72
4.7	Difference in electron density profile between H/H and H/P	75
4.8	relaxed rod: Fermi-level stabilised at mid-gap states	77
4.9	Iso-surface of the HOMO viewed (a) along the $[11\bar{2}0]$ direction, and (b) looking at the (0001) polar surface. (isovalue of $0.001 e/a_0^3$). . . .	77
4.10	Iso-surface of the LUMO viewed (a) along the $[11\bar{2}0]$ direction, and (b) looking at the (0001) polar surface. (isovalue of $0.001 e/a_0^3$). . . .	78
4.11	Density of states plot for: (a) b/b , before (top) and after (bottom) relaxation; (b) h/b ; (c) h/h; (d) b/h. In all cases relaxation has tended to move surface states away from the Fermi-level.	79
4.12	Structure of an unrelaxed, H terminated nanorod (0) and the relaxed structures of b/b to h/h. Red spheres represent Ga atoms, purple represents As, and grey represents H. The left hand picture looks down the $[11\bar{2}0]$ direction and the right hand picture looks at a view of the (0001) polar surface.	80
4.13	The magnitude of the dipole moment increases linearly with nanorod length for nanorods of cross-sectional area 3.56nm^2	81

4.14	The magnitude of the dipole moment increases with nanorod cross sectional areas (shown in (a)) for nanorods of length 12.8nm. The relationship is not linear over the whole of this range, as the polar surface charge density falls off quickly, asymptotically approaching a finite value (shown in (b))	82
4.15	LDOS plot of (a) a slab in the middle of the rods; (b) the Ga-rich end (top) and the As-rich (bottom)	83
4.16	Laterally averaged and Gaussian-smoothed charge distribution along the length of the nanorod for rods of AlN, GaN and GaAs. The data is the same in (a) and (b), but (b) is has a magnified “charge density” axis.	86
4.17	LDOS plots of the cation-rich (top) and anion rich (bottom) polar surfaces of AlN, GaN and GaAs.	87
4.18	Dependence of the polarisability of a material on the frequency of oscillation of the field.	89
4.19	Change in the electron density when a field of $E = 0.05$ V/nm is applied. Grey is no change, white is negative change, black is positive change. Dipole moment changed by 154 D (from -658 D to -504 D).	90
4.20	Change in cross-sectional-averaged electron density resulting from the field	91
4.21	LDOS plots of the Ga-rich polar surface (top) and As-rich polar surface (bottom) (a) with no applied electric field; and (b) in an applied uniform electric field of $E = +0.05$ V/nm	94
4.22	The change in polarisability density nanorod H/H-r with length and cross-sectional area.	97
4.23	LDOS (a) without and (b) with Efield.	98

4.24	(a) the potential across the rod. The potential difference remains almost identical, despite the applied field. The nanorod's internal field has increased slightly -0.089 V/nm to -0.10 V/nm. (b) Difference between charge densities, at $E=0.05\text{V/nm}$ and $E=0$. A net charge of $+0.41e$ has been transferred from one end to the other as a result of the field, resulting in a change in dipole moment from -818 D to -610 D.	98
4.25	Two aligned nanorods separated by a distance s . We are looking down on the (0001) surface of the nanorods.	100
4.26	The variation of the dipole moment of one of the pair of nanorods, with separation, in the aligned and anti-aligned configurations. . . .	101
4.27	The LDOS of an isolated rod (black) and of one of the pair of nanorods separated by a distance 2.57 nm between the rod's central axes (red) in the cases where the rods are (a) aligned; and (b) anti-aligned . . .	102
4.28	Difference in electron density along the length of the rod between one of the pair of nanorods separated by a distance 2.57 nm between the rod's central axes, and the isolated rod, in the cases where the rods are (a) aligned; and (b) anti-aligned. The charge transfer from the the left hand end to the right of the rod is 0.1e in (a), and -1.15 in (b).	103
4.29	The potential across the nanorod when aligned with (at separation 2.57 nm), anti-aligned with, and isolated from another nanorod. The decay of the internal field is most prominent when rods are anti-aligned, and least prominent when aligned.	104
4.30	Typical band-edge diagram of the nanorods in this work. The top and bottom surfaces here represent the end polar surfaces of the nanorod.	106
B.1	(a) lower bound, $\mu_H = -0.6281$ (b) a mid-range value, $\mu_H = -0.5845$ (c) upper bound, $\mu_H = -0.5683$	116
C.1	Geometry relaxation of rod with H-terminated As-rich polar surface and bare Ga-rich surface	119

List of Tables

4.1	Dipole moment d_z , net charge of the left-hand half Q_L and electric field at the mid-point E_m of the nanorods.	68
4.2	Configurations and dipole moments d_z before and after structural relaxation of the four nanorods.	78
4.3	Some properties of AlN, GaN, and GaAs in nanorod and in bulk. . .	87
4.4	Electronic polarisabilities of nanorod H/H-r and of a volume of bulk of the same size, resolved in three directions - $[11\bar{2}0]$ (x-direction), $[1\bar{1}00]$ (y-direction) and $[0001]$ (z-direction). The bulk value is calculate from the optical susceptibility $\chi(\infty)$. Error values represent one standard deviation.	93
4.5	Polarisabilities	96
A.1	Convergence data for the total energy and lattice parameters	111
B.1	Total energy of molecule, calculated in DFT, and the enthalpy of vaporisation plus fusion per molecule, obtained from Ref. [6]	115

Introduction

Inorganic nanocrystals have been the subject of intense study over the past decade, both from the perspective of fundamental physics, and technological application. Innovations in the techniques of nanoscale fabrication, together with the wide range of novel and technologically useful properties which these materials exhibit, and are predicted to exhibit, have been the main reasons for the explosion of interest.

Nanoscale materials can have very different electronic, optical and chemical properties from those of their larger scale equivalents [7]. One source of difference is the much larger surface to volume ratio seen in nanocrystals compared with bulk, resulting in a correspondingly larger degree of influence for the nanocrystal's surfaces on its properties. A second reason is the effect of quantum confinement, which becomes significant when the size of a nanoparticle in any dimension is comparable to the De Broglie wavelength of electrons in the material – typically between 10 nm and 1 μm in semiconductors. The ability to control the size of nanoparticles, their shape, and the constitution of their surfaces can allow a very great degree of control over their properties. For example, the band gap in CdS can be tuned between 2.5 and 4 eV [8] and its melting temperature between 1600 °C and 400 °C [9], only by reducing the size of nanocrystal, without changing the composition. Recent advances in our ability to manipulate these factors make nanocrystals highly promising materials for technological exploitation in a very wide range of areas.

Another extremely useful feature of nanocrystals, particularly nanoparticles in solution, is their ability, under specific conditions, to self-assemble into more complicated structures that would otherwise be extremely difficult or impossible to

fabricate. Specifically, nanoparticles have been observed to assemble in 3D lattice structures [10], 2D sheets, 1D ‘perl neckless’-like structures [11], and complex quasi-0D structures displaying a wide variety of different shapes and symmetries [12]. The collective properties of these quasi-lattice structures are, in general, different from those displayed by the individual nanoparticles and bulk samples, and are often controllable by fine tuning the environmental conditions. These collective properties have been exploited in sensor technology and potentially in a wide variety of other areas of application [13].

As we gain finer control over the properties of nanocrystals and the structures they assemble into, the number of applications will inevitably increase. At present, they have found application or potential application in: *in vivo* imaging in biology [14], light-emitting diodes [15], lasing devices [16], solar cells [17], hydrogen fuel cells [18], battery technology [19], magnetic information storage [20] and targeted cancer drug therapies [21]. This list is not exhaustive.

A great deal has been achieved over the past decade in the realms of nanoparticle synthesis, characterisation and application [22]. A large body of empirical knowledge has been developed, linking specific environmental conditions and fabrication methodologies to particular nanoparticle shapes and some of their properties. However, due to the limitations of current experimental technique, relatively little is known about the link between a nanocrystal’s detailed morphology and its electronic structure and properties. An illustrative example, and one which this Thesis directly addresses, relates to the charge distribution in semiconductor nanocrystals.

The charge distribution in a nanocrystal is a crucially important characteristic which is interlinked with all its other properties. Some nanocrystals exhibit no net charge and no significant dipole moment, other nanocrystals are known to carry charge and some semiconductor nanocrystals are found to exhibit a large dipole moment, even in the electronic ground state. Experiment doesn’t provide a sufficient level of control over the nanocrystal’s structure, or a precise enough level of knowledge about the individual nanocrystal’s state or electronic structure to be able to disaggregate which of the multitude of factors is most critical in determining a

nanocrystal's charge distribution, and how they are interlinked.

The limitations of current experimental technique place an increased importance on other means of investigation, the chief of which is *first-principles* quantum-mechanical computational simulation. However, accurate methods based on density functional theory (DFT) have been limited in the extent to which they have been able to fulfil their promise in this area, mainly because of the the scaling of computational resources with system size in conventional algorithms ($O(N^3)$ scaling, where N is the number of electrons in the system), their poor suitability to 0D systems, and the lack of sufficient computational power. As a result it has not been possible to access realistically large nanocrystal sizes using these methods. However, new linear-scaling DFT methods have been developed which, together with the improvements in computational power, which continues to increase exponentially with time, enable the study of more realistically sized nanocrystals.

The first aim of this Thesis is to use linear-scaling DFT to investigate the factors determining the charge distribution in nanocrystals, with particular focus on polar semi-conductor nanorods in colloidal solution. We also investigate their dielectric properties, which are likely to be a very important considerations for nanocrystal superstructures.

The structure of this Thesis is as follows: In chapter 1 we review the relevant background topics and the literature on nanocrystals and their polarity; in chapter 2 we review the theory behind conventional implementations of density functional theory; in chapter 3 we discuss linear-scaling methods, and specifically, the approach to linear-scaling taken in the ONETEP code; in chapter 4 we present details of our simulations of nanorods, together with our findings.

1

Semiconductor nanocrystals: structure, surfaces and charge distribution

Semiconductor nanocrystals are small lumps of semiconductor crystal of size in each dimension typically in the range 1 nm to 1 μm . In this chapter we focus on describing what is currently known about the structural properties, electronic structure and ground-state charge distribution in semiconductor nanocrystals, and the relevant issues surrounding these topics.

1.1 Colloidal nanocrystals

There is a wide range of different types of nanocrystals which can be synthesised, varying in size, shape, purity and surface adsorbates. They can be synthesised using a host of different techniques, from physical methods, like pulsed laser ablation, to mechanical methods, like bead milling, to chemical synthesis, or ‘wet’ chemistry [23]. This Thesis is concerned with investigating phenomena seen in chemically synthesised nanocrystal colloids, since these are closest to the sizes and surface conditions we have intended to replicate.

Chemical synthesis of nanocrystals involves the precipitation of a solid phase out of a liquid solution of reactants. This is typically a multi-staged process involving

1. Semiconductor nanocrystals: structure, surfaces and charge distribution

a nucleation stage, a growth stage, and a stabilisation stage, with careful control of the reactant composition and temperature throughout the process.

Crucial to the synthesis of colloidal nanocrystal and to the ability to control the nanocrystal size and shape, is the use of specific molecules known as stabilising agents. These are surfactants which serve to prevent inter-nanocrystal agglomeration and to reduce the reactivity of the nanocrystal surfaces, slowing or halting nanocrystal growth¹. These molecules typically consist of a hydrophilic, polar head, and a hydrophobic hydrocarbon chain tail. The stabilising agents commonly used in colloidal syntheses include alkyl thiols, amines, carboxylic and phosphonic acids, phosphines, phosphine oxides, phosphates, phosphonates, and various coordinating solvents (e.g. ethers) [25].

Evaluating nanocrystal structure involves identifying the composition and crystallinity of the nanocrystal core, the shape and size of the nanocrystal, and the morphology and composition of its surfaces. Powder X-ray diffraction (XRD) and electron diffraction are used to determine the crystal phase of the core, transmission electron microscopy (TEM) can be used to study nanocrystal shapes and sizes. Polar nanocrystals (crystals composed of two atomic species) are found in two crystalline phases, cubic zincblende, and hexagonal wurtzite. Zincblende nanocrystals tend to form roughly spherical (though faceted) shapes, and wurtzite nanocrystals tend to form anything from roughly spherical to more elongated shapes, which, if their aspect ratio is large enough, are called nanorods. Figure 1.1 shows the shape of a CdSe nanocrystal in the wurtzite structure with a low aspect ratio, which are observed using electron microscopy in Reference [1].

¹The final shape of nanocrystals produced by these chemical processes depends on many factors which contribute to the kinetics of growth. It is very difficult to negate kinetic barriers and bring a crystal into equilibrium with its surrounding environment. Nevertheless, theoretical predictions of a nanocrystal's equilibrium shape can be an important reference point for studying the energetics of nanocrystal formation. Equilibrium crystal shapes have traditionally been predicted using Wulff's construction [24], on the basis of knowledge of the free energies of the low energy facets of a crystal. However, there are problems with applying Wulff's construction at the nanoscale (1) Wulff's theorem is a continuum theory; (2) Edges and corners become extremely important considerations on the nanoscale; (3) fields which penetrate the entire crystal can cause the free energies of different surfaces on the nanocrystal to become interlinked.

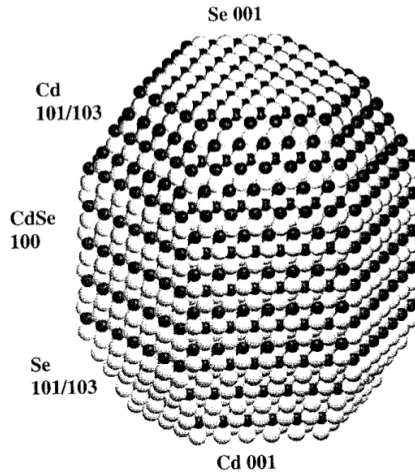


Figure 1.1 – Model of a stoichiometric, wurtzite CdSe nanocrystal obtained from the analysis of high resolution transmission electron microscopy images [1]. (figure is from Ref. [2]).

More difficult however, is the characterisation of the surface composition and morphology. To illustrate the difficulty, we use the example of a recent detailed study [26] which probes the surface chemistry of colloidal PbSe. The study uses a range of nuclear magnetic resonance (NMR) techniques to determine that the nanoparticle consists of a stoichiometric PbSe core, terminated by a Pb-rich surface, followed by a shell of surface ligands. The ligands are found to be composed 95-100% of oleic acid ligands, and 0-5% tri-n-octylphosphine, and the ligand density is 4.2 ligands per square nanometer of surface – equivalent to one ligand for every two surface Pb atom. This is the most detailed nanocrystal surface characterisation we have seen, yet it says little about the precise Pb/Se composition of the surfaces, and nothing about how the surfaces may be reconstructed, whether there is a symmetry to the way ligands attach to the surfaces, and whether there are differences between the different facets on the nanocrystal (e.g differences in ligand composition). These factors can be important in determining some of the properties of the nanocrystals, but are inaccessible to study using current experimental techniques, making it difficult to study their relative importance for the nanocrystal's polarity (among other properties).

This ‘accessibility gap’ can be partially filled using the methods of computational simulation. For example, a recent study [27] used X-ray absorption fine structure measurements combined with density functional theory calculations, to predict that the surfaces on the sides of CdSe nanorods are reconstructed Cd rich, with some unpassivated Se atoms. Other solely computational studies have determined that lattice constants can become smaller nearer nanocrystal surfaces, and that smaller nanocrystals tend to have smaller lattice constants compared with their bulk counterparts [28]. Still other computational studies have attempted to investigate the effect on the electronic structure of surface termination with ligands, one of which found that ligands tend to move the surface states away from the bandgap, rendering the surfaces ‘optically inert’ [29]. Many important questions remain open in these areas.

1.2 The quantum mechanics of semiconductor nanocrystals

A fundamental aspect of quantum mechanics is the idea of wave-particle duality, which says that every particle can be associated with a matter wave with a wavelength (the de Broglie wavelength) that is inversely proportional to its momentum. On scales comparable with a particle’s de Broglie wavelength, Newton’s equations of motion break down, and the system is described within the framework of quantum mechanics, and according to a new equation of ‘motion’: Schrödinger’s wave equation.

In this section, we consider three different quantum mechanical models, and their consequences for nanocrystals. Section 1.2.1 considers two atom-centric models: the valence bond model, in which it is assumed that atomic orbitals (AOs) interact in molecules to form individual ‘bond orbitals’ (BOs); and a tight-binding model, in which AOs can combine to form molecular orbitals which, in general, span the whole molecule. Section 1.2.2 considers the free-electron model of an extended

crystal, in which electrons are not bound strongly to individual atoms in the crystal.

1.2.1 From atoms to molecules to nanocrystals

In an atom, negatively charged electrons are said to ‘orbit’ a more massive positively charged nucleus. The electronic states (‘orbitals’) of the simplest atom, hydrogen, have been calculated analytically. In atoms with more than one electron, the Schrödinger equation is impossible to solve because of the electron-electron terms in the Hamiltonian. Fortunately, good approximations exist which simplify the problem. We discuss electronic structure calculations in more detail in Chapters 2 and 3.

In order to make the Schrödinger equation solvable for many electron atoms, the independent electron approximation is made, in which electrons are assumed to occupy their own one-particle atomic orbitals (AO), similar to those of the hydrogen atom. AOs are characterised by three quantum numbers: the principal quantum number n , the orbital angular momentum quantum number l , and the component of angular momentum along any given direction m_l . The state can be denoted $|n, l, m_l\rangle$, following Dirac’s notation. The Pauli exclusion principle ensures that each orbital can be occupied with no more than a pair of electrons of opposite spin. For example, the eight valence electrons in a neon atom occupy the states $|2, 0, 0\rangle$, $|2, 1, 0\rangle$, $|2, 1, -1\rangle$ and $|2, 1, 1\rangle$, with a pair of spin-antialigned electrons in each state. The state with $l = 0$ is known as an s -orbital and is spherically symmetric, and the three states with $l = 1$ are known as p_x -, p_y - and p_z -orbitals and have dumbbell-like shapes along the x , y and z axes respectively.

When two atoms are brought closer together so that their atomic orbitals overlap, the energies of the states change. These new states can no longer be thought of as belonging to only one atom, and are known as one-electron molecular orbitals (MOs). The energies of these molecular orbitals can be very well approximated in systems with a large number of electrons with *linear combinations of atomic orbitals* (LCAOs). For example, when two H atoms are brought together so that

their occupied s -orbitals overlap, the two MOs that are produced result from (1) adding the AO of atom 1 with the AO of atom 2 (2) subtracting AO of atom 1 from the AO of atom 2. The first MO has lower energy than the initial AOs, and the second MO has a higher energy. The two electrons occupy the lower energy molecular orbital in a spin-pair, resulting in a lowering of the total energy as compared with the isolated atoms, and the atoms are said to have *bonded*. The lower energy MO is known as the *bonding* orbital, and the higher energy one is the *anti-bonding* orbital.

If we consider the bonding axis between two atoms to be along the z direction, bonding and anti-bonding orbitals produced (within the framework of valence bond theory [30]) as a result of a linear combination of s orbitals, or p_z -orbitals are cylindrically symmetric around the bonding (z) axis, and are known as σ (bonding) and σ^* (anti-bonding) orbitals. Those produced as a result of linear combinations of p_x orbitals or p_y orbitals (which approach each other side-by-side) are not symmetric around the bonding axis and are known as π orbitals.

Valence bond theory can also explain the tetrahedral coordination of carbon in, for example, CH_4 by allowing a conceptual process called AO hybridisation in which the s -, p_x -, p_y - and p_z -AOs of the carbon atom are linearly combined so that they interfere constructively in some regions and destructively in others, in this case producing 4 notional ‘hybrid’ AOs with the same energy as each other and oriented towards the 4 corners of a tetrahedron. These ‘hybrid’ AOs, known as sp^3 hybrid orbitals (because they combine 1 s and 3 p orbitals), form σ and σ^* orbitals when combined with the s -orbital of the hydrogen atoms. The sp^n hybrids of a central atom depend on the coordination of the atom. In cases where a central atom is bonded with inequivalent atoms, or at unequal bond angles, it is necessary to consider more general sp^λ orbitals, where λ is non-integer [31].

We now turn to nanocrystals. Nanocrystals can be thought of as large molecules consisting of hundreds or thousands of atoms, to which valence bond theory can be applied. In GaAs nanocrystals, for example, the Ga atom has valence configuration $4s^24p^1$ and the As atom has $4s^24p^3$. After making the approximation that only valence electrons are involved in bonding, the tetrahedral co-ordination of

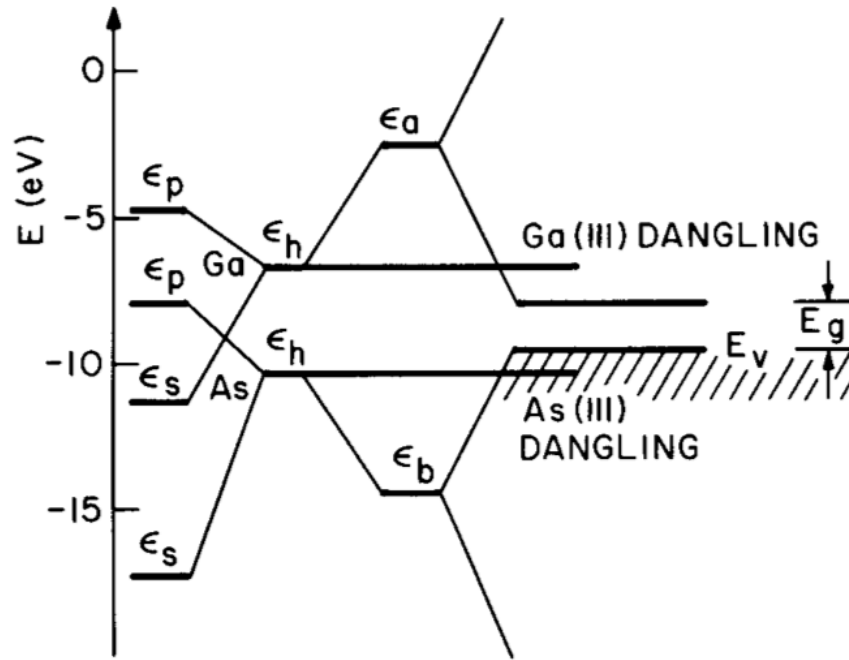


Figure 1.2 – Energy level diagram from Ref. [3], showing the hybridisation of atomic orbitals, the formation of bonding and anti-bonding orbitals, and the emergence of bands and surface dangling-bond states.

both Ga atoms and As atoms in bulk GaAs can be thought of as arising from the hybridisation of the s and p orbitals into four sp^3 hybrid orbitals on both the Ga atom and As atom. The linear combination of a hybrid AO on the As atom, with a hybrid AO on the Ga atom, gives rise to a bonding orbital with energy ϵ_b and an anti-bonding orbital with energy ϵ_a . Figure 1.2 describes the process schematically.

As the number of atoms in the nanocrystal increases, more bonding and anti-bonding orbitals are created. The energy separation between the high energy anti-bonding state and the low energy bonding state tends towards a constant as the number of atoms in the crystal increases. Crystals that are large enough to be in this constant regime have many states clustered over a range (bandwidth) around the original bonding and anti-bonding energies, with such small energy differences between these states, that it is simpler to say that an energy *band* has been created, with a continuum of allowable energies within these bands. The band which consists of anti-bonding molecular orbitals is the conduction band, and the one which consists

1. Semiconductor nanocrystals: structure, surfaces and charge distribution

of bonding MO is the valence band. We discuss how the band-structure changes with nanocrystal size in section 1.2.2

So far, we have discussed the delocalised MOs responsible for bonding within nanocrystals. However, the model we have been discussing (summarised in Figure 1.2) also provides insight in to the other important aspect of nanocrystals, its surfaces. In bulk crystal, bonding orbitals are formed from hybrid orbitals, but those hybrid orbitals on the surface atoms directed out from the surface remain unbound. These orbitals are known as *dangling bonds*. In Figure 1.2 we see that the dangling bond associated with a Ga surface atom has energy ϵ_h^{Ga} , which is just above the conduction band minimum, and the hybrid associated with an As surface atom has energy ϵ_h^{As} : just below the valence band maximum. In general, dangling hybrids may lie within the band-gap, and are typically quite near the Fermi level.

One can imagine cleaving the a crystal by cutting through bonding orbitals and increasing the total energy of the system (both surfaces) by $\epsilon_h^{Ga} + \epsilon_h^{As} - 2\epsilon_b$ for every bond that is cut, assuming that the two electrons in each bonding orbital are split between the Ga dangling hybrid of one surface, and the opposing As dangling hybrid of the other. In addition to this contribution to the surface energy, there is a possible contribution arising from the spin-pairing of the electrons on the As dangling hybrid, which lies below the Fermi level, and the reconstruction of the surface, which changes the AO constitution of the dangling hybrids and bonding orbitals of the surface atoms, moving them even further from the gap region. These additional contributions result in a large portion of the energy lost by creating the two surfaces being regained. One might predict from this picture that surfaces with more As atoms than Ga atoms, like the $(\bar{1}\bar{1}\bar{1})$ -As surface, would have a strong negative charge, due to the double occupation of the As dangling hybrids. However this crude non-interacting electron picture does not consider the increase in the Coulomb energy due to spin-pairing electrons, or due to the creation of uncompensated charge. It also does not consider that spatially repeating dangling hybrids will themselves broaden into surface bands on large enough surfaces. We discuss the problem of charged surfaces in Section 1.3.2.

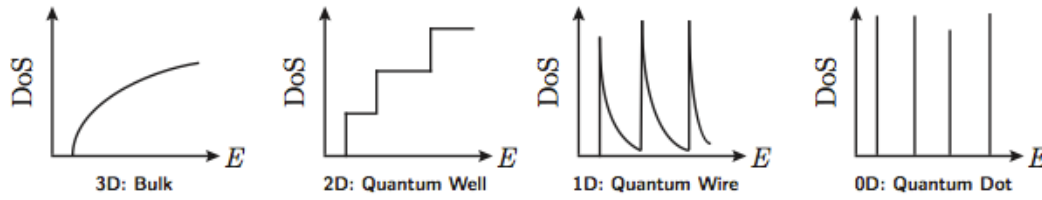


Figure 1.3 – Characteristic density of states plots for dimensions three to zero. As the dimension is lowered, the density of states displays sharper features.

1.2.2 From bulk crystals to nanocrystals

A bulk solid can be modelled as a cubic system of length L , containing N non-interacting (i.e. chargeless) ‘electrons’. This model, known as the free electron model, surprisingly captures many of the physical aspects of real solids, despite not accounting for the interaction of electrons with the underlying crystal lattice, or with other electrons. Typically periodic boundary conditions are invoked in order to avoid having to account for effects associated with abrupt boundaries (surfaces). The solutions to this model are a discrete set of travelling waves, each indexed by a wavevector $\mathbf{k}$ which is interpreted as defining the momentum and the energy ($\propto k^2$) of the electron described by that wavefunction. We discuss wavevectors in more detail in Chapter 2, where we describe Bloch’s theorem. The set of wavevectors constitute a uniform grid of points in k -space with spacing between points equal to $2\pi/L$. As $L \rightarrow \infty$, the space between, $2\pi/L \rightarrow 0$, and the spectrum of allowed electron momenta and energies becomes continuous. Within this model, it is possible to calculate the analytical form of the density of states (DOS) as a function of k , or of energy, for structures of different dimensionality.

Figure 4.5 shows the DOS for structures with different dimensionality, ranging from 3 to 0 dimensions. As the dimension is lowered, the density of states displays sharper features, until the energy spectrum becomes discrete in 0D – the relevant dimensionality for atoms and nanocrystals.

In semiconductors, the occupied states in the valence band and unoccupied states in the conduction band are separated by the band gap. As the size of the

nanocrystal is reduced, these bands fragment into a set of points at discrete values of k . The dispersion relations of the conduction band at its minimum, and the valence band at its maximum, can be approximated as being parabolic, and therefore, free-electron like. This observation motivates the approximation that the energies of the states near the band-edges change due to confinement in the same way that a free-electron confined to a box varies with the size of the box. In a 3D infinite cubic well of length L , one-electron energy levels vary as $1/L^2$. Although this is a good first approximation for small nanocrystals, corrections arise when the Coulomb interaction between electrons, crystal anisotropy and spin-orbit coupling are considered. A more detailed model [32] based on the effective-mass approximation predicts the effective band-gap of a spherical particle, in the limit of strong confinement (i.e. nanoparticle size is smaller than the effective Bohr radius of an exciton in bulk) as

$$E'_g = E_g + \frac{\hbar^2 \pi^2}{2L^2} \left[\frac{1}{m_e^*} + \frac{1}{m_h^*} \right] - \frac{1.786e^2}{\epsilon L} - \frac{0.124e^4}{\hbar^2 \epsilon^2} \left[\frac{1}{m_e^*} + \frac{1}{m_h^*} \right]^{-1},$$

where the first term is the bulk band-gap, the second is the result of confinement on the kinetic energy, the third represents the Coulomb attraction between the electron and hole, and the last term is the result of spatial correlation between the electron and hole.

1.3 Charge distribution in nanocrystals

1.3.1 A review of the evidence

Direct experimental evidence. In 1997 it was observed that wurtzite CdSe nanocrystals in colloidal solution exhibit a large permanent dipole moment oriented along the polar [0001] axis [33]. Such dipole moments have since been observed in other colloidal polar nanocrystals – wurtzite and zincblende [34, 35]. This dipole moment has been the subject of a good deal of interest due to the expectation that it has consequences for nanocrystal optical properties, and could play a role in nanocrystal

self-assembly [11].

There have been a substantial number of contradictory studies relating to the origin of the dipole moment, some attributing it to the non-centrosymmetric nature of the wurtzite structure [36] which is associated with an intrinsic spontaneous polarization – an interpretation that is widely accepted [10, 37]. The dipole moments found in cubic polar nanocrystals have been attributed to charges distributed in surface states, and it was suggested that this may also be the case in wurtzite nanocrystals too [35].

An electrostatic force microscopy study of CdSe nanorods [38] observed no dipole moment in the samples studied, but did find that some nanorod lateral surfaces were charged – some positive, others negative, and others neutral. The study concluded that their readings must have arisen from nanorods whose lateral surfaces were not exactly parallel to the wurtzite c -axis, so that $\mathbf{P} \cdot \mathbf{n}$ was non-zero on some surfaces, and positive or negative depending on their relative orientation with the c -axis.

Evidence from theory and simulation. A theoretical study has revealed a strong enhancement of the polarity of nanorods compared with thin films of the same length along the [0001] direction [39], reinforcing the suspicion that something more than crystal symmetry is involved in determining the dipole moment.

Other studies have attributed importance to different factors, including nanocrystal shape [40], molecular passivation of nanocrystal surfaces [40, 41], surface reconstruction [41, 42] and a piezoelectric effect caused by strain at the nanocrystal surfaces [43].

The question of origin and the important factors influencing the dipole moment is still unresolved, and is one of the primary topics of concern in this Thesis. We introduce the topic of surface charge more generally below.

1.3.2 Theory: from bulk surfaces to slabs to thin-films to nanocrystals

Cutting a binary compound along a particular plane will create a surface. If this surface is constituted of an equal number of each of the two species of atom, it is known as a *stoichiometric* surface, and if there is more of one type of ion than the other, it is referred to as a *polar* surface².

An ionic model of a crystal, in which the metal ion has a nominal charge of $+p$, and the non-metal has nominal charge of $-p$, would predict that polar surfaces (like wurtzite's [0001]) are charged, and stoichiometric surfaces (like zincblende's [001]) are neutral.

However, classical electrostatic arguments show that semi-infinite charged surfaces cost an infinite amount of energy per unit area and therefore cannot be stable [4, 44, 45]. To see this we consider the energy stored in the electric field E produced by an infinite plane (containing the x and y axes) of positive (or negative) charge. The field is constant in space all the way up to $z = \pm\infty$, and stores an energy per unit volume of $u = \frac{1}{2}\epsilon|E|^2$, therefore the energy per unit area of the plane of charge is infinite and the plane is unstable. Overall charge neutrality is a condition for stability in semi-infinite surfaces.

It has been known for a few decades that polar surfaces exhibit quite drastic stabilising reconstructions, including the addition or removal of surface atoms, resulting in a complex surface defect structure [45].

Ionic models [46] have been used to successfully predict the number of ions which must be removed or added in order to create a stoichiometric (and therefore, charge neutral) semi-infinite surface. For example, the wurtzite [0001]-metal ion (cation) surface with a frozen bulk-like termination (or the zincblende [111]-metal

²For clarity, we point out that there are four different uses of the word 'polar' in this field which can easily be confused: (1) polar crystals can refer to crystals constituted of two atomic species; (2) polar surfaces can refer to surfaces with a surplus of one atomic species over others (this is the usage of 'polar surfaces' in this Thesis); (3) polar surfaces can also refer to the surface dipole which exists in all crystal surfaces (this is not a usage used in this Thesis); (4) *nanocrystal polarity* is used synonymously with the nanocrystal's dipole moment.

1. Semiconductor nanocrystals: structure, surfaces and charge distribution

ion surface), can be made stoichiometric by the removal of $1/4$ of the surface metal ions. Alternatively, it can be made stoichiometric by adding $1/4$ of a layer's worth of non-metal ions (anions). The notional process of removing this electrostatic instability by creating charge neutral surfaces is known as 'polarity compensation'.

In a purely ionic crystal, polarity compensation is synonymous with the creation of stoichiometric surfaces, however, real macroscopically large crystals, exhibiting covalent bonding character, in real environments have alternative potential polarity compensation mechanisms available. In general, there are three mechanisms for polarity compensation,

1. modification of the surface ionic composition, resulting in terracing or reconstructions, depending on how vacancies or adatoms order;
2. partial filling of the electronic surface states – resulting in 'metallic' surfaces;
3. adsorption of charged foreign atoms, ions or molecules.

In order to account for covalency in crystals, a theoretical scheme known as the *electron counting model* has been devised [46], in which bonding orbitals are assumed, which carry two electrons each. In this model, a surface is stable only if bonding orbitals are completely (not partially) filled – a rule known as the *auto-compensation principle* [47].

For example, in bulk GaAs, each Ga (As) atom shares its 3 (5) electrons between its 4 bonds, so that the two electrons which constitute each bond are made from a contribution of $3/4$ electrons from Ga and $5/4$ electrons from As. The unreconstructed (111)-Ga surface has 'dangling' bond states with only $3/4$ electrons to fill each of these states, which does not fulfil the auto-compensation principle. If these $3/4$ ($5/4$) electrons per dangling bond were removed (added) to the surface, the auto-compensation principle could be satisfied, but the surfaces would be highly charged, and polarity would not have been compensated. However, both the auto-compensation principle, and the polarity-compensation principle are satisfied if $1/4$ of the Ga atoms are removed from the surface layer – a prediction that has indeed

been confirmed.

The electron-counting model has limitations: in many cases, ‘dangling bonds’ have no physical significance; there is no reason to assume that the number of electrons per bond contributed by each atom is conserved between bulk and surface; and it is difficult to generalise this model for ternary crystals [48].

A model has been proposed called the bond-transfer model [49] which is closely related to a previous model based on Pauling’s partial charge model [3], in which valence charge is thought of as being distributed between the atoms, with more of the valence charge being attributed to the anion than the cation, to a degree determined in part by the relative electronegativities of the two atoms. For example, in bulk MgO the six bonds surrounding each atom are equivalent, so the charges on each atom can be written.

$$Q_{\text{Mg}} = 2 - 6\Delta, \quad (1.1)$$

$$Q_{\text{O}} = -2 + 6\Delta, \quad (1.2)$$

where the formal ionic charges of Mg and O (+2 and -2 respectively) are corrected by a parameter Δ which describes the electron transfer per bond, resulting from the hybridisation between the orbitals of an atom with the 6 first-neighbour atoms. An analytic form for Δ has been derived from a tight-binding model, and it is a function of the resonance integrals associated with the bonds in question [48]. It is possible to generalise this model to consider partially filled surface states, specifically, to predict what level of partial-filling is required in order to produce neutral surfaces.

An important result which emerges from these considerations is that a relaxation of the surfaces alone can not eliminate the polar instability. The charge needed for polarity compensation is a function of the geometry of bulk ion cores, and the hybridisation of bonds in the bulk crystal, and changes in surface covalency do not add net charge [48, 50].

Polarity in finite systems. In thin-films (2D slabs) or 0D nanocrystals, there is no divergence of the electrostatic part of the surface energy, and therefore surfaces

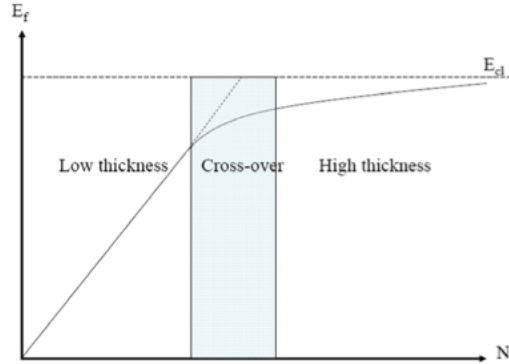


Figure 1.4 – Dependence of the 2D-slab’s formation energy on slab thickness. There are three identifiable regimes. Figure from Ref. [4]

can be charged. For example the formation energy of a thin-film with charged surfaces has a contribution arising from the electric field produced by the excess surface charge, which increases in significance in thicker films and therefore changes the energetics of the surface. Calculations have confirmed the dependence of surface energetics on slab thickness: one study has shown the variation in the stability of different surface termination with thickness [51].

It is plausible, then, that polar thin films may be able to be taxonomised into three size regimes [4], illustrated in Figure 1.4: (1) very thin films, which have surface charge independent of thickness [52], and therefore have a linearly increasing formation energy per atom with thickness (like the stored energy of a parallel plate capacitor has a linear dependence on thickness); (2) macroscopically sized slabs with neutral surfaces, which therefore have no dependence of their stability (formation energy per atom) on thickness; and (3) a cross-over region. The cross over region is likely to exhibit partial polarity compensation. It may, for example, exhibit a kind of surface reconstruction which isn’t observed in the macroscopic case, which only partially reduces surface charge. If there is such a cross-over region, its characteristic thickness is likely to be very small, due to the rapid increase of electrostatic potential in polar films.

This taxonomy is general enough to plausibly apply to 0D nanocrystals also, though it may be expected that the characteristic nanocrystal sizes which apply in

the cross-over region are larger than for thin films, due to their reduced dimensionality.

1.3.3 Nanocrystal superstructures and self-assembly

Nanocrystals, including polar nanocrystals have been observed to self-assemble into 3D lattice structures [10], 2D sheets, 1D ‘perl necklass’-like structures [11], and complex quasi-0D structures displaying a wide variety of different shapes and symmetries, under specific conditions [12]. There are a number of different types of forces which are thought to play a role, to varying degrees, in these self-assembly processes, including Van der Waals (VDW) forces, magnetic and electrostatic forces, repulsive steric forces, solvation forces, depletion forces and capillary forces to name only a few [13, 53].

In polar nanocrystals, dipole-dipole interactions have been shown to be far larger than VDW forces at medium distances (~ 5 times the size of the nanocrystal) [54]. There have been attempts to empirically model nanoparticle self-assembly using different assumptions about a nanocrystal’s charge distribution. There have been point-dipole approximations [55], approximations which combine a uniform spherical charge distribution with a point dipole and VDW forces [56], and most recently, an attempt to take in to account the non-centro-symmetric nature of a real nanocrystal charge distributions by approximating it by two opposite point charges placed on opposing surfaces of the nanoparticle [10].

These models have been successful at replicating some of the self-assembled superstructures seen in real solutions, though they are known to show significant error, particularly at small inter-particle spacings (e.g. at high nanoparticle concentrations). One of the studies has predicted the existence of an interesting nanoparticle superlattice phase, in which nanoparticles adopt an anti-ferroelectric ordering – a phase which has not yet been observed so far as we have been able to tell [10].

The quality of the predictions of these models depends on the accuracy of their assumptions. One major reason for the crudeness of these approximations is

1. Semiconductor nanocrystals: structure, surfaces and charge distribution

that very little has been discovered from experiment about the precise nature of the charge distribution in nanocrystals. In fact, a recent study which has tried to measure this directly using electron force microscopy (EFM) failed to observe a dipole moment at all [38]. On top of this, very little is known about the dielectric properties of nanocrystals, which are also likely to be very important in superstructures made of closely packed nanocrystals with non-uniform charge distributions.

It is one of the main topics of this Thesis to investigate the charge distribution and dielectric properties of nanocrystals, in part to inform better empirical models of nanoparticle superstructures and self-assembly.

2

Introduction to *ab initio* computational materials science

Most low-energy physical and chemical processes can be described in principle by solving the equations of non-relativistic quantum mechanics. In practice however, these equations are not analytically solvable for all but the most simple of systems so if we hope to use these equations to describe real systems, we must solve them numerically by modelling the system on a computer.

Computational modelling of systems and processes is done throughout science and engineering, for example the modelling of epidemics in biology, or modelling a working part in a car in engineering. Quantum mechanical calculations described here are different from these in the sense that they are ‘first-principles’ (or *ab initio*) calculations i.e. the model which is being solved doesn’t use empirical parameters (which usually serve to approximate the effect of a collection of more fundamental mechanisms) which are adjusted for the specific system which is being modelled, but instead the calculation requires only the specification of the atoms present in the system.

There have been a variety of different *ab initio* methods developed over the years. Solving the many-body Schrödinger equation (MBSE) directly has proved to be a completely impractical method for systems with many atoms, because the

computational cost of doing so increases exponentially with the number of electrons (or typically N^5 to N^8 in correlated wavefunction methods e.g the configuration interaction CISD method scales with $O(N^6)$). In 1964 Hohenberg and Kohn showed that there is a unique ground-state electron density associated with the external potential of the MBSE and this proof paved the way for the recasting of the MBSE in the form of a total-energy functional of the unique groundstate electron density $n_0(\mathbf{r})$ (a function of only 3 variables instead of the $3N$ variables of the ground-state wavefunction). This recasting in terms of the ground-state electron density, known as *density-functional theory* (DFT), led to the development of methods which scale with the cube of the number of particles, which is much more acceptable for practical implementation.

A group of methods which have been particularly successful in accurately modelling crystals with relatively small computational cost are based on density-functional theory and have been termed *total-energy pseudopotential methods*. With modern computers it is possible to model systems containing thousands of atoms using these methods. In this Chapter, we describe *ab initio* theory with particular focus on plane-wave DFT.

More recent developments in DFT methods have taken advantage of the property of locality inherent in quantum many-body systems to produce codes which scale linearly with system size. With these developments we are now able model systems containing thousands, to tens of thousands of atoms with the high accuracy which DFT allows, and we will be able to take full advantage of advances in computing power in the future to model much larger systems. We describe linear-scaling methods in Chapter 3.

2.1 Many-body quantum mechanics

2.1.1 The many-body Schrödinger equation

The general Hamiltonian for a system of interacting electrons and nuclei is,

$$\hat{H} = -\frac{1}{2} \sum_i \nabla_i^2 + \sum_{i>j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{i\alpha} \frac{Z_\alpha}{|\mathbf{r}_i - \mathbf{R}_\alpha|} - \frac{1}{2} \sum_\alpha \frac{\nabla_\alpha^2}{M_\alpha} + \sum_{\alpha>\beta} \frac{Z_\alpha Z_\beta}{|\mathbf{R}_\beta - \mathbf{R}_\alpha|} \quad (2.1)$$

in atomic units (where $e = \hbar = a_0 = m_e = 1$). Here, $\mathbf{r}$ represents electronic positions, $\mathbf{R}$ are positions of nuclei, Latin subscripts represent electrons, Greek subscripts represent nuclei. Z_α is the charge on nucleus α , M_α is the mass of nucleus α . The first and fourth terms are the kinetic energy operators of electrons and nuclei respectively; the second term represents the electron-electron Coulomb interaction, the third term is the electron-nucleus Coulomb interaction; and the fifth term is the nucleus-nucleus Coulomb interaction. This Hamiltonian is a non-relativistic one, so it omits spin-orbit coupling, magnetic effects and mass-velocity effect, although these corrections are small under most circumstances and can be added using perturbation theory. The fundamental, non-relativistic equation of motion of these electrons and nuclei is the time-dependent Schrödinger equation

$$i \frac{\partial \Psi(\{\mathbf{r}_i\}, \{\mathbf{R}_\alpha\}, t)}{\partial t} = \hat{H} \Psi(\{\mathbf{r}_i\}, \{\mathbf{R}_\alpha\}, t) \quad (2.2)$$

where $\Psi(\{\mathbf{r}_i\}, \{\mathbf{R}_\alpha\}, t)$ is the many-body wave function for the electrons and nuclei. None of the potential energy terms in equation 2.1 are explicitly time dependent, so we can perform a separation of the time and position variables in the many-body wavefunction without loss of generality, to obtain the solution

$$\Psi(\{\mathbf{r}_i\}, \{\mathbf{R}_\alpha\}, t) = \Phi(\{\mathbf{r}_i\}, \{\mathbf{R}_\alpha\}) e^{-iEt} \quad (2.3)$$

where E is a separation constant, thereby reducing the problem to solving the non-relativistic time-independent MBSE

$$\hat{H} \Phi(\{\mathbf{r}_i\}, \{\mathbf{R}_\alpha\}) = E \Phi(\{\mathbf{r}_i\}, \{\mathbf{R}_\alpha\}) \quad (2.4)$$

2.1.2 The Born-Oppenheimer approximation

At any instant in time, the electrostatic forces felt by electrons and nuclei are of the same order of magnitude, since the charges on electrons and nuclei have the same magnitude. So one would expect the changes in momenta caused by these interactions to be similar for both electrons and nuclei. Nuclei however, are typically 10^4 to 10^5 times more massive than electrons, which means that the average speed of electrons must be about the same order of magnitude greater than that of the nuclei. Since the characteristic timescale of electronic motion ($\propto \frac{1}{|v|}$) is much smaller than that of nuclei, we can make the approximation, without a great loss of accuracy, that for any particular nuclear configuration, the electrons move instantaneously into their ground-state configuration for that static configuration of nuclei. This separation of the nuclear and electronic motion is called the Born-Oppenheimer (or adiabatic) approximation [57]. Mathematically, it can be derived as follows. Motivated by the separation of variables method used to solve linear partial differential equations, we first write the wavefunction $\Phi(\{\mathbf{r}_i\}, \{\mathbf{R}_\alpha\})$ as

$$\Phi(\{\mathbf{r}_i\}, \{\mathbf{R}_\alpha\}) = \psi(\{\mathbf{r}_i\}; \{\mathbf{R}_\gamma\})\phi(\{\mathbf{R}_\gamma\}) \quad (2.5)$$

where in $\psi(\{\mathbf{r}_i\}; \{\mathbf{R}_\alpha\})$, the ionic coordinates $\mathbf{R}_\alpha$ are parameters and not dynamical variables. Substituting into equation (2.4) we find that the equation for ψ is

$$\left\{ -\frac{1}{2} \sum_i \nabla_i^2 + \sum_{i>j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{i\alpha} \frac{Z_\alpha}{|\mathbf{r}_i - \mathbf{R}_\alpha|} \right\} \psi(\{\mathbf{r}_i\}; \{\mathbf{R}_\gamma\}) = \mathcal{E}(\{\mathbf{R}_\gamma\}) \psi(\{\mathbf{r}_i\}; \{\mathbf{R}_\gamma\}) \quad (2.6)$$

This is the Schrödinger equation for the *electron coordinates only*, which can be solved in principle for a fixed set of nuclear positions $\{\mathbf{R}_\alpha\}$. After substituting in this expression, the full Hamiltonian applied to the full wavefunction becomes

$$\left\{ -\frac{1}{2} \sum_\alpha \frac{\nabla_\alpha^2}{M_\alpha} + \sum_{\alpha>\beta} \frac{Z_\alpha Z_\beta}{|\mathbf{R}_\beta - \mathbf{R}_\alpha|} + \mathcal{E}(\{\mathbf{R}_\alpha\}) \right\} \Phi(\{\mathbf{r}_i\}, \{\mathbf{R}_\gamma\})$$

$$\begin{aligned}
&= \psi(\{\mathbf{r}_i\}; \{\mathbf{R}_\gamma\}) \left\{ -\frac{1}{2} \sum_{\alpha} \frac{\nabla_{\alpha}^2}{M_{\alpha}} + \sum_{\alpha > \beta} \frac{Z_{\alpha} Z_{\beta}}{|\mathbf{R}_{\beta} - \mathbf{R}_{\alpha}|} + \mathcal{E}(\{\mathbf{R}_{\gamma}\}) \right\} \phi(\{\mathbf{R}_{\gamma}\}) \\
&- \sum_{\beta} \frac{1}{2M_{\beta}} \{ 2\nabla_{\beta} \phi(\{\mathbf{R}_{\gamma}\}) \cdot \nabla_{\beta} \psi(\{\mathbf{r}_i\}; \{\mathbf{R}_{\gamma}\}) + \phi(\{\mathbf{R}_{\gamma}\}) \nabla_{\beta}^2 \psi(\{\mathbf{r}_i\}; \{\mathbf{R}_{\gamma}\}) \} \quad (2.7)
\end{aligned}$$

The first term on the right-hand side of this expression is called the adiabatic term and the term containing $\mathcal{E}$ is called the adiabatic contribution of the electrons to the energy. The remaining terms in the above expression are called non-adiabatic terms and they contribute very little to the energy, because of the factor $\frac{1}{2M_{\beta}}$ (or, writing in the electron mass m_e explicitly, the factor becomes $\frac{m_e}{2M_{\beta}}$) which is very small (of order 10^{-4} to 10^{-5}). If the first of the non-adiabatic terms becomes appreciable, it gives rise to the electron-phonon interaction, one of the manifestations of which is the formation of Cooper pairs (pairs of electrons bound by the exchange of a virtual phonon) which can lead to the emergence of superconductivity. These non-adiabatic terms also induce scattering of electrons by the motion of the nuclei.

The Born-Oppenheimer approximation (otherwise known as the adiabatic approximation) is to throw away all the non-adiabatic terms so that equation (2.7) becomes a Schrödinger equation in the nuclear coordinates only, with Hamiltonian

$$\left\{ -\frac{1}{2} \sum_{\alpha} \frac{\nabla_{\alpha}^2}{M_{\alpha}} + \sum_{\alpha > \beta} \frac{Z_{\alpha} Z_{\beta}}{|\mathbf{R}_{\beta} - \mathbf{R}_{\alpha}|} + \mathcal{E}(\{\mathbf{R}_{\gamma}\}) \right\} \phi(\{\mathbf{R}_{\gamma}\}) = E \phi(\{\mathbf{R}_{\gamma}\}) \quad (2.8)$$

This approximation is crucial because it allows the separation of electronic and nuclear motion. For a particular nuclear configuration, the ground-state electronic energy is $\mathcal{E}(\{\mathbf{R}_{\gamma}\})$. Varying the ionic position maps out a potential energy surface for the nuclei, for which the motion of the ions can be solved. Newtonian mechanics typically suffices to solve for the motion of the nuclei.

2.1.3 The variational principle

We have separated the many-body problem so that a Hamiltonian operator for electrons only can be identified (from equation 2.6) as

$$\hat{H} = -\frac{1}{2} \sum_i \nabla_i^2 + \sum_{i>j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{i\alpha} \frac{Z_\alpha}{|\mathbf{r}_i - \mathbf{R}_\alpha|} \quad (2.9)$$

Consider the expectation value of this Hamiltonian:

$$\langle E \rangle = \frac{\langle \psi | \hat{H} | \psi \rangle}{\langle \psi | \psi \rangle} = E[\psi] \quad (2.10)$$

which is a functional of the electronic wave-function. If we make a small change to the wave-function so that $|\psi\rangle \rightarrow |\psi\rangle + |\delta\psi\rangle$, the change in the energy functional is

$$\begin{aligned} \delta E[\psi] &= E[\psi + \delta\psi] - E[\psi] \\ &= \frac{\langle \psi + \delta\psi | \hat{H} | \psi + \delta\psi \rangle}{\langle \psi + \delta\psi | \psi + \delta\psi \rangle} - \frac{\langle \psi | \hat{H} | \psi \rangle}{\langle \psi | \psi \rangle} \\ &= \frac{\langle \delta\psi | \hat{H} | \psi \rangle + \langle \psi | \hat{H} | \delta\psi \rangle}{\langle \psi | \psi \rangle} - \frac{\langle \psi | \hat{H} | \psi \rangle}{(\langle \psi | \psi \rangle)^2} (\langle \delta\psi | \psi \rangle + \langle \psi | \delta\psi \rangle) + O(\delta\psi^2) \\ &= \frac{1}{\langle \psi | \psi \rangle} \left[\langle \delta\psi | (\hat{H} - E[\psi]) | \psi \rangle + \text{c.c.} \right] + O(\delta\psi^2) \end{aligned} \quad (2.11)$$

where c.c. is the complex conjugate of the previous term. Thus, we see that $\delta E = 0$ when $|\psi\rangle$ is an eigenstate of the Hamiltonian, so that the eigenstates can be found by finding the stationary points of $E[\psi]$. Here, we are concerned with finding the ground state $|\psi_0\rangle$ and the corresponding ground-state energy E_0 . Suppose we have a state close to the ground-state, but with a small error. The error can be expanded as a linear combination of eigenstates so that

$$|\psi\rangle = |\psi_0\rangle + \sum_{n=1}^{\infty} c_n |\psi_n\rangle \quad (2.12)$$

Substituting in to the energy functional, we obtain:

$$\begin{aligned}
E[\psi] &= \frac{E_0 + \sum_{n=1}^{\infty} |c_n|^2 E_n}{1 + \sum_{n=1}^{\infty} |c_n|^2} \\
&= E_0 + \sum_{n=1}^{\infty} |c_n|^2 (E_n - E_0) + O(|c_n|^4)
\end{aligned} \tag{2.13}$$

Hence we can see two important things from these equations:

1. $E[\psi] \geq E_0$ with equality only when $|\psi\rangle = \alpha|\psi_0\rangle$, ($\alpha \in \mathbb{C}$)
2. The error in the estimate of the ground-state energy is proportional to the square of the error in the eigenstate i.e. second order in c_n .

These are very important results as they show that to calculate the ground state energy, we can try to minimise the functional $E[\psi]$, and the energy that we get gives an upper bound on the true value of E_0 , and that even a relatively poor estimate of the ground-state wave-function can give relatively good estimates of E_0 .

2.2 Density functional theory

After the Born-Oppenheimer approximation, the non-relativistic Hamiltonian for an N electron system is written in equation 2.9 or, written more concisely,

$$\hat{H} = \hat{T} + \hat{V}_{ee} + \hat{V}_{en} \tag{2.14}$$

In this section we describe the approach taken in density functional theory to solving this equation.

2.2.1 Hohenberg-Kohn Theorems

These are two theorems which justify on rigorous theoretical grounds the approach taken in density functional theory (DFT) of recasting the problem of finding the ground-state solution of the many-body problem described by the MBSE, in terms of minimising a functional of the electron density. I will just state these theorems here

and not prove them because in section 2.2.2 I shall discuss the framework due to Levy, which proves the same things as the Hohenberg-Kohn theorems but in more generality. The proofs are quite simple and can be found in reference [58].

Theorem 1

The external potential $V_{en}(\mathbf{r})$ (here, the instantaneous Coulomb attraction between the electrons and nuclei in the electronic Hamiltonian) is determined uniquely (to within an additive constant) by the ground state electron density $n_0(\mathbf{r})$ i.e. there is a one-to-one mapping between $V_{en}(\mathbf{r})$ and the ground state density $n_0(\mathbf{r})$.

Theorem 2

For any external potential $V_{en}(\mathbf{r})$, there is an associated universal functional of the electron density $E[n]$, whose global minimum is the ground-state energy of the system E_0 , and the density corresponding to this ground-state energy, is the ground-state density $n_0(\mathbf{r})$.

Comments on the theorems

These theorems are proved by Hohenberg and Kohn to apply to densities $n(\mathbf{r})$ which are ground-state densities of some external potential (known as *v-representable* densities). Theorem 2 states a variational principle, namely that $E_0 = \min_n E[n]$ for all possible *v-representable* densities. While it is true that the ground state density n_0 must be *v-representable*, we would require a variational search strategy which would only search the space of *v-representable* densities, but for all practical variational search strategies this is not guaranteed. In fact, it has been shown that many “reasonable” trial densities arising in minimisation methods are not *v-representable* [59]. A second problem with these theorems is that the derivation only holds for non-degenerate ground-states. Fortunately, Levy was able to reformulate DFT in a way which extends the domain of possible electron densities $n(\mathbf{r})$ from that of *v-representable* densities, to all densities arising from some anti-symmetric *N*-body wavefunction, a condition known as *N-representability*, and in a way that

is valid even for degenerate ground-states. This formulation is known as the Levy constrained search formulation [60].

2.2.2 Levy's Constrained Search Formulation

Levy's theory [60] starts by making a distinction between the ground-state wave-function Ψ_0 , and a second wave-function Ψ , both of which integrate to the same ground-state density $n_0(\mathbf{r})$. Since Ψ_0 is the ground-state, the variational principle dictates that

$$\langle \Psi | \hat{H} | \Psi \rangle \geq \langle \Psi_0 | \hat{H} | \Psi_0 \rangle = E_0 \quad (2.15)$$

where $\hat{H}$ is the many-body electron Hamiltonian in equation 2.14. The electron-nucleus interaction potential energy operator $\hat{V}_{en}$ in this Hamiltonian can be written in the form

$$\hat{V}_{en} = \sum_{i\alpha} \frac{Z_\alpha}{|\mathbf{r}_i - \mathbf{R}_\alpha|} = \sum_{i=1}^N V_{nuc}(\mathbf{r}_i) = \int V_{nuc}(\mathbf{r}) \sum_{i=1}^N \delta(\mathbf{r} - \mathbf{r}_i) d\mathbf{r} = \int V_{nuc}(\mathbf{r}) \hat{n}(\mathbf{r}) d\mathbf{r} \quad (2.16)$$

where $V_{nuc}(\mathbf{r}_i) = \sum_{\alpha} \frac{Z_\alpha}{|\mathbf{r}_i - \mathbf{R}_\alpha|}$. The expectation value of this operator is now just

$$\left\langle \Psi \left| \int V_{nuc}(\mathbf{r}) \hat{n}(\mathbf{r}) d\mathbf{r} \right| \Psi \right\rangle = \int V_{nuc}(\mathbf{r}) n(\mathbf{r}) d\mathbf{r} \quad (2.17)$$

We see in equation 2.17 that the functional $\langle \Psi | V_{en} | \Psi \rangle$ is a functional only of the density, and not the wave-functions, and therefore the term can be eliminated from the inequality in equation 2.15, leaving

$$\langle \psi | (\hat{T} + \hat{V}_{ee}) | \psi \rangle \geq \langle \psi_0 | \hat{T} + \hat{V}_{ee} | \psi_0 \rangle \quad (2.18)$$

The meaning of equation 2.18 is that the ground-state wave-function ψ_0 is the one which not only integrates to the ground-state density $n_0(\mathbf{r})$ (since ψ also integrates to $n_0(\mathbf{r})$), but also minimises the kinetic plus electron-electron repulsion energy. Thus, defining a functional (known as the universal functional) as

$$F[n] = \min_{\psi \rightarrow n} \langle \psi | \hat{T} + \hat{V}_{ee} | \psi \rangle \quad (2.19)$$

where $F[n]$ searches all ψ yielding the density n , allows the ground-state energy to be expressed as

$$E_0 = \min_n \left[F[n] + \int V_{nuc}(\mathbf{r})n(\mathbf{r})d\mathbf{r} \right] \quad (2.20)$$

$$= \min_n E[n] \quad (2.21)$$

where

$$E[n] = F[n] + \int V_{nuc}(\mathbf{r})n(\mathbf{r})d\mathbf{r} \quad (2.22)$$

which is a search through all N-representable densities i.e. densities that can be obtained from any anti-symmetric wave-function. Levy's formalism not only resolves the representability and degeneracy problems of the Hohenberg-Kohn theorems, but also provides a prescription for constructing this energy functional (from equation 2.22). All that remains is to find an accurate form for the functional $E[n]$.

2.2.3 The Kohn-Sham approach

An exact expression for the two terms in equation (2.19) in terms of the electron density is not known. Kohn and Sham suggested [61] the problem could be overcome by splitting the universal functional $F[n]$ up into three additive terms as follows

$$F[n] = T_{KS}[n] + E_H[n] + E_{xc}[n] \quad (2.23)$$

The second term equation 2.23, $E_H[n]$ is an approximation of the electron-electron interaction energy, and is the electrostatic energy of a classical continuous 'electron' cloud distribution (the Hartree energy)

$$E_H[n] = \frac{1}{2} \int \int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' \quad (2.24)$$

The first term in equation 2.23, $T_{KS}[n]$, is constructed, following Levy, as

$$T_{KS}[n] = \min_{\Psi_{KS} \rightarrow n} \langle \Psi_{KS} | \hat{T} | \Psi_{KS} \rangle \quad (2.25)$$

which is the kinetic energy functional for an auxilliary system of ‘fictitious’ non-interacting particles Ψ_{KS} , interacting with a mean-field potential V_{KS} , obeying the Hamiltonian

$$\hat{H}_{KS} = \sum_i \left(-\frac{1}{2} \hat{\nabla}_i^2 + \hat{V}_{KS}(\mathbf{r}_i) \right) \quad (2.26)$$

so that the energy of this system of fictitious particles (known as the Kohn-Sham energy) is

$$E_{KS}[n] = \min_{\Psi_{KS} \rightarrow n} \langle \Psi_{KS} | \hat{H}_{KS} | \Psi_{KS} \rangle = T_{KS}[n] + \int V_{KS}(\mathbf{r}) n(\mathbf{r}). \quad (2.27)$$

V_{KS} is chosen so that the ground-state density of this non-interacting system is equal to $n(\mathbf{r})$ (which can always be done since $n_0(\mathbf{r})$ is a v -representable density). A justification for why V_{KS} is chosen like this, and how it is calculated is given when we discuss the Kohn-Sham equations below. This independent electron Hamiltonian (equation 2.26) is separable into N one-electron Schrödinger-like equations of the form

$$\left(-\frac{1}{2} \hat{\nabla}_i^2 + \hat{V}_{KS}(\mathbf{r}_i) \right) \psi_i = \varepsilon_i \psi_i \quad (2.28)$$

where i labels electrons. Ψ_{KS} is an antisymmetric wavefunction made up of a suitable sum of products of these wavefunctions, specifically, a *Slater-determinant* of form,

$$\Psi_{KS}(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = \frac{1}{\sqrt{N!}} \sum_P (-1)^{\zeta_P} \psi_{(P \times 1)}(\mathbf{r}_1) \psi_{(P \times 2)}(\mathbf{r}_2) \dots \psi_{(P \times N)}(\mathbf{r}_N) \quad (2.29)$$

where the sum runs over all permutations (where P labels the permutation) of the set of orbitals $\{\psi_i\}$, and ζ_P is the parity of the permutation P . Substituting Ψ_{KS} , in to the Hamiltonians of equation 2.28, we discover that the expectation value for the Kohn-Sham Hamiltonian (i.e. the Kohn-Sham energy) is just

$$E_{KS}[n] = \sum_i^{N/2} \varepsilon_i \quad (2.30)$$

The final term in Kohn and Sham's casting of the universal functional in equation 2.23, is defined by equations 2.19 and 2.23 as being

$$E_{xc}[n] = F[n] - T_{KS}[n] - E_H[n] \quad (2.31)$$

One can see that $E_{xc}[n]$ contains the corrections to the Hartree energy necessary to fix the electron-electron interaction energy, and corrections to the Kohn-Sham kinetic energy to fix the kinetic energy of the fully interacting system. Thus, this functional accounts for correlation between the spatial position of electrons due to their Coulomb repulsion, which the Hartree term treats only in a mean-field way. It also contains corrections due to correlations resulting from the fermionic nature of electrons: the requirement that the wavefunction be anti-symmetric under the exchange of any two particles imposes a correlation between electrons of parallel spin which, together with the Coulomb correlation, serves to screen the electrostatic repulsion of electrons and lower the interaction energy. For this reason $E_{xc}[n]$ is known as the exchange-correlation functional.

The utility of separating the functional $F[n]$ in equation (2.23) into a sum of three terms can be seen when we notice that the first two terms are easy to calculate, and the hope is that the problematic exchange-correlation term contributes only a small amount to the total energy in many systems of interest, and can be dealt with via simple approximations. We discuss a simple yet highly successful mean-field approximation to this functional in the section on the local density approximation below.

The Kohn-Sham equations

The ground-state energy of the interacting many-body system can be found by minimising equation 2.22 with respect to the density, subject to the constraint of normalisation. Written mathematically, the following condition must be satisfied

$$\frac{\delta}{\delta n(\mathbf{r})} \left(E[n] - \mu \int n(\mathbf{r}) d\mathbf{r} \right) = 0 \quad (2.32)$$

where μ is the Lagrange multiplier used to impose the normalisation constraint. Carrying out this functional derivative yields

$$\frac{\delta T_{KS}[n]}{\delta n(\mathbf{r})} + \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + V_{nuc}(\mathbf{r}) + \frac{\delta E_{xc}[n]}{\delta n(\mathbf{r})} = \mu$$

subject to the normalisation condition.

Similarly, the stationary condition which needs to be satisfied in order to minimise the energy of the system of fictitious particles with respect to the electron density, whose Hamiltonian is given in equation 2.26, is

$$\frac{\delta}{\delta n(\mathbf{r})} \left(E_{KS}[n] - \mu_{KS} \int n(\mathbf{r}) d\mathbf{r} \right) = 0$$

where μ_{KS} is the Lagrange multiplier, again imposing normalisation. This yields the Euler equation

$$\frac{\delta T_{KS}[n]}{\delta n(\mathbf{r})} + V_{KS} = \mu_{KS} \quad (2.33)$$

Now if we compare the Euler equation for the interacting system (2.32) with that of the non-interacting system (2.33), we see that they both give the same ground-state density if the Kohn-Sham potential is

$$V_{KS}[n](\mathbf{r}) = \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + V_{nuc}(\mathbf{r}) + \frac{\delta E_{xc}[n]}{\delta n(\mathbf{r})}$$

to within an additive constant ($= \mu_{KS} - \mu$). This constant has no effect on the calculation of T_{KS} (and the only purpose of $V_{KS}[n]$ is so that we can calculate T_{KS} ,

so we set the constant to zero).

More concisely, this can be written as

$$V_{KS}[n](\mathbf{r}) = V_H[n](\mathbf{r}) + V_{nuc}(\mathbf{r}) + V_{xc}[n](\mathbf{r}) \quad (2.34)$$

where the definitions of V_H and V_{xc} are obvious by comparison with the previous expression. By manipulating equations 2.34, 2.22 and 2.27, one can relate the Kohn-Sham energy $E_{KS}[n]$ to the total energy of the interacting system $E[n]$:

$$E[n] = E_{KS}[n] - E_H[n] - \int d\mathbf{r} n(\mathbf{r}) V_{xc}[n](\mathbf{r}) + E_{xc}[n]$$

since

$$E_H[n] = \frac{1}{2} \int d\mathbf{r} n(\mathbf{r}) V_H[n](\mathbf{r})$$

One can see that although the densities of the fully-interacting and non-interacting systems are the same, their energies are different, and last three terms on the RHS of this expression which represent the difference are commonly termed the ‘double-counting’ terms (even though it is only $E_H[n]$ which is ‘double counted’ - due to the factor of $\frac{1}{2}$ in front of the expression for the Hartree energy).

These equations imply that so long as V_{KS} is constructed according to the prescription in equation (2.34), then the ground-state density of the system of fictitious particles will coincide with the ground-state density of the fully interacting system. This means that the problem reduces to: (1), solving the set of one-electron non-interacting Schrödinger-like equations 2.28 for the Kohn-Sham orbitals ψ_i ; (2), constructing an electron density from these, using

$$n(\mathbf{r}) = 2 \sum_{i=1}^{N/2} |\psi_i|^2;$$

and (3), substituting this $n(\mathbf{r})$ into the expression for the total-energy.

However, the Kohn-Sham potential $V_{KS}[n](\mathbf{r})$, which appears in these non-

interacting Schrödinger equations is itself a functional of the density, and the density constructed at stage (2) is only the ground-state density of the non-interacting system if it is equal to the density used to calculate this Kohn-Sham potential. In general, following the above procedure using an initial guess trial density $\tilde{n}$ used in $V_{KS}[\tilde{n}](\mathbf{r})$, will yield a different density $n(\neq \tilde{n})$, so the procedure will need to be iterated to self-consistency (i.e. until $n = \tilde{n}$) to within an acceptable error (tolerance).

The only problem left is that we don't know the analytic form for the exchange-correlation energy $E_{xc}[n]$. This is usually dealt with by using either the *local density approximation* (LDA) or *generalised gradient approximations* (GGAs). This work uses only LDA.

The local density approximation

The reason why DFT has proved to be so useful is because it is possible to make simple approximations for V_{xc} which turn out to produce remarkably accurate values for the total energies of many many-electron systems. The particular approximation used in this work is called the local density approximation (LDA).

In the LDA, the contribution to the exchange correlation energy from a particular volume element $d\mathbf{r}$ at position $\mathbf{r}$ in space, is taken as the value it would have if all of space was filled with a homogeneous electron gas of density $n(\mathbf{r})$, i.e.

$$E_{xc}[n(\mathbf{r})] = \int d\mathbf{r} n(\mathbf{r}) \epsilon_{xc}(n(\mathbf{r})) \quad (2.35)$$

where $\epsilon_{xc}(n(\mathbf{r}))$ is the energy per electron of the homogenous electron gas of density equal to $n(\mathbf{r})$. The exchange-correlation potential $V_{xc}(\mathbf{r})$ is then

$$V_{xc}[n(\mathbf{r})] = \epsilon_{xc}(n(\mathbf{r})) + n(\mathbf{r}) \left. \frac{d\epsilon_{xc}(n(\mathbf{r}))}{dn} \right|_{n=n(\mathbf{r})} \quad (2.36)$$

Notice that function $\epsilon_{xc}(n(\mathbf{r}))$ is not a functional of $n(\mathbf{r})$, but just a function of the value of n at position $\mathbf{r}$. The function $\epsilon_{xc}(x)$ has been calculated analytically for the homogeneous electron gas [62], and in the high density limit it has be

calculated numerically using quantum Monte Carlo simulations and a few accurate parameterisations exist [63, 64].

In the limit of slowly varying density $n(\mathbf{r})$, the LDA gives the exchange-correlation potential exactly, however most systems of interest have rapidly varying density, so the fact that it can predict many properties of solids to an accuracy of within a few percent is very surprising. The reason why it should produce such accurate energies is yet to be fully understood, however it is thought to be in part because it obeys a sum rule – a normalisation condition on the exchange-correlation hole (xc-hole – the region of depleted electron density around a position $\mathbf{r}$, caused by the presence of an electron at $\mathbf{r}$), which says that an electron scoops out a region of depleted electron density which integrates to a magnitude of exactly one electron – and in part because only the spherical average of the xc-hole is important for the energy. The fact that LDA often performs better than a GGA in which the sum rule has not been enforced, is strong evidence as to the importance of the sum rule. Quantum Monte-Carlo (QMC) studies of the xc-hole [65] have shown that the LDA gets the shape of the xc-hole very wrong in regions of low density, since the LDA scoops out a spatially large spherically symmetric hole (in order that the total amount of charge scooped out is $|e|$) whereas the calculated xc-hole gives much more weight to nearby high density regions. But of course regions of low density don't contribute much to the xc energy, so this is another reason why the LDA approximation gives surprisingly accurate results.

2.3 Periodic supercells

Kohn-Sham DFT gives a way of reducing the many-body problem of interacting electrons, to a problem of non-interacting fermions in an effective potential (represented by the Hamiltonian (2.26)). In order to model bulk crystal however, in which there are a macroscopically large number of electrons moving in the potential of a macroscopically large number of ions, we would need to calculate a huge number of wavefunctions (one for each electron), and we would require an unfeasibly large

basis set in order to expand any one of these wavefunctions. The first of these problems is overcome by performing the calculation on a finite, but periodic system and applying Bloch's theorem to the electronic wavefunctions, outlined in section 2.3.1 below, and the approximation used to overcome the second of these problems are outlined in sections 2.3.2 and 2.3.3 below.

It is also often easier to study small, finite systems by imposing artificial periodicity, so that the simplifying benefits of Bloch's theorem can be exploited. The finite system is contained within a simulation cell which is 'padded' with vacuum which is periodically repeated throughout space – known as a supercell. The supercell must be large enough to ensure that the interaction between periodic images is insignificant.

2.3.1 Bloch's theorem

The electronic wavefunctions of a periodic solid i.e. a solid in which electrons are moving in a static potential $U(\mathbf{r})$, where $U(\mathbf{r}) = U(\mathbf{r} + \mathbf{R})$, for all $\mathbf{R}$ in a Bravais lattice, can be written in the form of a plane wave times a function ($u_{n\mathbf{k}}(\mathbf{r})$) with the periodicity of the Bravais lattice (for a proof of this, see [66]):

$$\psi_{n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{n\mathbf{k}}(\mathbf{r}) \quad (2.37)$$

where

$$u_{n\mathbf{k}}(\mathbf{r}) = u_{n\mathbf{k}}(\mathbf{r} + \mathbf{R}) \quad (2.38)$$

These two equations imply that

$$\psi_{n\mathbf{k}}(\mathbf{r} + \mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}} \psi_{n\mathbf{k}}(\mathbf{r}) \quad (2.39)$$

where the set of complex numbers $\{e^{i\mathbf{k}\cdot\mathbf{R}}\}$ can be viewed as eigenvalues of translation operators $\hat{T}_{\mathbf{R}}$ which commute with the Hamiltonian and are therefore good quantum numbers, making the $\mathbf{k}$ s good quantum vectors. This set of equations is known as Bloch's theorem.

The cell-periodic part of the wavefunction can be expanded in a basis of discrete plane waves (a discrete Fourier transform) so that we obtain

$$\psi_{n\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} c_{n,\mathbf{k}+\mathbf{G}} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} \quad (2.40)$$

where $\mathbf{G} \cdot \mathbf{R}$ is an integer multiple of 2π for all lattice vectors $\mathbf{R}$ i.e. $\mathbf{G}$ is a reciprocal lattice vector.

In an infinitely extending solid, the space of allowable $\mathbf{k}$ -vectors is continuous over the whole of reciprocal space, so the use of Bloch's theorem changes the problem from one in which we require an infinite number of wavefunctions for the infinite number of electrons, to one in which we require a finite number of wavefunctions (equal to the number of electrons in the simulation cell), but at an infinite number of $\mathbf{k}$ points. Another problem is the fact that an infinite number of Fourier components are required to expand the electronic wavefunction (equation 2.40). The problem of infinite number of $\mathbf{k}$ -points is overcome in section 2.3.2, and the problem of infinite $\mathbf{G}$ -vectors overcome in section 2.3.3.

2.3.2 $\mathbf{k}$ -point sampling

Another consequence of Bloch's theorem is that any point in $\mathbf{k}$ space which lies outside the first Brillouin zone is equivalent to a $\mathbf{k}$ point inside the first Brillouin zone. This is shown by considering two points in $\mathbf{k}$ -space, $\mathbf{k}$ and $\mathbf{k}'$ related by $\mathbf{k}' = \mathbf{k} + \mathbf{G}$.

$$\psi_{n\mathbf{k}'}(\mathbf{r}) = e^{i\mathbf{k}'\cdot\mathbf{r}} u_{n\mathbf{k}'}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} [u_{n\mathbf{k}'}(\mathbf{r}) e^{i\mathbf{G}\cdot\mathbf{r}}] = e^{i\mathbf{k}\cdot\mathbf{r}} \tilde{u}(\mathbf{r}) = \psi_{n'\mathbf{k}}(\mathbf{r}) \quad (2.41)$$

since $u_{n\mathbf{k}'}(\mathbf{r}) e^{i\mathbf{G}\cdot\mathbf{r}}$ is a cell periodic function so that $e^{i\mathbf{k}\cdot\mathbf{r}} [u_{n\mathbf{k}'}(\mathbf{r}) e^{i\mathbf{G}\cdot\mathbf{r}}]$ is a Bloch state with wavevector $\mathbf{k}$. This means that a state with a $\mathbf{k}$ vector outside the first Brillouin zone is 'aliased' to a state lying within the first Brillouin zone, but at a different n . The consequence of this observation is that we only need be concerned with the

occupied states at each $\mathbf{k}$ point within the first Brillouin zone. This is a step towards making the problem tractable, however there is still the problem that there are an infinite number of $\mathbf{k}$ point in the first Brillouin zone. The occupied states at each $\mathbf{k}$ point contribute to the Kohn-Sham potential meaning that, in principle, an infinite number of calculations are needed to compute this potential.

In practice however, since the electronic wavefunctions vary smoothly over the first Brillouin zone [67], the electronic wavefunctions at $\mathbf{k}$ points that are very close together will be almost identical, so it is possible to represent a wavefunction over a finite region of $\mathbf{k}$ -space using only one $\mathbf{k}$ -point, and therefore one can represent the wavefunction $\psi_{n\mathbf{k}}$ on a finite set of, well chosen grid points. The less rapidly varying the wavefunctions are in $\mathbf{k}$ -space, the fewer the number of grid points which need to be sampled in order to get an accurate description of the wavefunctions. The Kohn-Sham potential, and total-energy are more difficult to calculate if the system is metallic because a dense set of $\mathbf{k}$ points is needed to distinguish accurately the position and shape of the Fermi surface. Methods for making efficient choices of these $\mathbf{k}$ -points have been developed [68].

A corollary to the last point is that, in general, the smaller the first Brillouin zone is, the less $\psi_{n\mathbf{k}}$ will vary across it, and therefore we can get away with sampling many fewer points. The volume of the first Brillouin zone is inversely proportional to the volume of the simulation cell, therefore for very large systems, sampling at only one point in the first Brillouin zone (usually the point $\mathbf{k}=(0,0,0)$, the Γ point), is sufficient.

2.3.3 The plane-wave basis set

In practice, continuous quantities (such as the electronic density) in the simulation cell need to be represented on a grid of points in real-space. The density of this grid of points determines the maximum frequency of oscillation of any real-space quantity. By discretising real-space, and using discrete Fourier transforms, one imposes a long range periodicity in reciprocal space, with a period that is inversely propor-

tional to the distance between grid points in real-space. This period imposes an upper limit on the magnitude of $\mathbf{G}$ -vectors which is representable on the reciprocal space grid. The approximation made when real-space is discretised, is to truncate the sum in equation 2.40 so that only $\mathbf{G}$ -vectors of magnitude smaller than the inverse of the real-space grid spacing are included. Fortunately, one can justify this approximation on physical grounds, and one can easily tune the approximation to obtain an arbitrary degree of accuracy just by changing the density of real-space grid points.

By operating on the wavefunctions in equation 2.40 with the kinetic energy operator, we find that the kinetic energy term of the Hamiltonian is proportional to $|\mathbf{k} + \mathbf{G}|^2$. One would expect then, that at high enough values of $|\mathbf{k} + \mathbf{G}|$, the expansion coefficients $c_{n,\mathbf{k}+\mathbf{G}}$, will become very small, since such contributions would cost a large kinetic energy. This means that imposing a cutoff on the number of $\mathbf{G}$ -vectors to include in the basis will not be a bad approximation, so long as the cutoff is large enough. In practice, the cutoff is placed on the quantity $|\mathbf{k} + \mathbf{G}|$ and is $|\mathbf{k} + \mathbf{G}| < G_{max}$. This results in a real-space grid spacing of $2\pi/G_{max}$ (however, in plane-wave pseudopotential methods, this real-space grid is actually doubled to avoid aliasing when squaring the wavefunctions to calculate the density).

Most plane-wave DFT methods use the parameter E_{max} , which is related to G_{max} by $E_{max} = \frac{\hbar^2}{2m} G_{max}^2$.

2.4 Pseudopotentials

A further approximation which is used is based on the observation that in many atoms, the electrons lying near the nucleus can be extremely tightly bound to the nucleus, with binding energies reaching hundreds of electron volts. Because of these extremely high binding energies, they can be assumed to be relatively unaffected by the surrounding chemical environment and can be assumed to move with the nucleus at all times.

It is therefore convenient to replace the core electrons and the nuclei with

a much weaker effective *pseudopotential* [69] that acts on a set of pseudo wave functions, illustrated schematically in Figure 2.1

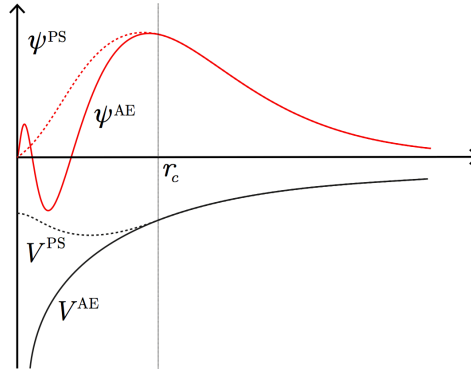


Figure 2.1 – A schematic illustration of a pseudopotential V^{PS} and pseudo-wavefunction ψ^{PS} (dashed black and red lines respectively) compared to the all-electron (AE) system on which they are based (solid lines). At radii larger than r_c both the pseudopotential and pseudo-wavefunction exactly reproduce the AE system.

The pseudopotentials are constructed so that the pseudo wave-function has no radial nodes in the core region, and so that the pseudo wave-functions exactly match the all-electron wavefunction outside the core region.

The great advantage of using this approximation is that if we wish to use Bloch's theorem to expand wavefunctions using a discrete set of plane waves, a very large number of plane waves would be needed in order to expand the wavefunctions of all the electrons. The reason for this is twofold: (1) The nuclear potential can become very steep near the nucleus of an atom ($\sim \frac{1}{r}$) and this strong confinement of the electrons can cause the electronic wavefunctions, by the Heisenberg uncertainty principle, to have high kinetic energy (i.e. strong curvature of the wavefunctions) so that a large basis set is needed in order to represent these wavefunctions; and more significantly (2) the valence electrons, in order to maintain orthogonality with the core electrons (the Pauli exclusion principle), must oscillate rapidly within the core region, which again means that a very large basis set is required. Pseudo-wavefunctions which result from pseudopotentials have no radial nodes in the core region, and therefore require a much smaller basis set (i.e. a much smaller cutoff-

energy) to represent them.

There is a subcategory of pseudopotentials that are very widely used, due to their good transferability to many chemical environments, known as norm-conserving pseudopotentials. Hamann, Schluter and Chiang derived four criteria for creating these norm-conserving pseudopotentials [70]:

1. That the all-electron and pseudo valence eigenvalues agree for the chosen reference system;
2. That the all electron and pseudo valence wavefunctions agree beyond a chosen cutoff radius r_c ;
3. That the pseudo valence wavefunction is nodeless within the cutoff-radius ;
4. that:

$$\int_0^{r_c} |\psi_i^{PS}|^2 r^2 dr = \int_0^{r_c} |\psi_i^{ae}|^2 r^2 dr \quad (2.42)$$

where ψ_i^{ae} are the states of the all-electron system.

The general procedure for constructing these pseudo-potentials is as follows [71]: first, a suitable pseudo wavefunction is constructed which satisfies the norm-conservation criteria above. Next, the pseudo-wave function is used to generate a ‘screened’ pseudopotential for each angular momentum state l , using the inverted radial Schrödinger equation

$$V_{scr,l}^{PS} = \varepsilon_l - \frac{l(l+1)}{2r^2} + \frac{1}{2rR_l^{PS}(r)} \frac{d^2}{dr^2} [rR_l^{PS}(r)] \quad (2.43)$$

and finally, the screening caused by the valence electrons is removed by subtracting the Hartree ($V_H^{PS}(r)$) and exchange-correlation ($V_{xc}^{PS}(r)$) potentials calculated from the valence pseudo-wavefunctions, leaving

$$V_{ion,l}^{PS}(r) = V_{scr,l}^{PS}(r) - V_H^{PS}(r) - V_{xc}^{PS}(r) \quad (2.44)$$

The pseudo-wavefunctions are non-unique, and this extra variational freedom is used to try to construct a pseudopotential which is as smooth as possible inside the core region. There are many different schemes for constructing such pseudo-wave functions.

These equations clearly show that the potential is different for different angular momenta. In general, the total pseudopotential operator can be defined as

$$\hat{V}_{ion}^{PS}(r) = V_{local}(r) + \sum_l \hat{V}_{nonlocal,l}(r) \quad (2.45)$$

where V_{local} is a local potential and

$$\hat{V}_{nonlocal,l}(r) = [V_{ion,l}^{PS}(r) - V_{local}(r)]\hat{P}_l \quad (2.46)$$

is the semi-local part, with $\hat{P}_l = \sum_{m=-l}^l |lm\rangle\langle lm|$ being projection operators which project out the l th angular momentum component from the wavefunction.

The local potential can be arbitrarily chosen, but in practice, the semi-local part is truncated at some value of l , and the local potential is chosen to get the scattering for the higher angular momentum components roughly correct. Kleinman and Bylander showed [72] that this semi-local potential can be transformed into separable form as

$$\hat{V}_{nonlocal,l} = \sum_{m=-l}^l \frac{|\hat{V}_{nonlocal,l}\phi_{lm}\rangle\langle\phi_{lm}\hat{V}_{nonlocal,l}|}{\langle\phi_{lm}|\hat{V}_{nonlocal,l}|\phi_{lm}\rangle} \quad (2.47)$$

where $|\phi_{lm}\rangle$ is an eigenstate of the atom pseudo Hamiltonian. This form reduces the number of integrals that need to be carried out to a level that scales linearly with the size of the basis set.

2.4.1 Non-linear core corrections

The norm-conservation criterion ensures a reasonable degree of transferability of pseudopotentials. However, the general scheme for constructing pseudopotentials

described in the previous section makes an approximation which can be the source of poor transferability in many cases. The density $n(\mathbf{r})$ can be written as a sum of core and valence contributions. In simple pseudopotential methods, the all-electron density is replaced by the density associated with the pseudovalence orbitals $n^{PS}(\mathbf{r})$ only. This implies a linearisation of the energy functional $E[n]$, however, this functional contains the exchange-correlation functional $E_{xc}[n]$, which is a non-linear functional. Thus, this linearisation is only valid if the valence charge density in the core region does not change much in different chemical environments, which isn't always the case.

Louie *et al* [73] resolved this problem in a scheme which takes the non-linear nature of $E_{xc}[n]$ into account by adding the core charge density explicitly i.e. replacing $n^{PS}(\mathbf{r})$ with $n^{PS}(\mathbf{r}) + n^{core}(\mathbf{r})$. This is known as a non-linear core correction (NLCC). The core charge is calculated once, when the pseudopotential is generated, and it is then stored and used to reconstruct the full exchange-correlation potential during the DFT calculation, which is used to generate the corrected exchange correlation energy using

$$E_{xc}[n^{PS}; n^{core}] = E_{xc}[n^{PS} + n^{core}] - E_{xc}[n^{core}]$$

2.5 Summary

In this chapter we have described the basic theorems and approximations which form the basis of the plane-wave DFT approach to solving the many-body Schrödinger equation. The first approximation that was made was the adiabatic (or Born-Oppenheimer) approximation, which separates out nuclear and electronic motion and therefore greatly simplifies Schrödinger equation. The theorems which provide a rigorous grounding for the DFT approach are the two Hohenberg-Kohn theorems, which imply that all properties of the system are completely determined by the ground-state density, and that there exists an energy functional of the density, which is minimal for the ground state density. Kohn and Sham proposed a form for

this functional by replacing the many-body problem with an auxiliary independent-particle problem, where the electron-electron interaction of the full problem is included, through the effective potential of this auxiliary system. This formulation is powerful because it makes possible mean-field approximations to the electron-electron interaction that have proven very successful.

The use of periodic boundary conditions, Bloch's theorem and plane-waves, together with Brillouin zone sampling, reduces the problem of solving the Kohn-Sham equations for an infinite crystal, with an infinite number of electronic states, to a problem of calculating a finite number of states at a finite number of points in the first Brillouin zone, and the use of the supercell approximation allows us to extend these techniques to the study of isolated systems.

The pseudopotential approximation offers a further computational saving. Assuming that electrons in the atomic core shells are so tightly bound to the nucleus that they are not significantly perturbed by altering the chemical environment of the atom, pseudopotentials allow the removal of these electrons from calculations without a significant loss in accuracy.

The computational effort required to carry out calculations of this type scales with order N^3 , where N is the number of atoms, or the size of the supercell. Consequently, system sizes are limited to ~ 1000 atoms. In the next section we discuss linear-scaling methods, which allow the study of much large systems, and are better placed to take full advantage of improvements in computational power in the future.

3

Linear scaling

Conventional plane-wave DFT [74] has an $O(N^3)$ scaling. This is because there are $O(N)$ extended Kohn-Sham orbitals, which are of size $O(N)$, and they have to remain orthogonal with $O(N)$ other orbitals. It is the extended nature of the Kohn-Sham orbitals which forces us, firstly to have to use the full number of basis functions to represent each orbital, and secondly to have to explicitly check the orthogonality of each orbital against every other one. Linear scaling methods take advantage of the ‘nearsightedness’ [75, 76] inherent in quantum many-body systems in equilibrium, by constructing a scheme based on a truncated density matrix. In this chapter we will outline the ONETEP methodology [77, 78] for achieving linear scaling.

3.1 The density matrix

ONETEP can be classified as a *density matrix method* [79], which is a group of methods in which the energy functional is minimised with respect to the elements of the single particle density matrix, usually represented in terms of a set of localised functions. We start by introducing the single-particle¹ density operator for a many body system which can be written in terms of the orthonormal single particle

¹The single-particle approximation is made throughout, as we follow Kohn-Sham DFT

eigenstates of the Kohn-Sham Hamiltonian $\{\psi_i\}$:

$$\hat{\rho} = \sum_i f_i |\psi_i\rangle\langle\psi_i| \quad (3.1)$$

so that in the position representation, the single-particle density matrix becomes

$$\rho(\mathbf{r}, \mathbf{r}') = \langle \mathbf{r} | \hat{\rho} | \mathbf{r}' \rangle = \sum_i f_i \psi_i(\mathbf{r}) \psi_i^*(\mathbf{r}') \quad (3.2)$$

where $\{f_i\}$ are occupation numbers which, at temperature $T = 0$ K, take the value 1 for states with energy below the chemical potential, and 0 for states above. The electron density is defined as

$$n(\mathbf{r}) = 2\rho(\mathbf{r}, \mathbf{r}) \quad (3.3)$$

where the factor of 2 is because of spin-degeneracy. In this representation, the non-interacting Kohn-Sham kinetic energy, which appears in equation 2.25, can be written

$$T_{KS}[n] = 2\text{Tr}(\hat{\rho}\hat{T}_{KS}) \quad (3.4)$$

where $\hat{T}_{KS} = -\frac{1}{2}\nabla^2$, and similarly for the interaction energy of the electrons with the nuclei $E_{en}[n]$ (in practice $E_{en}[n]$ is the interaction energy of electrons with the pseudopotentials) is just

$$E_{en}[n] = 2\text{Tr}[\hat{\rho}\hat{V}_{ion}^{PS}] \quad (3.5)$$

where $\hat{V}_{ion}^{PS}$ is defined in equation 2.45.

The other terms in the Kohn-Sham equations, E_H and E_{xc} can be calculated from n given by equation 3.3.

3.2 Constraints on the density matrix

3.2.1 Normalisation

The total number of electrons is given by

$$N = \int n(\mathbf{r}) d\mathbf{r} = 2 \int \rho(\mathbf{r}, \mathbf{r}) d\mathbf{r} = 2 \sum_i f_i = 2 \text{Tr}[\hat{\rho}] \quad (3.6)$$

In the canonical ensemble formalism, N is required to remain constant. This constraint can be imposed in the usual way by introducing a Lagrange multiplier λ and performing an unconstrained minimisation of the functional

$$L[n] = E[n] + \lambda(2\text{Tr}[\hat{\rho}] - N) \quad (3.7)$$

3.2.2 Idempotency

In the ground-state, the single particle density operator must be idempotent (i.e. $\hat{\rho}^2 = \hat{\rho}$) which forces the occupancy numbers $\{f_i\}$ to be either 0 or 1, enforcing the Pauli exclusion principle and the orthonormality of Kohn-Sham orbitals simultaneously:

$$\begin{aligned} \rho^2(\mathbf{r}, \mathbf{r}') &= \int d\mathbf{r}'' \rho(\mathbf{r}, \mathbf{r}'') \rho(\mathbf{r}'', \mathbf{r}') = \sum_{ij} f_i f_j \psi_i(\mathbf{r}) \psi_j^*(\mathbf{r}') \int d\mathbf{r}'' \psi_i^*(\mathbf{r}'') \psi_j(\mathbf{r}'') \\ &= \sum_{ij} f_i f_j \psi_i(\mathbf{r}) \psi_j^*(\mathbf{r}') \delta_{ij} = \sum_i f_i^2 \psi_i(\mathbf{r}) \psi_i^*(\mathbf{r}') = \rho(\mathbf{r}, \mathbf{r}'), \quad \text{iff } f_i^2 = f_i \quad \forall i \end{aligned} \quad (3.8)$$

where i labels all the permutations of bands and wave vector pairs (i.e. bands n and wave vectors $\mathbf{k}$). This constraint cannot be imposed explicitly without incurring a scaling larger than linear, so it needs to be imposed some other way. We outline how idempotency is imposed in ONETEP in sections 3.6 and 3.9.

3.3 Locality in quantum systems

Linear scaling methods must be formulated in a way which enables one to exploit the locality or “nearsightedness” [75, 76] inherent in quantum many-body systems. In plane-wave DFT methods, the electronic states are described in terms of Kohn-Sham orbitals, which are Bloch states which extend across the whole system. The ability to exploit the phenomenon of locality is lost in this formulation. Wannier showed [80] that one could obtain a set of functions $w_{n\mathbf{R}}$ which is localised in space, by performing a unitary transformation (in particular, a Fourier transform) on the Bloch states. Since the transformation from Bloch states to Wannier functions is a unitary one, Wannier functions also form an orthonormal basis set for the same Hilbert space, and the density matrix can be written in terms of these Wannier functions as

$$\rho(\mathbf{r}, \mathbf{r}') = \sum_{n\mathbf{R}} f_n w_{n\mathbf{R}}(\mathbf{r}) w_{n\mathbf{R}}^*(\mathbf{r}') \quad (3.9)$$

It has been shown analytically in one dimension [81] that in a periodic system in the tight-binding limit, these Wannier functions are exponentially localised with an exponent related to the square root of the band gap. More recent numerical and analytical studies [82, 83] have shown exponential localisation for Wannier functions in three dimensions in insulators, with an exponent related to the size of the band-gap. Wannier functions in metals are found to have polynomial locality.

Within this Wannier function representation of the density matrix, one can see that it might be possible to exploit the ‘nearsightedness’ of the density matrix by imposing a cutoff radius on these Wannier functions so that the Wannier functions are artificially set to zero after some radius r_c around the average position of that function, so that the elements of the density matrix in equation (70) which have $|\mathbf{r} - \mathbf{r}'| > r_c$ are zero. That way, the information stored in the density matrix scales proportionately with the number of electrons in the system and not with its square – as in the case of a non-truncated density matrix.

3.4 Non-orthogonal orbitals and the density kernel

Some linear-scaling methods use a fixed set of local atomic-like orbitals as basis functions with which to represent the density matrix [84–89], or Gaussians [90, 91] or other fixed functions. The approach that is adopted in the ONETEP method instead optimises the local functions [92] in the environment of system being studied, which ensures that a minimal number of functions need be used, reducing the computational cost due to an artificially large matrix representation, but increasing the computational cost due to the need to optimise the functions. These localised functions $\{\phi_\alpha(\mathbf{r})\}$ [93], are very closely related to the extended Kohn-Sham orbitals $\{\psi_i(\mathbf{r})\}$. They are defined by

$$\psi_i(\mathbf{r}) = \phi_\alpha(\mathbf{r})M_i^\alpha \quad (3.10)$$

where $\mathbf{M}$ is a linear transformation. Repeated greek indices imply a summation over that index. Substituting equation 3.10 into the orthogonality relation $\langle\psi_i|\psi_j\rangle = \delta_{ij}$ yields

$$(M^\dagger)_i^\alpha S_{\alpha\beta} M_j^\beta = \delta_{ij} \quad (3.11)$$

where $\mathbf{S}$ is the overlap matrix between the localised functions

$$S_{\alpha\beta} = \langle\phi_\alpha|\phi_\beta\rangle \quad (3.12)$$

A set of dual vectors $\{\phi^\alpha\}$ are also defined as

$$|\phi^\alpha\rangle = |\phi_\beta\rangle(S^{-1})^{\beta\alpha} \quad (3.13)$$

so that the contravariant duals are orthogonal to the covariant orbitals

$$\langle\phi^\alpha|\phi_\beta\rangle = \delta_\beta^\alpha \quad (3.14)$$

The $\mathbf{S}$ matrix acts as a metric for the non-orthogonal basis. The orthogonal Kohn-Sham basis could be cast in a similar form, but the metric would be the identity matrix, so there would be no distinction between the orbitals and their duals. Substituting equation (3.10) into the definition of the density matrix in the Kohn-Sham position representation (equation 3.2) one obtains the density matrix in terms of these highly localised orbitals

$$\rho(\mathbf{r}, \mathbf{r}') = \phi_\alpha(\mathbf{r}) K^{\alpha\beta} \phi_\beta^*(\mathbf{r}') \quad (3.15)$$

where $K^{\alpha\beta}$ is called the density kernel [79] and is defined as

$$K^{\alpha\beta} = \langle \phi^\alpha | \hat{\rho} | \phi^\beta \rangle = \sum_i M_i^\alpha f_i (M_i^\dagger)^\beta \quad (3.16)$$

In ONETEP, only the Γ -point is sampled in the Brillouin zone since, as we saw in section 2.3.2, the volume of the first Brillouin zone is inversely proportional to the volume of the simulation cell, so for very large systems (for which ONETEP is designed), wavefunctions vary very little with $\mathbf{k}$ over the first Brillouin zone, and sampling at only one point in the first Brillouin zone, is sufficient to accurately sample the wavefunctions in reciprocal space.

The advantage of this is that it allows us to insist that the local orbitals are always real without loss of generality, making the complex conjugate symbol in equation 3.15 redundant.

The localised orbitals $\{\phi_\alpha\}$ are non-orthogonal, Wannier like orbitals which have been given the name non-orthogonal generalised Wannier functions, or NGWFs [92].

Boundary conditions are imposed on the NGWFs so that they are localised within some radius r_{cut} . Since the NGWFs are non-orthogonal, truncating them alone does not result in a strictly localised density matrix. A cutoff is directly imposed on the density kernel $\mathbf{K}$ so that the elements of $\mathbf{K}$ corresponding to localised orbitals $\phi_\alpha(\mathbf{r})$ and $\phi_\beta(\mathbf{r}')$ are set to zero when the distance between $\mathbf{r}$ and $\mathbf{r}'$ is greater

than a cutoff length r_K . These cutoffs on the NGWF and kernel ensure that the information contained in the density matrix scales linearly with system size.

The localised nature of these NGWFs, and the need to optimise them in the calculation, requires a localised primitive basis set in order to represent them. The basis set used to represent these orbitals is described in section 3.7.

3.5 Fractional occupancies and idempotency

In section 3.2.2 we found that it is necessary that the density matrix be idempotent if the occupancies are to remain physical (i.e. 0 or 1). Iterative methods (described in section 3.6) have been developed in which the density matrix can approach having perfect idempotency, however the use of these methods requires that we have an energy functional that can cope with fractional occupancies. Fractional occupancies cannot be treated within the Kohn-Sham energy functional $E[n]$ since the Kohn-Sham kinetic energy term assumes that either a state is fully occupied or not. A generalisation of the Kohn-Sham energy functional, $E_J[n]$ was proposed by Janak [94], which is capable of treating fractional occupancies. He defines a new kinetic energy term

$$T_J[n] = 2 \sum_{i=1}^M f_i \int d\mathbf{r} \psi_i^*(\mathbf{r}) \left(-\frac{1}{2} \nabla^2\right) \psi_i(\mathbf{r}) \quad (3.17)$$

where the sum runs over states up to $i = M$, where in general $M \geq \frac{N}{2}$ and the occupancies are in the range $0 \leq f_i \leq 1$. Because of the new kinetic energy term, a new exchange-correlation term $E_{xc,J}[n]$ is needed to incorporate it. The minimisation of the new functional $E_J[n]$ must be performed with respect to both the occupancies $\{f_i\}$ and the orbitals $\{\psi_i\}$ and subject to the usual constraint of fixing the total number of electrons to N (normalisation). The Janak functional $E_J[n]$ is only variational with respect to changes in $\{f_i\}$ when all the values of f_i are in the interval $[0, 1]$, otherwise the energy can be lowered without bound by letting $f_i \rightarrow \infty$ for states with energy below the chemical potential and letting $f_i \rightarrow -\infty$ for states above

the chemical potential (whilst the normalisation remains correct). So the eigenvalues of the density matrix (the occupancies) must be kept in the interval $[0,1]$ to stop the density matrix following these unphysical “run-away” solutions. This is ensured in the ONETEP method by using a *penalty functional* [95] for density matrices which are very far from idempotent, and iterative methods incorporating the *McWeeny purification transformation* [96] for density matrices which are closer to being idempotent. The latter of these is outlined in the next section. Just like for the Kohn-Sham energy functional, the Janak functional is minimised when $f_i = 1$ for all states below the Fermi energy, and $f_i = 0$ for all states above the Fermi energy, which is consistent with what we expect at ‘absolute zero’ temperature.

3.6 Imposing idempotency

In order to keep the occupancies in the appropriate interval to avoid runaway solutions of the Janak functional, and to ensure that the density matrix converges to an idempotent one as quickly as possible, it is necessary to impose idempotency ‘weakly’. In linear-scaling density-matrix schemes, we only have access to the matrix representation $\mathbf{K}$ of the density operator and not the eigenvalues themselves, so we need a scheme for imposing idempotency that can deal with density matrices in a non-orthogonal representation in linear-scaling effort.

This is achieved using an iterative method related to that proposed by McWeeny [79]. Consider the transformation

$$\boldsymbol{\rho}_{n+1} = W(\boldsymbol{\rho}_n) = 3\boldsymbol{\rho}_n^2 - 2\boldsymbol{\rho}_n^3 \quad (3.18)$$

where $\boldsymbol{\rho}$ is the density operator in matrix representation in an orthogonal basis. Since it is clear that $\boldsymbol{\rho}_{n+1}$ and $\boldsymbol{\rho}_n$ commute, we can get an idea of what this transformation does by looking at the common diagonal representation of these matrices, where the occupancies are the diagonal elements. In this case the relationship above can be expressed in terms of these occupancies

$$f_{n+1} = W(f_n) = 3f_n^2 - 2f_n^3 \quad (3.19)$$

So long as the occupancies lie in the interval $[-\frac{1-\sqrt{5}}{2}, \frac{1+\sqrt{5}}{2}]$, ρ_{n+1} will be more idempotent than ρ_n , and furthermore, if all of the occupancies are in the range $[-\frac{1}{2}, \frac{3}{2}]$ then the eigenvalues of ρ_{n+1} will be in the range $[0, 1]$, which is what is required to avoid runaway solutions. It also ensures that eigenvalues around zero converge to zero quadratically, and eigenvalues around one converge to one in the same way. Li, Nunes and Vanderbilt [96] developed an $O(N)$ scheme which incorporates this McWeeny purification transformation to imposing idempotency whilst finding the ground state total-energy. This method is described in section 3.9.

3.7 The psinc basis set

The optimisation of localised NGWFs during the calculation requires that the orbitals be expanded in terms of a primitive basis of localised functions.

The choice of basis in ONETEP is a set of real, periodic, bandwidth-limited delta-functions $\{D_i(\mathbf{r})\}$, which are referred to as *psinc* functions (periodic cardinal sine functions) [92, 97], so that an NGWF can be written

$$\phi_\alpha(\mathbf{r}) = \sum_i D_i(\mathbf{r}) C_{i\alpha} \quad (3.20)$$

where $C_{i\alpha}$ are expansion coefficients.

If $\{D_i(\mathbf{r})\}$ were simply a set of delta functions making up a grid, error would be incurred on Fourier transform because a delta function cannot be represented faithfully in reciprocal space, since it has infinite bandwidth. Additionally, the delta functions are undesirable since they can't be differentiated. The basis functions used in ONETEP are *coarse-sampled* and *bandwidth-limited* delta functions defined

$$D_i = \frac{1}{M} \sum_{\mathbf{G}}^{\mathbf{G}_{max}} e^{i\mathbf{G} \cdot \mathbf{r}} \quad (3.21)$$

where the sum is over all the reciprocal lattice vectors of the simulation cell, up to and including $\mathbf{G}_{max}$, and M is the total number of grid points. Just as in plane-wave DFT, the value of $\mathbf{G}_{max}$ is determined by an energy cutoff parameter, and the value of M is chosen such that there is one function centred on each grid point, which takes a value of unity on that grid point and zero on every other grid point (see Figure 3.1). Since this is a bandwidth limited function, it must be periodic, and this construction ensures that its period coincides with the length of the unit cell so that its periodic image due to the Born-Von Kármán boundary conditions produces no discontinuity, even in any of the derivatives, at the supercell boundary. The value of the functions vary between grid points, however, in order to ensure that these values remain real, the number of grid points in a unit cell along each dimension must be odd.

These psinc functions are related to plane-waves by a Fourier transform, which enables the kinetic energy to be calculated as accurately and efficiently as in plane-wave DFT [92]. We describe how in section 3.8.

Due to this formal equivalence between conventional plane-wave DFT methods and the ONETEP method it is possible to improve the accuracy up to full plane-wave accuracy by increasing $\mathbf{G}_{max}$, $r_{\mathbf{K}}$ and r_{cut} .

3.8 FFT box technique

The calculation of the kinetic energy for each NGWF involves expanding the NGWF in terms of the psinc basis functions, then carrying out a fast Fourier transform (FFT), multiplying by the wave-vector squared, then carrying out another FFT back to real space. The FFTs required are global operations in which every point in the system is involved so that calculating the kinetic energy for a single NGWF scales as $O(M \log M)$ (where M is the number of grid points, which is roughly proportional to the number of atoms, at least for a bulk solid), making the calculation for the whole system scale higher than $O(M^2)$.

This problem is overcome in ONETEP by exploiting the strict locality of

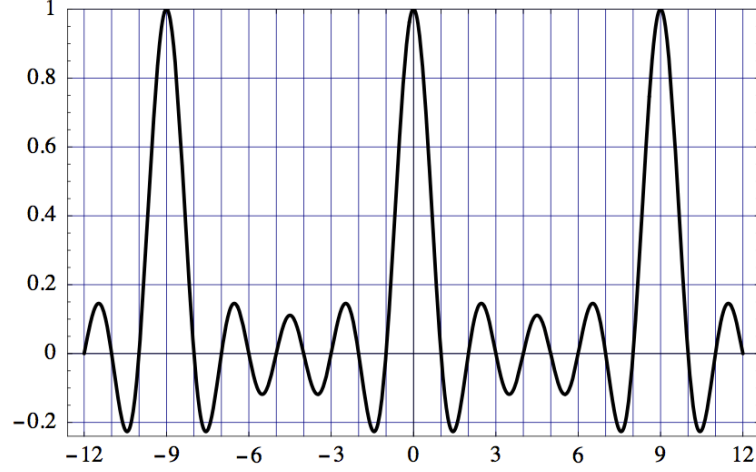


Figure 3.1 – One dimensional psinc function with $M=9$. The values are unity on the grid point on which it is centred and zero on all other grid points within the cell (which is $M=9$ grid points wide). From Ref. [5]

the NGWFs by using only the subspace of the full psinc basis associated with a particular NGWF and its overlapping neighbours, making the calculations of the kinetic energy of each NGWF, for example, independent of system size – thereby recovering linear scaling. This is achieved in the ONETEP method by the *FFT box technique* [98] in which, for each NGWF, FFTs and other calculations are performed within a box whose sides are parallel to the sides of the simulation cell, and whose dimensions are just big enough to fit any of the NGWFs and all of its overlapping neighbours within it. Its dimensions are set at the beginning of the calculation and are universal throughout the calculation. It has the same spacing of grid points as the simulation cell, and its origin always lies on a grid point which in general is different for each matrix element. The details of this construction are important in order to guarantee the consistent action of operators and the hermiticity of the calculated Hamiltonian.

The approximation made by using this FFT box method is effectively to coarsely sample the frequency components of the NGWF in reciprocal space. But since NGWFs are strictly localised in real space, we expect them to be broad and smooth in reciprocal space, meaning that coarse sampling is likely to be, and in fact

turns out to be, a good approximation. The error due to this approximation can be made arbitrarily small, if need be, by increasing the size of the FFT box.

The FFT box technique is also used in order to construct the density from the NGWFs in linear scaling effort. The density ($n(\mathbf{r}) = 2K^{\alpha\beta}\phi_{\alpha}(\mathbf{r})\phi_{\beta}(\mathbf{r})$) must be constructed on a grid which is at least twice as fine as the grid on which the NGWFs are represented to avoid aliasing: the density is proportional to the product of the squares of the NGWFs, meaning that it contains Fourier components which are up to twice as big as those of the NGWFs. The individual contributions to the charge density $\phi_{\alpha}(\mathbf{r})\phi_{\beta}(\mathbf{r})$ are localised in real space and thus can be calculated on an FFT grid. A given pair of overlapping NGWFs are projected into their FFT box and then Fourier interpolated onto a finer grid (of at least twice the resolution) to avoid aliasing i.e. the NGWFs are Fourier transformed, then padded with zeros in reciprocal space, then Fourier transformed back into real space, but on a finer grid. The NGWFs are then multiplied together and projected from their FFT boxes onto the full simulation cell fine grid to construct the density.

Sometimes matrix elements of the form $\langle\phi_{\alpha}|\hat{O}|\phi_{\beta}\rangle$ need to be calculated, where $\hat{O}\phi_{\beta}$ is delocalised over the whole FFT box, for example matrix elements of the kinetic energy operator. Such integrals are calculated simply by taking the dot product between psinc grid points in common between ϕ_{α} and $\hat{O}\phi_{\beta}$, and therefore only points within the localisation sphere of ϕ_{α} within the FFT box are used [99].

The fact that FFTs are not done over the full simulation cell, but only within boxes located around atoms, means that large regions of vacuum can be added to the simulation cell for near to no computational cost. This means that the cross-over point (i.e. the point at which it becomes computationally more efficient to use a linear-scaling method than a traditional cubic-scaling method), is particularly low for low dimensional structures like molecules, nanorods etc, in which a substantial amount of vacuum would need to exist in the supercell.

3.9 Finding the ground-state energy

3.9.1 Functionals to be minimised

The function of the ONETEP code is to find the ground-state total energy E_0 , by minimising the total energy functional with respect to all the degrees of freedom of the density (i.e. the density kernel and the NGWFs)

$$E_0 = \min_{\{K^{\alpha\beta}\}, \{\phi_\alpha(\mathbf{r})\}} E[\{K^{\alpha\beta}\}, \{\phi_\alpha(\mathbf{r})\}] \quad (3.22)$$

subject to a normalisation constraint and subject to the density matrix remaining idempotent. Notice that since the total ground-state energy depends only variationally on the density (see section 2.1.3), so it must depend variationally on the density kernel (if weakly idempotent) and NGWFs too.

Li, Nunes and Vanderbilt [96] suggested a convenient method by which this minimisation could be carried out, with the necessary constraints of normalisation and idempotency, while retaining $O(N)$ scaling. At a fixed chemical potential μ , minimising the grand potential

$$\Omega[\rho] = E - \mu N_e = \text{tr}[\rho(H - \mu)] \quad (3.23)$$

with respect to ρ without constraints gives the same minimum as the total-energy functional under the normalisation constraint. With this in mind, they constructed a new functional Ω_{LNV} , which, when minimised without constraint, imposes the idempotency condition as well as normalisation. The functional suggested makes use of McWeeny's 'purification transformation' discussed in section 3.6, and is

$$\Omega_{LNV}[\rho] = \text{tr}[(3\rho^2 - 2\rho^3)(H - \mu)] \quad (3.24)$$

Adapting this method for non-orthogonal orbitals [100], the functional becomes (in

ONETEP's notation)

$$\Omega_{LNV}[\mathbf{K}] = 2tr[(3\mathbf{KSK} - 2\mathbf{KSKSK})(\mathbf{H} - \mu\mathbf{S})] \quad (3.25)$$

where $\mathbf{K}$ is the density kernel, which is just the density matrix in the representation of the non-orthogonal orbitals and their duals, and $\mathbf{S}$ is the overlap matrix defined in equation 3.12. Notice that if we work in an orthogonal basis set which diagonalises the Hamiltonian, $\mathbf{S}$ becomes the identity matrix, $\mathbf{K}$ becomes the density matrix, and we recover equation 3.24.

An alternative functional was suggested by Millam and Scuseria [101] which also imposes idempotency (again via the ‘purification transformation’) and normalisation on the density matrix:

$$\Omega_{MS}[\rho] = tr[(3\rho^2 - 2\rho^3)H] + \mu'(tr[\rho] - N_e) \quad (3.26)$$

This functional has also been implemented as an option in ONETEP.

If, however, the density matrix is very far from being idempotent (i.e. the density matrix has eigenvalues outside the range $[-\frac{1-\sqrt{5}}{2}, \frac{1+\sqrt{5}}{2}]$), the McWeeny purification transformation is not convergent and another method must be used which is globally convergent. The so-called ‘penalty functional’ method was suggested by Haynes and Payne [95] to solve this problem. They suggest that the functional which must be minimised is

$$Q[\rho] = E[\rho] + \alpha P[\rho] \quad (3.27)$$

where α is a scalar and $P[\rho]$ is the penalty functional:

$$P[\rho] = tr[(\rho^2 - \rho)^2] \quad (3.28)$$

which has minima at $f_n = 0$ and $f_n = 1$. A problem with this method is that the local minima of the total functional $Q[\rho]$ do not occur when the occupancies are exactly

0 or 1, like is the case for the functional $P[\rho]$, so that minimisation of this functional will give the wrong energy and a not-exactly idempotent density matrix. Although correction expressions exist for the energy, this doesn't matter in ONETEP since the penalty functional method is used to rapidly get the density matrix roughly near idempotency and then the LNV method takes over. A recent assessment of different density-matrix optimisation schemes can be found in Ref. [102].

Expressions can be derived for the gradient of the functionals P and Ω in terms of the degrees of freedom of the density matrix $\{\phi_\alpha\}$ and $\{K^{\alpha\beta}\}$ so that methods such as the conjugate gradients method can be used to search for the minimum. The strategy used to minimise these functionals is the subject of the next subsection.

3.9.2 Minimisation strategy

In the last section we saw that ONETEP finds the ground-state energy by performing an unconstrained minimisation of the LNV grand potential:

$$E_0 = \min_{\{K^{\alpha\beta}\}, \{\phi_\alpha(\mathbf{r})\}} E[\{K^{\alpha\beta}\}, \{\phi_\alpha(\mathbf{r})\}] = \min_{\{K^{\alpha\beta}\}, \{\phi_\alpha(\mathbf{r})\}} \Omega_{LNV}[\{K^{\alpha\beta}\}, \{\phi_\alpha(\mathbf{r})\}] \quad (3.29)$$

In practice, this minimisation is performed in two constrained search stages which form a nested loop [78]. In the inner loop, the LNV functional is minimised with respect to the density kernel in the representation of the fixed NGWFs:

$$F[\{\phi_\alpha\}] = \min_{\{K^{\alpha\beta}\}} \Omega_{LNV}[\{K^{\alpha\beta}\}, \{\phi_\alpha(\mathbf{r})\}] \quad (3.30)$$

In the outer loop $F[\{\phi_\alpha\}]$ is minimised with respect to the NGWFs at a fixed density kernel:

$$E_0 = \min_{\{\phi_\alpha\}} F[\{\phi_\alpha\}] \quad (3.31)$$

The current density kernel, however, no longer satisfies the minimisation condition

with these new NGWFs, so the two stages need to be iterated to self-consistency.

3.10 Isolated supercells: the cutoff-Coulomb interaction

In this work we will be interested in carrying out calculations on isolated nanorods which are likely to have a large permanent dipole moment. ONETEP uses (at least at the present time) the supercell approximation, in which periodicity is imposed on an isolated system, with vacuum ‘padding’ surrounding the system so as to reduce the interaction between periodic images. This introduces a finite size error whereby the total energy varies with the size of the supercell. In systems with a large aspect ratio and a large dipole moment (e.g a polar nanorod), very large ‘padding’ regions are necessary in order to reduce this error to an acceptable level.

One method for eliminating this problem – referred to here as the *cutoff-Coulomb* method [103, 104] – has been implemented in ONETEP [105]. It involves truncating the Coulomb potential to zero after it reaches a certain distance away from the nanorod. We describe this method below.

The Hartree potential is given in real space by

$$V_H(\mathbf{r}) = \int d\mathbf{r}' n(\mathbf{r}') v_C(|\mathbf{r} - \mathbf{r}'|)$$

where $n(\mathbf{r})$ is the density and $v_C(|\mathbf{r}|) = 1/|\mathbf{r}|$. It is calculated more easily in reciprocal space where it is diagonal,

$$\tilde{V}_H(\mathbf{G}) = \frac{1}{\Omega} \tilde{n}(\mathbf{G}) \tilde{v}_C(\mathbf{G})$$

where $\tilde{v}_C(\mathbf{G}) = 4\pi/\mathbf{G}^2$, $\mathbf{G}$ are the reciprocal space coordinates and Ω is the volume of the simulation cell. It is then Fourier transformed back into real space.

Similarly, the Hartree energy is given by the expression

$$E_H = \frac{1}{2} \int d\mathbf{r} n(\mathbf{r}) V_H(\mathbf{r}) = \frac{1}{2} \mathcal{F}^{-1}(\tilde{n} \cdot \tilde{V}_H) = \frac{1}{2\Omega} \sum_{\mathbf{G}} |\tilde{n}(\mathbf{G})|^2 \tilde{v}_C(\mathbf{G})$$

We would like to modify the Coulomb potential term $v_c(\mathbf{r})$ so that it is forced to zero at some distance large enough so that any infinitesimal volume of charge from the charge distribution feels the proper $1/r$ potential from every other infinitesimal volume of charge in the charge distribution, but small enough so that no part of the charge distribution feels the Coulomb potential from any part of the charge distribution in the periodic image charge distribution. This can be done by replacing $v_c(\mathbf{r})$ with a new ‘spherical cutoff-Coulomb’ potential $v_{CC}(\mathbf{r})$ defined by

$$v_c(\mathbf{r}) \rightarrow v_{CC}(\mathbf{r}) = \frac{\Theta(R - r)}{|\mathbf{r}|}$$

where $\Theta(R - r)$ is the Heaviside step function which defines a sphere of radius R within which $\Theta(R - r) = 1$ (and therefore $v_{CC}(\mathbf{r}) = v_c(\mathbf{r})$) and outside which $\Theta(R - r) = 0$. The Fourier transform of $v_{CC}(\mathbf{r})$ can also be calculated analytically, and is

$$\tilde{v}_{CC}(\mathbf{G}) = \frac{4\pi(1 - \cos(\mathbf{G} \cdot \mathbf{R}))}{\mathbf{G}^2}$$

The region after which the Coulomb potential is cut to zero is a sphere in this case, but it need not be. It can be any shape, so long as one can calculate the numbers $\tilde{v}_{CC}(\mathbf{G})$, either analytically, or numerically using FFTs. Spherical cutoff regions are good for when the system is a zero dimensional structure (e.g. a nanoparticle), however for long structures like nanorods, it is desirable to use a cylindrical cutoff region. Derivations of $\tilde{v}_{CC}$ for different cutoff-region geometries can be found in Ref. [104], and an analysis of the merits of this, and other methods for dealing with the periodic-interaction problem in linear scaling DFT can be found in Ref. [105] (to be published in 2011).

3.11 Summary

The exponential decay of the density matrix, which is an inherent feature of insulating systems, means that it is a good approximation to assume that the density matrix $\rho(\mathbf{r}, \mathbf{r}')$ is short ranged in $|\mathbf{r} - \mathbf{r}'|$, and therefore that the information contained within it is of $O(N)$, where N is the number of electrons in the system. ONETEP, like other linear-scaling DFT methods, is designed in a way which allows the exploitation of this approximation. By adopting a local orbital representation of the density matrix, matrix elements very far from the diagonal are very small, and can be neglect. This produces very sparse matrices, that can then be manipulated in linear-scaling effort using efficient sparse matrix algebra. The strict localisation region on the orbitals in ONETEP, and the cutoff parameter on the density kernel are designed for this purpose. ONETEP uses a non-orthogonal basis, which in principle allows for greater localisation of orbitals since localisation and orthogonality are competing requirements, and therefore sparser matrices and a lower linear-scaling prefactor.

The innovation of the FFT-box technique allows matrix element of the hamiltonian to be calculated without having to carry out an FFT on the entire simulation cell, which scales as order $O(N \log N)$ for each matrix element, but rather only over a region encompassing the local orbital and its immediate neighbours – which scales as $O(1)$

Idempotency can be imposed implicitly in linear scaling effort via one of the self-consistent density matrix minimisation schemes that have been developed, (e.g. the LNV method), which ensure a relatively rapid and stable convergence of the density matrix to its idempotent ground state.

The cutoff-Coulomb method is used to eliminate the finite size error that arises when using the supercell approximation with charged or highly polar systems.

In the next chapter we will make use of these methods to investigate the polar properties and electronic structure of nanorods of realistic size.

4

Nanorod charge distribution and electronic structure

Nanocrystals in solution have been observed to exhibit a large dipole moment and there are a number of contradictory explanations as to its origin, which have been reviewed in chapter 1. As of today, attempts to investigate the charge distribution and the factors affecting it directly by experimental measurement have not shed enough light in this area to be able to satisfactorily resolve these contradictions in the literature. The main reason for this has been that current experimental technique doesn't allow a sufficient level of control over the nanocrystal's detailed structure, or a precise enough level of knowledge about an individual nanocrystal's state (electronic and ionic) to be able to disaggregate the multitude of factors which may be contributing to the dipole moment in a nanocrystal, and how they may be interlinked. Also, dielectric properties of nanocrystals have been little studied, yet are crucial for understanding the behaviour of nanorods in environments with large local electric fields (due to close proximity with other polar nanocrystals or charged ions in solution) or applied electric fields.

Computational modelling is an extremely useful tool for investigating some of these questions, since it allows us a level of knowledge and control over our system so that is not currently experimentally possible. Until recently, nanostructures of

realistic sizes have been beyond the reach of accurate quantum-mechanical methods. However, developments in linear-scaling density-functional theory (DFT) methods have now made possible the simulation of nanostructures comprising thousands of atoms with a high level of theory.

We take advantage of these methods, as implemented in the ONETEP code (chapter 3), to accurately simulate polar semiconductor nanorods of realistic sizes, with particular focus on how a nanorod’s charge distribution and electronic structure are affected by changes in the surface terminating species (section 4.1), surface geometry relaxation (section 4.2), nanorod size (section 4.3), semiconductor type (section 4.4), external electric fields (section 4.5), and interactions with other nanorods (section 4.6).

4.1 The effect of surface termination

In this section we model idealised nanorods, constructed with perfect bulk-like surfaces with no defects. The surfaces which make up the long sides of the nanorod (the ‘lateral’ surfaces) are exactly parallel with the long axis of the nanorod and with the wurtzite c -axis (otherwise known as the $[0001]$ direction, or the polar direction, or the z -direction in this work). The surfaces which terminate the nanorod along its long direction (the ‘end’ surfaces) are the polar wurtzite (0001) surfaces. We vary, the surface termination on the polar and the lateral surfaces and investigate the effect on the polarity of the nanorod.

For our investigation, we choose nanorods of wurtzite GaAs because accurate pseudopotentials for Ga and As are obtainable which require relatively low computational cost while having the essential features that endow all polar nanorods (e.g. CdSe, ZnO) with an asymmetric distribution of charge, namely, there is a degree of ionicity to the bonding and the crystal structure lacks inversion symmetry. This allows us to access the extensive size regime in which a nanorod’s dipole moment increases linearly with its length and cross-section, as we discover in section 4.3.

Initially we model unrelaxed, stoichiometric nanorods of wurtzite (w-)GaAs.

Our nanorods are ‘grown’ parallel to the wurtzite c -axis with a length ~ 12.8 nm and hexagonal cross-sections of area ~ 3.56 nm: an example is shown in Fig. 4.1.

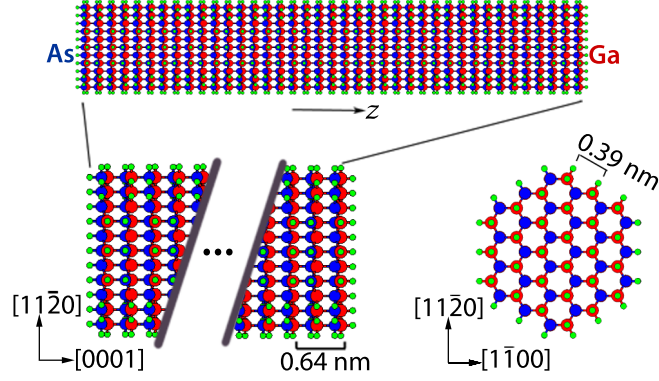


Figure 4.1 – An unrelaxed GaAs nanorod with Ga, As, and H atoms colored red, blue, and green, respectively. The Ga and As terminated ends are indicated.

Six different nanorods consisting of 2106–2862 atoms were created to represent a variety of lateral ($\parallel c$) and polar ($\perp c$) surface terminations, either bare or saturated by chemisorption. They are labelled H/H, H/B, H/P, B/H, B/B, B/P, to indicate how the lateral/polar surfaces are terminated, respectively. H signifies termination by hydrogen atoms and B signifies a bare surface. P denotes termination by pseudo-atoms [106], each with one electron and fractional nuclear charges of $+\frac{5}{4}$ and $+\frac{3}{4}$ for passivating surface Ga and As atoms respectively. By approximating the electronegativities of As and Ga they form bonds with the surface atoms of similar character to those found in the bulk and render the III-V semiconductor surfaces electronically inert [106]. Additionally, within the electron counting model discussed in section 1.3.2, these pseudo-atoms should not contribute any net charge to the surface, since the $\frac{5}{4}$ ($\frac{3}{4}$) electrons they contribute to each Ga-P $_{+5/4}$ (As-P $_{+3/4}$) bond, is canceled by their ionic charge. For both these reason, they make for a good reference terminating species. A H atom, on the other hand, passivates a surface dangling bond but also contributes a net negative (positive) charge of magnitude $\frac{1}{4}$ to the surface when it binds to a Ga (As) atom. In all of these models symmetry dictates that only the longitudinal component of $\mathbf{d}$, d_z , is non-zero.

For our calculations, a plane-wave energy cut-off of 450 eV, with a local or-

bital radius [107] of 0.53 nm for each atomic species is found to be sufficient to converge properties of interest. The density kernel is not truncated so that metallic and insulating structures are treated on an equal footing. We use norm-conserving pseudo-potentials with non-linear core corrections, and 3d electrons frozen into the core. Exchange and correlation are treated within the local density approximation. To eliminate interaction between a nanorod and its periodic images we have used a cylindrically truncated Coulomb interaction [104] instead of the usual Ewald interaction (see section 3.10). Preliminary calculations verifying our choices of simulation parameters and the suitability of our pseudo-potentials can be found in Appendix A

4.1.1 Results and discussion

Table 4.1 shows that both the magnitude and direction of $\mathbf{d}$ depend critically on the chemistry of both the lateral and polar surfaces and therefore cannot be dominated by the spontaneous polarisation of the wurtzite lattice. It is clear that $\mathbf{d}$ is not dominated by the spontaneous bulk-polarization of w-GaAs, which we have calculated using the Berry phase method implemented in the *Quantum Espresso* code, to be 0.005 C/m^2 , implying a contribution to d_z for nanorods of our size of $+62 \text{ D}$, if the polarizations are similar [42]. H/B and B/B both show positive d_z , meaning that the Ga-terminated end carries a net positive charge, while B/P has a very small d_z and the other rods exhibit negative d_z . These observations suggest that the synthesis conditions of nanorods and their chemical environments are likely to play crucial roles in determining $\mathbf{d}$, insofar as they affect the coverage of the surfaces with adsorbates. Furthermore, negative moments imply that surface chemistry effects have overwhelmed the contribution from the spontaneous polarisation of w-GaAs.

Figure 4.2 shows the calculated d_z of nanorods with fully H-terminated lateral surfaces, but with a varying coverage of H atoms on the polar surfaces. Each point represents a single sample chosen at random from the ensemble of nanorods with a given coverage. It is clear that d_z tends to decrease significantly as the hydrogen coverage increases and that $\mathbf{d}$ changes direction at around 56% coverage. This means

4. Nanorod charge distribution and electronic structure

Table 4.1 – Dipole moment d_z , net charge of the left-hand half Q_L and electric field at the mid-point E_m of the nanorods.

Nanorod	H/H	H/B	H/P	B/H	B/B	B/P
d_z (D)	-614	+330	-531	-235	+125	+41
Q_L (e)	+1.00	-0.56	+0.95	+0.39	-0.18	-0.08
E_m (V/nm)	+0.100	-0.050	+0.105	+0.030	-0.013	-0.005

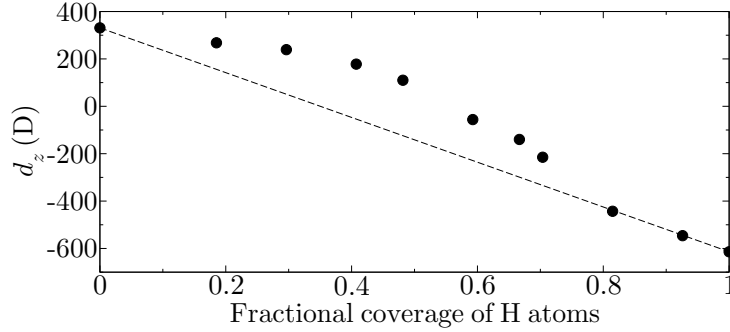


Figure 4.2 – Dipole moment d_z as a function of H atom coverage of the polar surfaces for a nanorod with fully H-terminated lateral surfaces.

that if one could control the degree of coverage of the polar surface (e.g. by varying the temperature, pressure, or the chemical potentials of the various species during synthesis) one could vary d_z over a wide range of values.

We have carried out a calculation which, though quite crude, lends some plausibility to this picture: at $T = 0$ K, we have calculated that over the range of realistic hydrogen chemical potentials μ_H for which water molecules are stable, the thermodynamically most stable polar surface hydrogen coverage of the data set in Figure 4.2, namely that which minimises $E - \mu_H n_H$, goes from 0 to around 70%. In principle, this therefore allows access to values of d_z of between -200 D and +330 D. Details of this calculation can be found in Appendix B.

In Figure 4.3, we plot the electron density profile along the length of the nanorod (the z direction in this work, $\parallel c$), averaged over the x and y directions. The large oscillations in the electron density over a unit cell make it very difficult to see how charge is distributed between unit cells. It is therefore useful to process the data in a way which averages out oscillations over the unit cell, but preserves

the cell by cell variations.

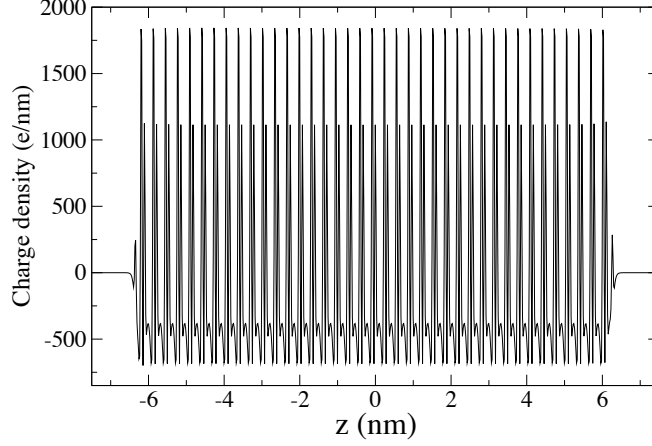


Figure 4.3 – Charge density profile along z , with x and y directions averaged out

In Figure 4.4(a) we plot the function

$$\tilde{\rho}(z) = \frac{1}{L\sqrt{2\pi}} \int \int \int \rho(x', y', z') e^{-\frac{(z-z')^2}{2L^2}} dx' dy' dz' ,$$

which is the laterally-averaged charge density profile along the length of the nanorod (the z direction in this work, $\parallel c$), convolved with a Gaussian of standard deviation ($L = 0.32$ nm) equal to half the length of a unit cell in the z direction. This smooths out the large variations in the density on the length scale of a unit cell, revealing how excess charge is distributed along the length of the nanorod: clearly it is spread over several nm from the ends of the nanorods. The amount of excess charge on the left-hand (As-terminated) end, Q_L , is shown in Table 4.1. This has been calculated by integrating the quantity $\tilde{\rho}(z)$ up to the middle of the nanorod.

The observed spread of the excess charge on the nanorod ends could be important for models of nanoparticle self-assembly that usually treat nanocrystals as point dipoles or assume that excess charge is perfectly localized on the polar surfaces (see section 1.3.3). These assumptions would lead to quantitatively, or even qualitatively, incorrect results at short distances. Furthermore, the delocalization of charge suggests that nanorods may be highly polarisable, which could significantly

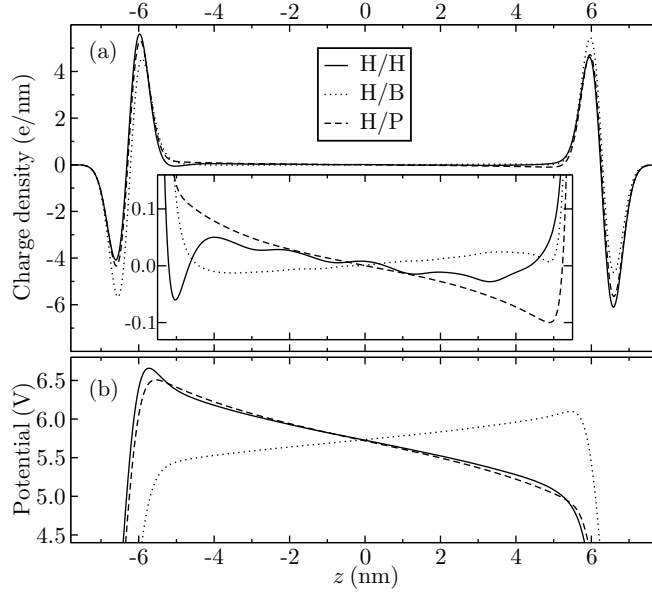


Figure 4.4 – (a) Charge density and (b) electrostatic potential as a function of position z along the long axes of the H/H, H/B, and H/P nanorods. Both quantities are integrated over the plane perpendicular to the long axis and smoothed on the length scale of a wurtzite unit cell.

affect the interactions between nanorods. We investigate their polarisability and interactions between nanorods in section 4.5 and 4.6.

Figure 4.4(b) shows the smoothed potential as a function of position along three different nanorods and Table 4.1 gives the values of the electric field at the centre of each nanorod, E_m . It is clear that the internal electric fields are not uniform but stronger near the polar surfaces than in the middle. The voltage drops we have observed are of order 0.1 V/nm, reaching as high as 0.6 V/nm near the polar surfaces of H/H. These fields are of similar magnitude to those observed in strained quantum wells and are expected to significantly affect optical absorption frequencies, selection rules and carrier recombination rates.

In Figure 4.5 we plot the density of states (DOS) for each of the six nanorods. The first initially surprising feature of these plots is that none of the nanorods appear to have a bandgap, and might therefore be interpreted as being metallic in character. However, there is a more subtle explanation of this feature, which requires us to consider the spatial distribution of these states, which we do by looking at the local

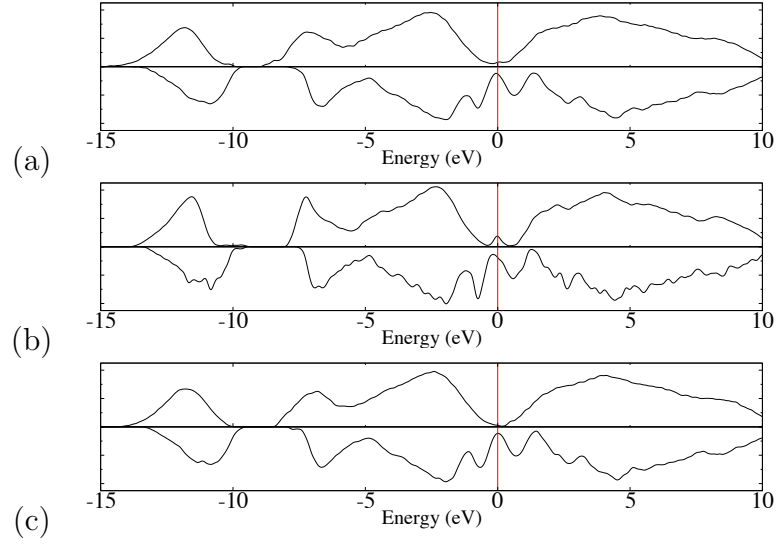


Figure 4.5 – Density of states for: (a) H/H (top) and B/H (bottom); (b) H/B and B/B; (c) H/P and B/P. The Fermi level is at 0eV. Nanorods with bare lateral surfaces have relatively many states near the Fermi-level.

density of states (LDOS).

Figure 4.6 shows the calculated ‘slab-wise’ LDOS for nanorods H/H, H/B, B/H, and a fourth nanorod which we label H/(67%H,33%B). This fourth rod has 67% hydrogen coverage on the polar surfaces, and is one of the rods from which the data in Figure 4.2 was generated. We define a slab LDOS as follows: each nanorod is nominally divided into 20 slabs in the z -direction, each consisting of four planes of atoms: two each of Ga and As. The slab LDOS is the sum of the contributions to the total DOS from the local orbitals centred on those atoms. Superposing these slab LDOS, as in Figure 4.6, shows that the electric field shifts the spectrum from slab to slab. This shift has the effect of smearing out the total DOS so that none of the nanorods studied has an electronic energy gap despite individual slabs having well-defined gaps.

Closer examination of Table 4.1 reveals two distinct trends: first, that all other things being equal, terminated polar surfaces result in the most negative d_z , and bare polar surfaces result in the most positive. Second, terminated lateral surfaces result in $\mathbf{d}$ of larger magnitude than bare lateral surfaces.

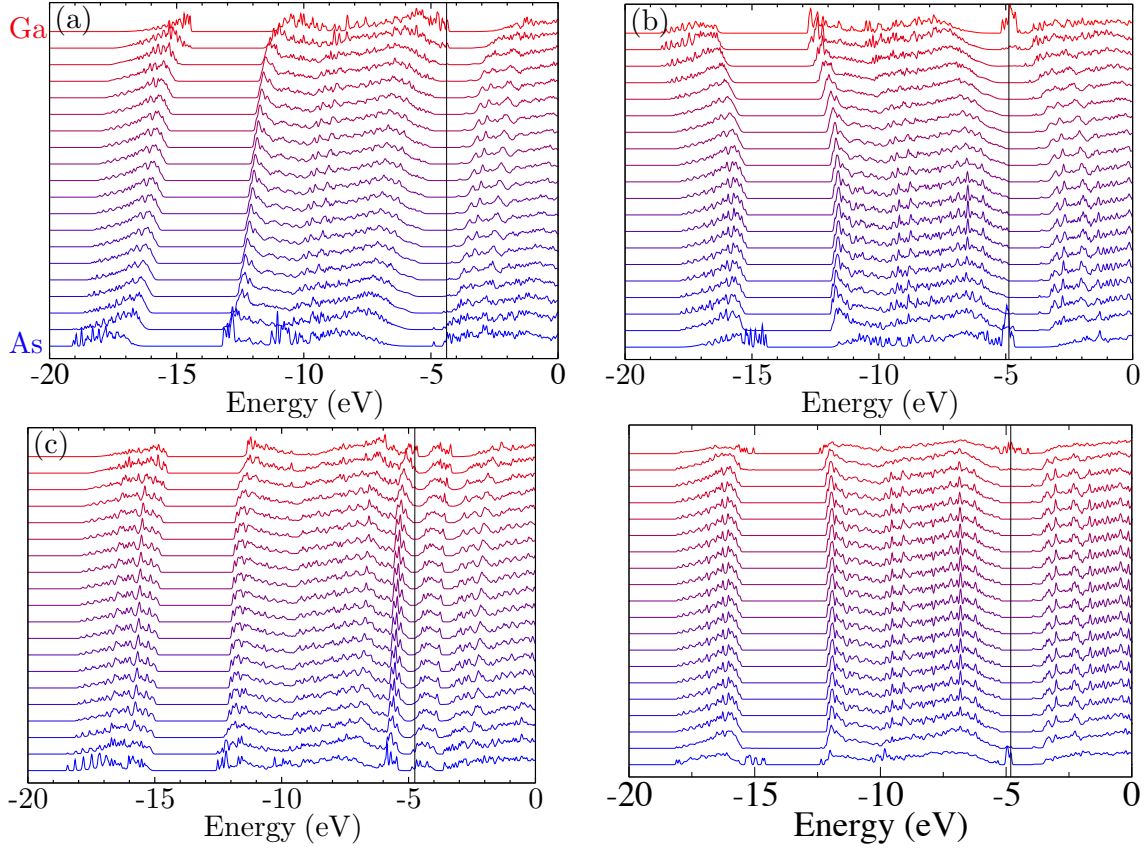


Figure 4.6 – Local densities of states for three nanorods: (a) H/H; (b) H/B; (c) B/H; (d) one of the nanorods of figure 4.2 with 67% hydrogen coverage on polar surfaces. The global Fermi level is plotted in each case. Red (blue) curves correspond to slabs close to the polar surfaces terminated by Ga (As).

Figure 4.6 helps us to understand these trends: polar surface dangling bond states on H/B can clearly be identified in Fig. 4.6(b) on the data sets associated with the first and last slabs of the nanorod. The Ga (As) dangling bond states lie mostly above (below) the Fermi level E_F , resulting in an excess of electrons on the As-terminated end and a more positive d_z . In contrast, Figure 4.6(a) reveals that both the Ga-H and As-H bonding states on the polar surfaces of H/H lie below E_F , resulting in an excess of electrons on the Ga-H terminated end and a depletion on the As-H end. This excess and depletion is because the $\frac{3}{4}$ of an electron nominally available from the Ga atom, and the one electron available from H are insufficient to fully populate the Ga-H bond. Conversely, the $\frac{5}{4}$ electrons from As provide a

surplus for the As-H bond. Hence the formation of these bonds redistributes charge between the two polar surfaces to produce a negative d_z .

Regarding the second trend: comparison of Figs. 4.6(a) and 4.6(c) shows that the nanorod with bare lateral surfaces (B/H) does not exhibit a large local energy gap as H/H does due to the presence of lateral surface dangling bond states in the gap. A large $\mathbf{d}$ is associated with a large internal electric field and slab-by-slab energy shift. However, for nanorods with lateral surface states close to E_F , the electric field pushes these states above E_F at one end and below at the other thereby reversing the charge build up on the ends and reducing the field. We conclude from these observations that a large local energy gap clear of surface states is a necessary condition for a large $\mathbf{d}$, and the quenching of surface states on unstable lateral surfaces, by the introduction of adsorbates, can be an important polarity enhancement mechanism in nanorods.

In all the nanorods studied, the Fermi-energy is found to coincide with the energies of states comprising a significant density, located close to the end surfaces of the nanorod. In the case of H/H (Figure 4.6(a)), these states are near the top of the valence band on the Ga-rich end, and the bottom of the conduction band on the As-rich end, and the potential energy difference for a test charge of e across the rod is therefore close to the band-gap energy. In the case of H/B, these states at the Fermi energy are mid-gap surface states, and the potential difference is smaller than the band-gap energy, and equal to the difference between the energies of the surface states on either end of the rod, relative to their local conduction band minimum. Accordingly, we find a lower dipole moment magnitude on H/B than H/H. In the case of the nanorod of figure 4.6(d), with a fractional polar surface coverage of 67% on each polar surface, these mid-gap states on opposite surfaces are much closer together in energy relative to their local band edges than those of H/B, resulting in a lower potential difference and a relatively small dipole moment (only -70 D).

A group of mid-gap states at the Fermi-energy is a common feature of many (particularly polar) semiconductor surface structures, and it plays a role in the phenomenon of Fermi-level pinning: the group of mid-gap states stabilises (“pins”)

the Fermi-energy at the surface at the position of their average energy. It is stabilised because the surface states are able to act as donors or acceptors, which get filled or emptied to compensate for any change that may affect the relative position of the Fermi-level (e.g. the application of a voltage across the nanorod).

We suggest a related picture for our nanorod system, which we will find further support for in subsequent sections; in any given nanorod structure the, direction of the dipole moment is dependent on the balance between the intrinsic polarisation of the crystal, the additional ionic charge on each end, and the number of occupied electronic surface states on each end, which may change when adatoms are added or removed, or for different reconstructions of the surfaces. However, there is a second determining factor, which is the tendency for the Fermi-level to become stabilised at energies associated with a group of mid-gap states, (or if there are no mid-gap states, by the band edges). There can be three reasons for the stability of the Fermi-level: (1) high density of states at the Fermi-level, so that small changes in the potential difference across the rod can generate large redistributions of charge to mitigate it; (2) a large energy barrier to occupying and un-occupying states on the ends of the rod because they redistribute in large discrete lumps of 1 (or 2 in non-spin polarised systems) electrons, and not (effectively) continuously as is the case in the limit of semi-infinite surfaces; and (3) highly polarisable occupied orbitals can screen out perturbations to the potential across the rod.

One of the implications of this picture is that the problem of nanorod polarity cannot be cast in terms of an ionic model, or a simple bond-electron counting model, since these models do not include constraints on the nano-rod potential difference imposed by the Fermi-level-stabilisation tendency proposed above. For example, the observation from Table 4.1 of the similarity in d_z , Q_L and E_m between rods H/H and H/P, despite the very different ionic charge of the terminating atoms on the polar surfaces seems surprising at first. A simple bond electron and ion counting argument predicts that the Ga (As) polar surface on H/P should have an additional charge of +0.25e (-0.25e) for each of the 27 bound pseudo-atoms over those of nanorod H/H, and should therefore have a greatly reduced dipole moment and potential difference

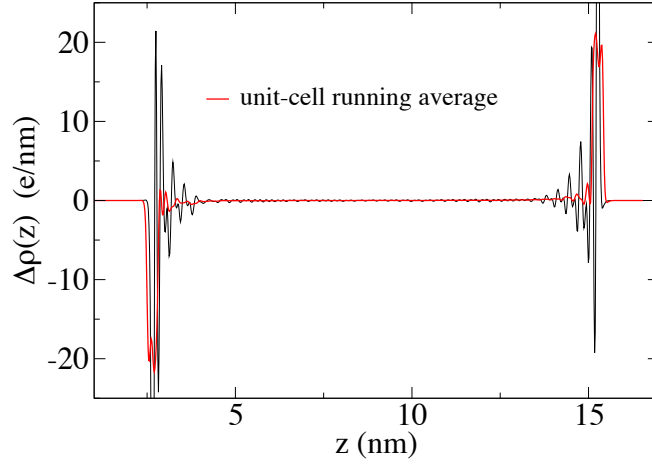


Figure 4.7 – Difference in electron density profile between H/H and H/P

across it. However, the LDOS of H/H (Figure 4.6(a)) shows that the Fermi-level is located within regions of high surface state density on either end of the rod, and is likely to be stabilised. Changing the ionic charge of the hydrogen atoms on the polar surfaces (to go from rod H/H, to H/P), can mean one of three things: either the potential difference across the nanorod must change, to reflect the extra ionic charge on the surfaces, which would result in occupancy shifting of the polar surface states, or the electrons can polarise so as to cancel out the extra field generated by the ionic charge, or some mixture of the two. We find that the potential difference across the nanorod does not change with the change in ionic charge – the Fermi-level is stabilised at nearly the same level relative to their local valence band maxima on each end. The electrons have been polarised by the extra ionic charge so as to screen out the extra field it produces. Figure 4.7 shows the difference in the electron distribution along the length of the rods between H/H and H/P. The polarisation of H/P relative to H/H generates an excess (shortage) of exactly 6.75 electrons on the Ga (As) end of the rod, exactly cancelling the excess (shortage) of ionic charge on that surface between rods H/H and H/P, thereby preserving the potential difference across the nanorods.

This interpretation implies that the polarisability of electrons is very high in nanorods of this size, given their ability to screen out most of the field from the

extra charge of the pseudo-ions. In section 4.5 we investigate the polarisability of nanorods in a uniform electric field, and we do indeed find a strong enhancement of the polarisability compared with bulk, which may well originate from the relative stability of the Fermi-level position at the nanorod ends.

4.2 The effect of structural relaxation

So far we have considered pristine nanorods. However, structural relaxations might be expected to play a large role for two reasons: some of the surfaces studied above may be unstable and, given our arguments above, reconstructions which change the surface composition are likely to have a large effect; and strain caused by the surface may also induce significant charge separation. Regarding the first concern, it has not been our intention here to determine the stability of a particular surface, but rather to investigate the link between surface termination and **d**, regardless of stability. Secondly, to assess the effects of surface-induced strain we performed structural relaxations on four much smaller nanorods of length ~ 3.5 nm, width ~ 1.2 nm and with different surface terminations (b/b, h/b, h/h, b/h, following the same convention as before). Additionally, we perform one very computationally expensive structural relaxation on the larger H/H rod of the previous section (length 12.8 nm, cross-sectional area 3.56 nm^2). We label the resulting geometry-relaxed structure H/H-r. In both cases, we converge the calculation so that the forces on all atoms are below $0.002E_h/a_0$

Before we consider the changes in electronic and ionic structure of the smaller rods, we first look at the electronic structure of H/H-r. In Figure 4.8 we plot the slab-by-slab LDOS of H/H-r in the same way as was done for H/H in Figure 4.6. Aside from the well known “healing” of the local band-gap on the end surfaces, we find the presence of a pair of nearly degenerate mid-gap surface states at precisely the Fermi-energy, 0.3 eV above the top of the (local) valence band, on the Ga-H terminated end of the nanorod. Both the HOMO and LUMO states, plotted in Figures 4.9 and 4.10 respectively, belong to this set of surface states on the Ga

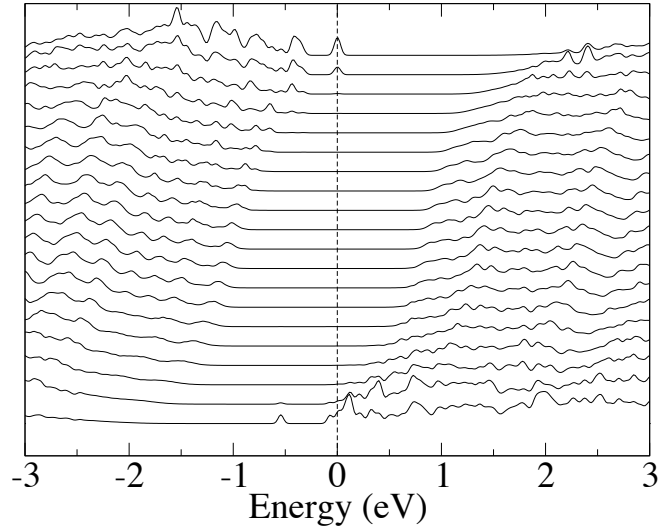


Figure 4.8 – relaxed rod: Fermi-level stabilised at mid-gap states

surface, and look very similar in shape, save for a symmetry relation.

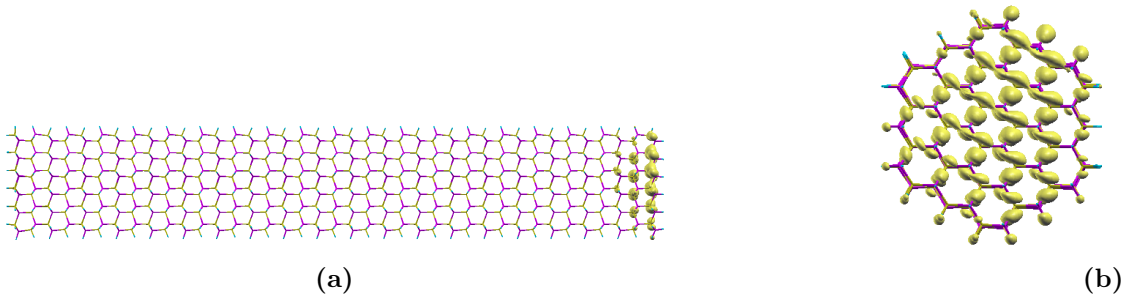


Figure 4.9 – Iso-surface of the HOMO viewed (a) along the $[11\bar{2}0]$ direction, and (b) looking at the (0001) polar surface. (iso-value of $0.001 \text{ e}/a_0^3$).

The presence of these mid-gap surface states is significant because it means that the Fermi-energy on the Ga-rich polar surface is stabilised at an energy closer to the conduction band minimum (the location of the Fermi-energy on the As polar surface), than it would be if this mid-gap state didn't exist (perhaps because of different reconstruction on the surface). The result is a lower potential difference across the nanorod than would be expected with no mid-gap states, and therefore a lower dipole moment. This illustrates a potentially important mechanism by which different polar surface reconstructions may produce different nanorod dipole

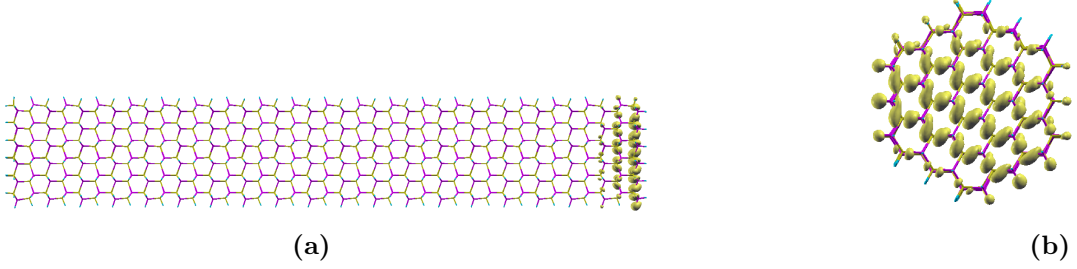


Figure 4.10 – Iso-surface of the LUMO viewed (a) along the $[11\bar{2}0]$ direction, and (b) looking at the (0001) polar surface. (isovalue of $0.001 \text{ e}/a_0^3$).

Table 4.2 – Configurations and dipole moments d_z before and after structural relaxation of the four nanorods.

Nanorod			h/h	h/b	b/h	b/b
$d_z \left\{ \begin{array}{l} \text{before} \\ \text{after} \end{array} \right\}$	relaxation		-131.64	+55.20	-22.81	+15.38
		(D)	-103.03	+67.99	-89.23	+21.52

moments.

Table 4.2 shows d_z before and after structural relaxation for the four smaller rods. Only rod b/h displays a large change in dipole moment on relaxation. No dipole moments changed direction, and the ordering of the magnitudes stayed the same. The changes in the DOS as a result of relaxation are shown in Figure 4.11. The increase in d_z on b/h coincides with the shifting of the lateral surface dangling bond states away from E_F on relaxation, as can be seen in the DOS in Figure 4.11(d). This healing of the energy gap allows for a larger $\mathbf{d}$ to be supported, consistent with the argument about the effect of lateral surfaces in section 4.1.

The relaxed structures from the geometry optimisation of rods b/b, h/b, h/h and b/h are shown in Figure 4.12. They reveal quite drastic relaxations on some surfaces (in particular, on bare polar surfaces) and only very small changes on others (namely, all hydrogen terminated surfaces). The severity of the relaxation on the polar surfaces of h/b and b/b indicates that those surfaces would likely not be stable at those compositions if brought into contact with a realistic environment. For example, the polar Ga surfaces of h/b appear to be ejecting Ga atoms from

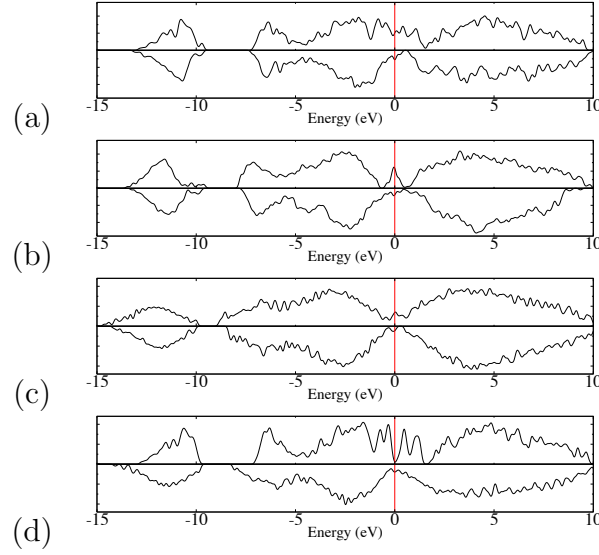


Figure 4.11 – Density of states plot for: (a) b/b , before (top) and after (bottom) relaxation; (b) h/b ; (c) h/h ; (d) b/h . In all cases relaxation has tended to move surface states away from the Fermi-level.

the surfaces. Hydrogen termination stabilises the polar surfaces against severe reconstruction (at least at a temperature of 0K in vacuum). We describe the relaxed geometries in more detail in Appendix C.

Semi-infinite bulk-like bare polar surfaces are known to be unstable to reconstruction, a feature known as the ‘polar instability’, described in chapter 1.3.2. Semi-infinite surfaces are only stable when neutrally charged, and they often reconstruct so as to adopt a closed-shell configuration which is simultaneously neutral. This is evidently not the case in quasi-0D nanocrystals, whose surfaces carry charge - there is no ‘neutrality’ condition to aid us in finding low energy structures. The problem of finding stable nanostructures, given a specific environment, is an extremely difficult one given the extent of variability in the space of nanocrystal morphologies, and it is beyond the scope of this thesis.

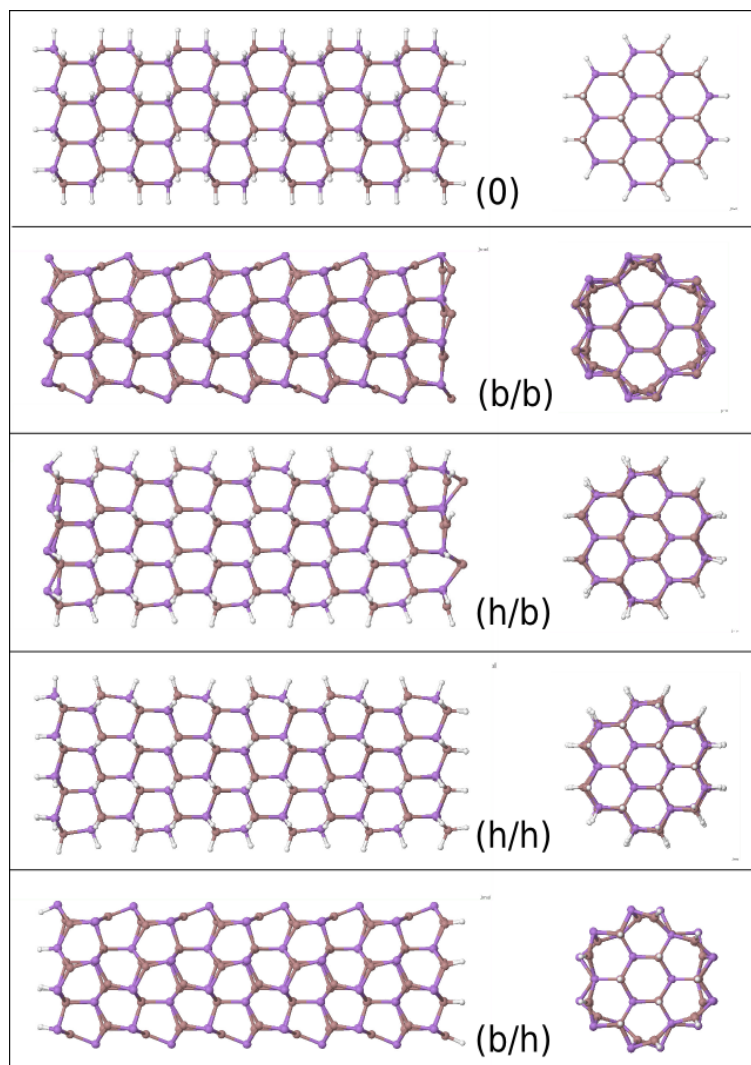


Figure 4.12 – Structure of an unrelaxed, H terminated nanorod (0) and the relaxed structures of b/b to h/h. Red spheres represent Ga atoms, purple represents As, and grey represents H. The left hand picture looks down the $[11\bar{2}0]$ direction and the right hand picture looks at a view of the (0001) polar surface.

4.3 The effect of nanorod size

In this section, we look at how the dipole moment of nanorod H/H varies with rod length and cross-sectional area. Figure 4.13 shows the results obtained by keeping the cross-sectional area at a ‘default’ value of 3.56nm^2 , and varying the nanorod’s length between a minimum value of 6.4nm to a maximum of 51.2nm , and in Figure 4.14a we fix the nanorod length at its ‘default’ value of 12.8nm , and vary the cross-sectional area between a minimum of 0.4nm^2 and a maximum of 9.88nm^2 (corresponding to a cross-sectional hexagon side of 5 unit cells in length).

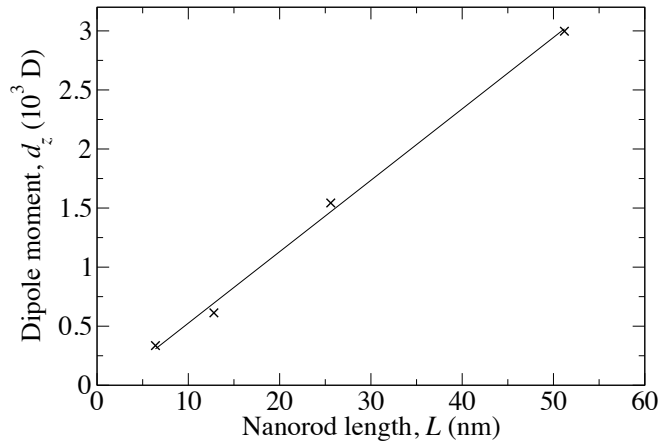


Figure 4.13 – The magnitude of the dipole moment increases linearly with nanorod length for nanorods of cross-sectional area 3.56nm^2 .

We find that the dipole moment increases proportionally with the length of the nanorod, over the range studied. This implies that the excess polar surface charge density on each end surface is independent of nanorod length, at least for rods of this cross-sectional area.

As cross-sectional area is increased (Figure 4.14a), the dipole moment increases too, however the data suggests that the relationship isn’t linear. In fact the average surface charge *density* on the polar end surfaces, plotted in Figure 4.14b, decreases rapidly with cross-sectional area, towards a constant value that is probably slightly above zero for nanorods of this length (because polar thin-films, unlike semi-infinite surfaces, are known to exist with uncompensated polarity). After a

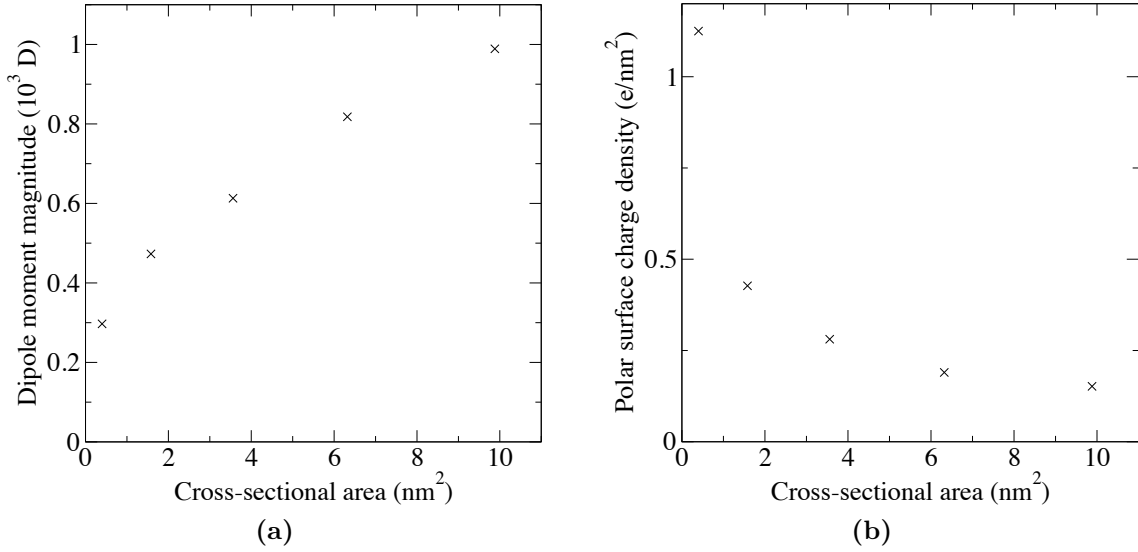


Figure 4.14 – The magnitude of the dipole moment increases with nanorod cross sectional areas (shown in (a)) for nanorods of length 12.8nm. The relationship is not linear over the whole of this range, as the polar surface charge density falls off quickly, asymptotically approaching a finite value (shown in (b))

cross-sectional area of about 4 nm^2 , the variation in the surface charge density is so small that the nanorod dipole moment may be said to increase linearly with cross-sectional area.

In Figure 4.15a we plot the LDOS associated with four consecutive planes of atoms (referred to as a ‘slab’ section 4.1) located far from the ends of nanorod so that no polar surface states are present in the LDOS. We find that the local band-gap is higher in thinner nanorods. This variability of the size of the band-gap with nanocrystal size is known to be the result of quantum confinement in small nanocrystals. We observe no such variability with nanorod length, since the extent of the nanorod along the long axis is too large for quantum confinement to be significant.

The LDOS plots of the two end polar surface slabs in Figure 4.15b illustrate that as the rod gets thicker, a greater energy window of surface states crosses the Fermi-energy on the Ga (As) rich surface, and become unoccupied (occupied), con-

4. Nanorod charge distribution and electronic structure

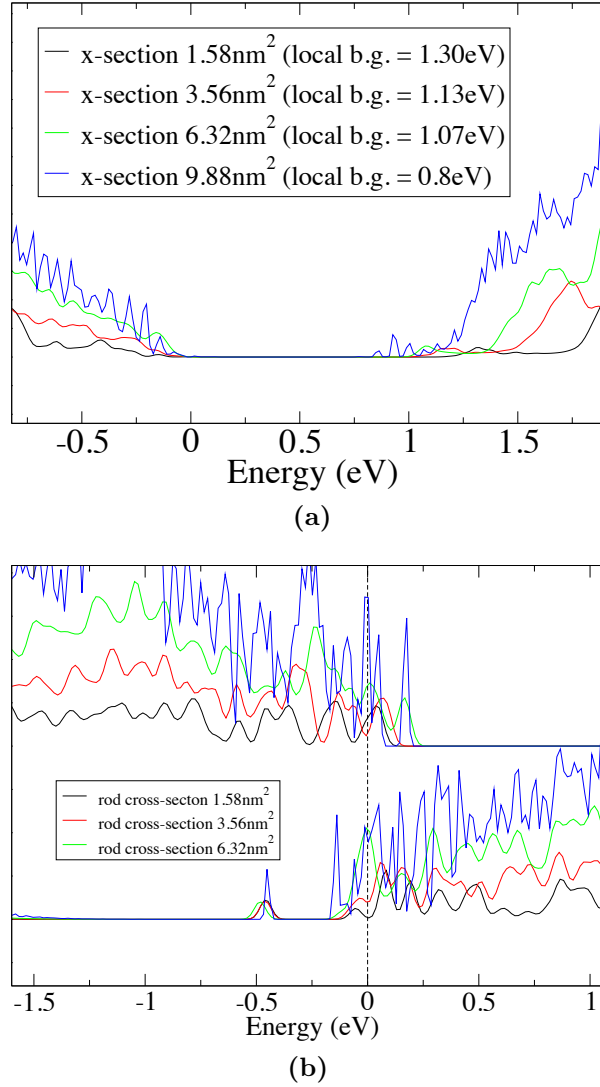


Figure 4.15 – LDOS plot of (a) a slab in the middle of the rods; (b) the Ga-rich end (top) and the As-rich (bottom)

tributing to a reduction in the surface charge density. This charge redistribution serves as a buffer against large change in the potential difference across the nanorod, though a significant change is permitted. This shift is bound to increase the instability of the surface as more and more density of states accumulates around the Fermi-level.

We will argue that there are two factors which play a role in the reduction of the polar surface charge density as the rods get thicker – the reduced decay rate (due to less ‘fringe’ loss) of the nanorod’s internal polarisation field, and the reduced

band-gap – the first being relatively more important.

We have seen in section 4.1 the important role that the band-gap can play in determining the potential difference across the nanorod, by stabilising the Fermi-energy at the band-edges (at least for rods with no mid-gap polar surface states). In the case of rod H/H, which we are using in this section, the Fermi-energy is stabilised close to the band-edges at either end of the rod by surface-states, and as the thickness of the nanorod increases, the potential difference between the band-edges decreases due to a reduction in the confinement energy. The contraction of the band-gap is indicative of a more general contraction of all states into a narrower energy window. Charge redistribution between the two surfaces may be necessary then, as the band-gap decreases, either to stabilise the system at a lower potential difference, or if the potential difference doesn't change, to respond to the shift of the polar surface states relative to the Fermi level, some of which will have crossed and changed their occupancy.

In section 4.1 we argued that the position of the Fermi energy with respect to the polar surface states is very stable for three reasons: (1) high density of states at the Fermi-level, so that small changes in the potential difference across the rod can generate large redistributions of charge to mitigate it; (2) a large energy barrier to occupying and un-occupying states on the ends of the rod because they redistribute in large discrete lumps of 1 electron; and (3) highly polarisable occupied orbitals can screen out perturbations to the potential across the rod. In thicker nanorods, the energy cost (per unit volume, per unit charge) of polarising the planes of charge in the crystal is larger, because the resulting polarisation field decays less rapidly due to reduced finite size effects (or 'fringe' effects). This means that the energy barrier against shifting the relative position of the Fermi-energy becoming relatively lower – hence why we find in Figure 4.15b that there has been relatively more occupancy shifting in thicker rods, which have less polarisable bonds.

We suggest that because the polarisation field decays more rapidly in thinner rods, more charge density must be accumulated on the polar surfaces in order to produce the required potential difference across the nanorod which is fixed by the

stabilised positions of the Fermi energy on both the end surfaces, and this is the main reason for the drop in surface charge density with thickness observed in Figure 4.14b.

In order to put these trends in perspective, it is useful to consider the asymptotic scaling of the energy with nanocrystal size. The asymptotic contribution from charged surfaces as the finite crystal gets longer depends on the aspect ratio of the crystal, and ranges somewhere in between that of separating point charges, which scales as $1/L$, and that of a separating infinitely extending charged parallel-planes, which scales as L . Simultaneously long and thick nanorods would carry a very high surface energy per unit area if they maintained charged surfaces, and therefore charged surfaces are unlikely to exist. For nanorods that are thick enough so that the internal electric field is relatively constant over a specified range of nanorod lengths, the increase in the potential difference across the rod would generate a redistribution of charge as the Fermi level runs into a band, so that the nanorod's dipole moment would not scale linearly with length over the specified range of lengths. We do not see this effect in our nanorods because the decay in the internal field is fast enough to ensure that the potential difference remains nearly constant across all the nanorod lengths: the field in the rods of length 12.8 nm, 25.6 nm and 51.2 nm are found to decay to values of 0.1 V/nm, 0.035 V/nm and 0.009 V/nm respectively in the rod centres.

4.4 The effect of semiconductor type (GaAs, GaN, AlN)

In this section we investigate how the polar behaviour of nanorods depends on which semiconductor the nanorod is made of. We calculate the charge distribution of three rods – one made of GaAs, another of GaN, and a third of AlN. The rods are of type H/P (lateral surfaces fully covered in with hydrogen atoms, and polar surfaces fully covered with pseudo-hydrogen atoms), all have the same number of atoms (2862, as in previous sections) , and are constructed of the same number of unit cells in

4. Nanorod charge distribution and electronic structure

each direction. Atoms are located at their bulk equilibrium values, as calculated in the CASTEP plane-wave-DFT code, meaning that the GaAs rod is longer than the GaN rod, which in turn is longer than the AlN rod, because of the differences in bulk lattice parameters.

Figure 4.16(a) and (b) show the Gaussian-smoothed, laterally averaged distribution of charge along all three rods, calculated in the same way as in section 4.1.

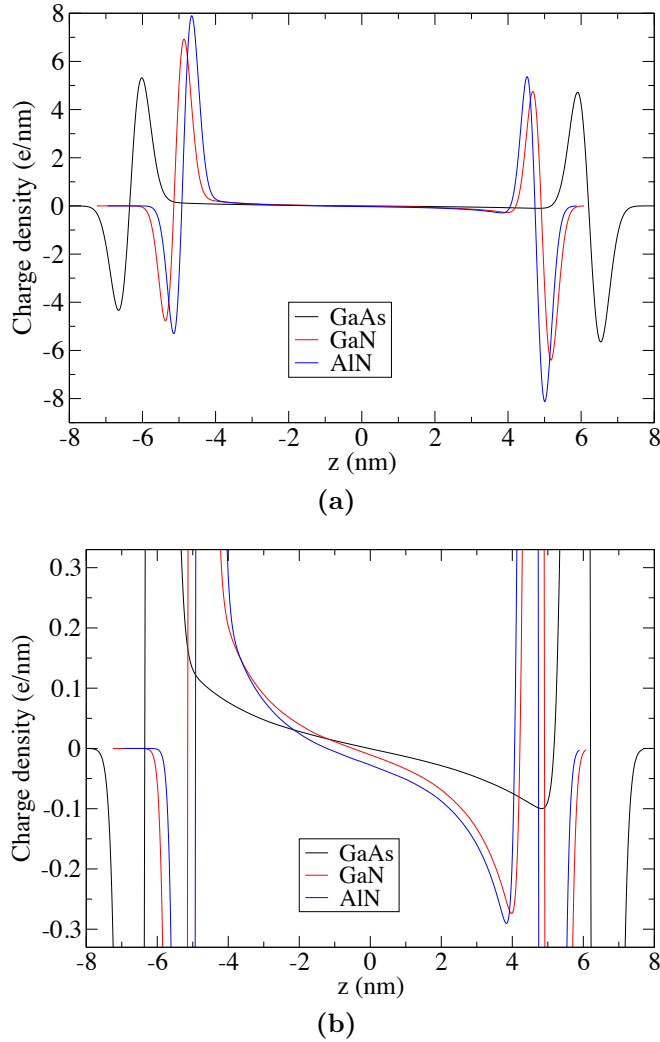


Figure 4.16 – Laterally averaged and Gaussian-smoothed charge distribution along the length of the nanorod for rods of AlN, GaN and GaAs. The data is the same in (a) and (b), but (b) has a magnified “charge density” axis.

The main characteristics of these rods and their charge distributions are sum-

4. Nanorod charge distribution and electronic structure

	AlN	GaN	GaAs
Rod volume (nm ²)	21.87	23.32	45.66
Rod d_z (D)	-713	-682	-531
Rod Q_L (e)	1.61	1.50	0.95
Rod Q_L decay constant	0.48	0.80	1.02
Rod polarisation (C/m ²)	-0.11	-0.098	-0.039
Bulk sp. polarisation (C/m ²)	0.073	0.029	0.005
Bulk permittivity, ϵ_r	9	9.7	13.1
Experimental Bulk bandgap (eV)	6.2	3.4	1.4
DFT bulk lattice parameter a (Å)	3.075	3.154	3.935
DFT bulk lattice parameter c (Å)	4.941	5.132	6.486

Table 4.3 – Some properties of AlN, GaN, and GaAs in nanorod and in bulk.

marised in Table 4.3, along with reference information about the bulk properties of these semiconductors (their lattice parameters, band-gaps, relative permittivities and spontaneous bulk polarisations).

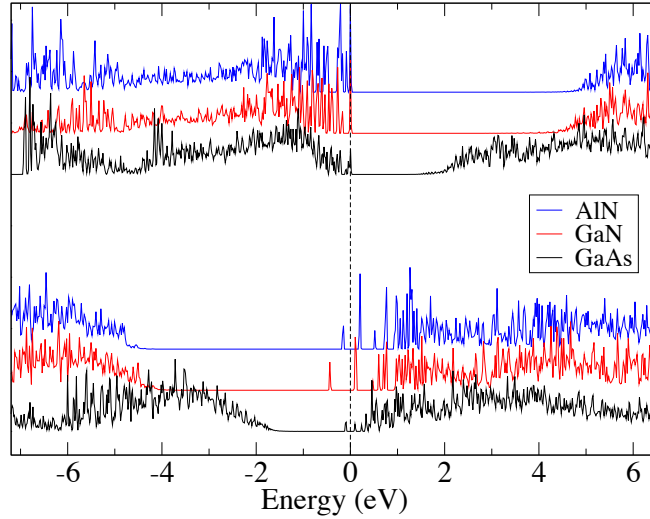


Figure 4.17 – LDOS plots of the cation-rich (top) and anion rich (bottom) polar surfaces of AlN, GaN and GaAs.

We find that both the dipole moment and the polarisation (dipole moment density) is positively (negatively) correlated with semiconductor band-gap (permittivity), with the nanorod of the largest band-gap semiconductor (AlN) supporting the largest polarisation, and the nanorod of the lowest (GaAs) supporting the lowest.

There is a similar correlation with the localisation of the excess charge on the ends, with excess charge being more localised (i.e. exhibiting a larger decay constant) in nanorods of higher band-gap materials.

The LDOS plots in Figure 4.17 of the polar surfaces of the GaAs, GaN and AlN rods reveal why the polarity has increased with band-gap for these particular nanorod structures: the Fermi-level is stabilised by the high density of mid-gap surface states which lie just above the valence band edge on the Ga/Al end of the rod, and below the conduction band edge on the As/N end. The difference in energy between these states on the two ends of the rod, relative to their respective local valence band maximum, is much larger in the AlN rod compared with the GaAs rod, so the potential difference across the rod is much larger.

4.5 The effect of applied electric fields

4.5.1 Introduction

When an electric field is applied to atoms, molecules or crystals – perhaps as a result of close proximity to a charged ion or surface, or due to exposure to electromagnetic waves – their charge distributions become distorted, and their dipole moments change. For small fields, the displacement of charge from the equilibrium position is proportional to the strength of the applied field.

$$d_i(\omega) = \sum_j \alpha_{ij}(\omega) E_j(\omega)$$

where the indices label the three cartesian basis vectors. The tensor $\alpha_{ij}(\omega)$ is known as the linear-polarisability. It is a tensor, not a scalar, because it is possible in general for a field in, say, the x -direction, to induce contributions to the dipole moment in the y or z directions. The polarisability of a material due to an oscillating electric field of frequency ω depends on whether the charged particles which make up the molecule or crystal can accelerate fast enough to keep up with the oscillations in the field. Only electrons are efficiently polarised by optical frequencies, whereas both

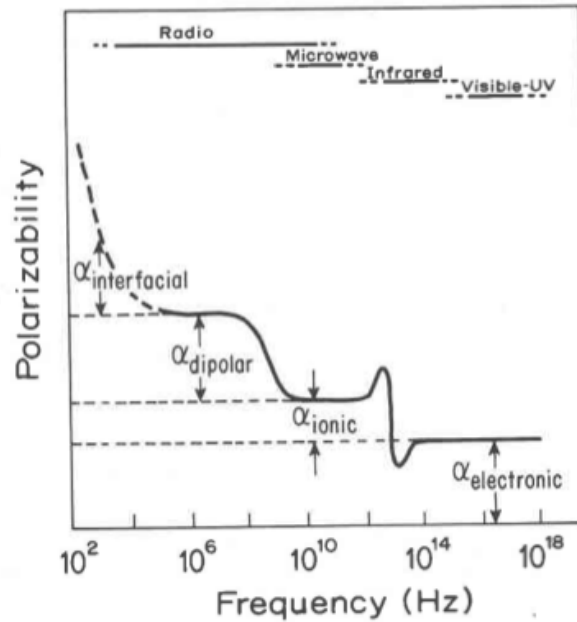


Figure 4.18 – Dependence of the polarisability of a material on the frequency of oscillation of the field.

electrons and ions are efficiently polarised in static fields.

In a colloidal solution of nanocrystals, there are four basic polarisation mechanisms: (1) the distortion of the electronic charge distribution within a nanocrystal relative to the ionic positions (or pseudo-potential positions, in our case) defines the electronic polarisability; (2) the relative displacement of ions from each other within a nanocrystal defines the ionic polarisability; (3) the change in orientation of permanent dipoles which can freely rotate, defines the dipolar polarisability; and (4) the effect of the rearrangement of surface charges defines the interfacial polarisability. Each of these mechanisms has its own characteristic time-scale, and is not activated by electric fields which oscillate on a shorter time-scale, as illustrated in figure 4.18.

In this section we are interested in the ‘static’ ($\omega = 0$) linear-polarisability of a single nanorod, with a focus on how it differs from bulk crystal, and its dependence on size and surface termination.

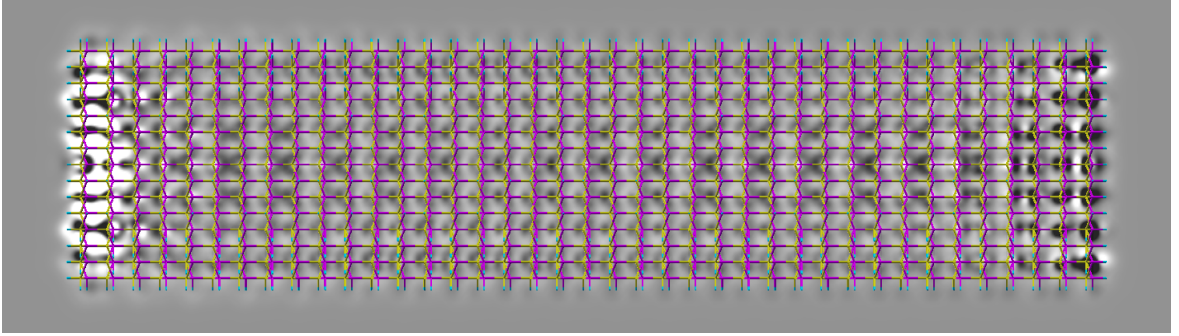


Figure 4.19 – Change in the electron density when a field of $E = 0.05$ V/nm is applied. Grey is no change, white is negative change, black is positive change. Dipole moment changed by 154 D (from -658 D to -504 D).

4.5.2 Electronic and ionic polarisability

Electrons with fixed ions

We model nanorod H/H-r (fully hydrogen terminated, geometry relaxed in zero field) using the same simulation parameters as in previous sections and we apply a uniform electric field of varied strengths to the nanorod by adding a saw-toothed potential energy term to the Kohn-Sham equations.

First, we apply a large field of $E = +0.05$ V/nm, which is nearly half the magnitude of the isolated nanorod's intrinsic electric field (calculated as $E = -0.11$ V/nm), in the direction opposite to its permanent dipole moment. The ion cores are initially fixed at their $E = 0$ equilibrium positions, and we consider only the response of the electrons to the field. We will consider the response of the ions in the next part of this section. We find that the magnitude of the nanorod's dipole moment decreases by 154 D, from a value of -658 D at $E = 0$, to -504 D at $E = 0.05$ V/nm. Figure 4.19 shows the difference in the charge density generated by the electric field, along a $(11\bar{2}0)$ cross sectional plane which contains the long $[0001]$ axis of the nanorod. In Figure 4.20, we average out variations in the x and y directions and show change in the charge distribution along the z axis ($[0001]$ direction).

We find a significant polarisation of the electron density due to the field,

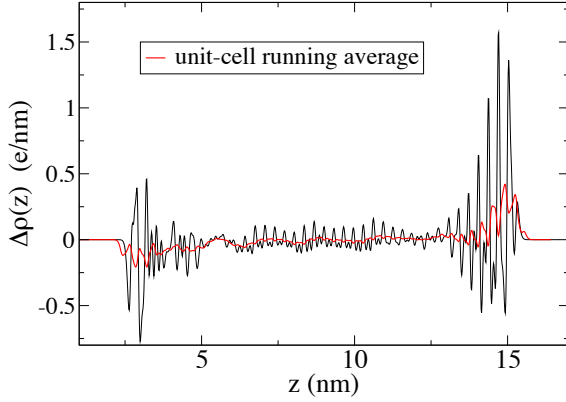


Figure 4.20 – Change in cross-sectional-averaged electron density resulting from the field

even in the middle of the rod. The excess charge on the polar surfaces resulting from the polarisation is $\pm 0.3e$, which can be seen to be spread over $\sim 2\text{nm}$ (~ 6 atomic layers) on either end of the rod.

The electric field in the middle of the nanorod is changed very little by the application of the external electric field, changing from $E_{\text{intrinsic}} = -0.10\text{V/nm}$, to $E_{\text{intrinsic}} = -0.11\text{V/nm}$. In effect, the charge has been rearranged in a way which has screened out the applied field.

Ionic polarisation

In bulk materials, the measure of how easily a material polarises in response to an electric field is the susceptibility $\chi(\omega)$. We calculate the ‘optical’ ($\omega = \infty$) and ‘static’ ($\omega = 0$) susceptibilities of bulk wurtzite GaAs, using CASTEP: a plane-wave DFT code and find them to be:

$$\chi(\infty) = \begin{pmatrix} 11.16 & 0 & 0 \\ 0 & 11.16 & 0 \\ 0 & 0 & 11.51 \end{pmatrix}$$

$$\chi(0) = \begin{pmatrix} 12.73 & 0 & 0 \\ 0 & 12.73 & 0 \\ 0 & 0 & 13.38 \end{pmatrix}$$

in the basis of the primitive unit cell of wurtzite. $\chi(\infty)$ is associated with the po-

larisability of the electrons only (electrons are assumed to respond instantaneously in DFT), and $\chi(0)$ includes the the polarisability of both electrons and ions. Accordingly, $\chi(\infty)$ can also be interpreted as the ‘static’ polarisability of bulk wurtzite GaAs *if* the positions of the ions are kept fixed. Therefore, focusing specifically on the susceptibility along the wurtzite c-axis, we see that it is dominated by the electronic susceptibility, which alone contributes $11.51/13.38 = 86\%$ to the total susceptibility.

Turning now to our nanorod model, we perform a geometry optimisation of the H/H-r nanorod in the same uniform electric field as in subsection 4.5.2 (i.e. $E = +0.05$ V/nm). The dipole moment of this relaxed structure is calculated as -496 D, a change of 162 D from the zero-field value of -658 D, calculated in section 4.5.2. Comparing this with the change of 154 D observed in the fixed-ion model of the previous section, it is clear that the change in dipole moment resulting from a relaxation of the electrons only, contributes the greatest portion to the polarisability of the nanorod (at least in small fields), just like in bulk. Specifically, the electrons alone contribute $154/162 = 95\%$ to the total polarisability – a larger proportion than the 86% seen in bulk. We will see in section 4.5.3 that the electrons in the nanorod are a lot more polarisable than they are in bulk, and this could explain this higher percentage.

4.5.3 Polarisability: a comparison between nanorods and bulk

In this section, we calculate the electronic polarisability of the H/H-r nanorod along three orthogonal directions, $[11\bar{2}0]$ (x -direction), $[1\bar{1}00]$ (y -direction) and $[0001]$ (z -direction), and compare them to their bulk values. For a schematic representation of these directions, see Figure 4.1. We ignore the ionic polarisability due to: (1) its relative insignificance to the total polarisability in small fields compared with the electronic polarisability; (2) the large computational resources that would be required to calculate it. In order to calculate the nanorod’s electronic polarisability, we calculated the dipole moment of a rod at 5 different values of electric field E (-

4. Nanorod charge distribution and electronic structure

Polarisabilities ($10^3 \times e^2 a_0^2 / E_h$)		
	Bulk	Rod H/H-r
α_{xx}	271	51.2 ± 0.15
α_{yy}	271	51.0 ± 0.13
α_{zz}	280	583 ± 4

Table 4.4 – Electronic polarisabilities of nanorod H/H-r and of a volume of bulk of the same size, resolved in three directions - $[11\bar{2}0]$ (x-direction), $[1\bar{1}00]$ (y-direction) and $[0001]$ (z-direction). The bulk value is calculate from the optical susceptibility $\chi(\infty)$. Error values represent one standard deviation.

0.0002, -0.0001, 0.0000, 0.0001, 0.0002 $Ha/(ea_0)$), within the linear response regime.

The polarisability is the gradient $\alpha_{ij} = dd_i/dE_j$.

The bulk polarisabilities are calculated by resolving the electronic susceptibility, $\chi(\infty)$ along the x , y and z directions, and scaling it by the volume of the nanorod. Results are summarised in Table 4.4.

We find that the electronic polarisability of the nanorod is enhanced by a factor of ~ 2 compared with bulk, in the direction parallel to the long axis of the nanorod (the $[0001]$ direction), but is suppressed in the two directions in the cross-sectional plane of the nanorod.

In Figure 4.21 we plot the local density of states (LDOS) of the slab of atoms at either end of nano rod, in the same way as in section 4.1, first in no electric field, and second in a large uniform electric field of $E = 0.05V/nm$ (the same large field used in section 4.5.2) .

We find that the energy shift between the two end slabs of the nano rod is almost unaffected by the applied electric field – consistent with the observation in section 4.5.2 that the nanorod’s internal electric field is nearly unaltered, despite the large (half the magnitude of the rod’s internal field) applied electric field. There is very small rearrangement of the polar surface states relative to the Fermi-level and therefore some redistribution of occupancies. The potential difference between the two surfaces remains nearly unchanged.

The polarisabilities in table 4.4 are calculated over a range of small values

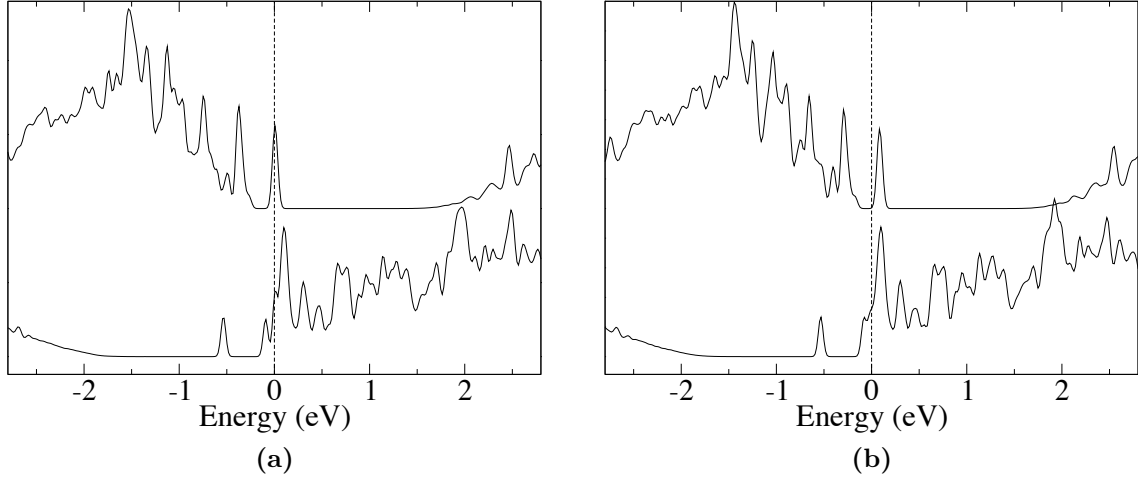


Figure 4.21 – LDOS plots of the Ga-rich polar surface (top) and As-rich polar surface (bottom) (a) with no applied electric field; and (b) in an applied uniform electric field of $E = +0.05$ V/nm

of the electric field, and over this range, we do not find any such rearrangement of the polar surface states relative to the Fermi-level, and no changes in occupancy of the surface states. This means that the potential difference between the nanorod ends, and therefore the resulting internal field (which remains at 0.10 V/nm over the range) is pinned over the range. The polarisation therefore arises from the polarisation of occupied orbitals throughout the nanorod, which serves to screen the applied field so as to preserve the potential difference and maintain the occupancies on the surface states. We discuss the possible origins of this increased polarisation later in this section.

As the applied electric field becomes larger, the energy cost of polarising the occupied orbitals to preserve the rod's internal potential difference becomes larger compared with the energy cost of changing the occupancies of the surface states, destabilising the position of the Fermi-level, and hence the occupancy shifting observed in Figure 4.21 becomes the preferred mechanism for stabilising the potential difference. This change in occupancy implies a redistribution of $+1e$ from the As-end to the Ga-end. In fact, we calculate from figure 4.20 a redistribution of only $+0.3e$, due to the polarisation response of the electrons in the nanorod bulk, which serves

to cancel out some of the charge redistribution from occupancy shifting. Though the potential difference across the nanorods vary very little between the rods, this small difference show up in the increase in the internal electric field in the centre of the nanorod, from -0.10 V/nm to -0.11 V/nm.

Polarisabilities α_x and α_y are lower in the nanorod than in bulk. This is not surprising for two overlapping reason: (1) the electrons which make up bonds are the most polarisable parts of a semiconductor, and in the finite x and y directions, the proportion of ‘missing’ bonds is high, due to the surfaces, reducing the number of highly polarisable elements in the crystal; and (2) in an extended system, the application of an electric field along a row of atoms induces a series of atomic dipoles in which the positive end of one induced dipole is attracted to the negative end of the neighboring dipole, thus reinforcing the polarization. The small size of system in the x and y directions, weakens this reinforcement effect, potentially reducing the polarisability.

Just as in section 4.1 between rods H/H and H/P, the potential difference across the rod H/H-r in zero field and the rod in a non-zero field remains nearly unchanged, and the Fermi-energy remains stabilised at almost the same position on either end of the rod. We have seen the internal electric field remain almost constant. The polarisability of a bulk semiconductor is typically not large enough to generate a depolarisation field that can almost completely screen out the external field. In the nanorod, it is large enough to do so, in the z direction. There are three possible sources for this enhanced polarisability: (1) the small width of the nanocrystal is likely, to result in a reduced energy cost of polarising the electrons, (it is harder to polarise planes of charge than point charges); (2) the lateral surfaces states may have a very large polarisability; (3) the position of the Fermi-energy relative to states on the polar surfaces is very stable, which fixes the potential difference across the nanorod and therefore determines the polarisability.

Though all three sources may play a role to varying degrees, we will find reason in section 4.5.5 to prefer the third, in the context of how the polarisability changes with nanorod length.

4. Nanorod charge distribution and electronic structure

Polarisabilities ($10^3 \times e^2 a_0^2 / E_h$)			
	H/H-r	H/B	B/B
α_z	583 ± 4	423 ± 9	556 ± 47

Table 4.5 – Polarisabilities

4.5.4 The effect of surface termination

In this section we calculate and compare the polarisabilities of nanorods with different polar and lateral surface terminations: specifically, rods H/H-r, H/B and B/B from section 4.1. Results are summarised in table 4.5.

The termination on both the polar and lateral surfaces appears to have very little influence on the polarisability.

4.5.5 The effect of size

In this section we calculate the polarisability per unit volume (polarisability density) in the z direction of nanorods of different size. This allows us not only to directly compare the values we get from nanorods of different lengths and widths, but also to compare it with the GaAs bulk susceptibility, which is also a polarisability density.

In figure 4.22a we find that the polarisability density increases linearly with nanorod length. If it were the case, as proposed at the end of section 4.5.3, that the enhancement in the polarisability of a nanorod in the z -direction compared with bulk was the result of the decreased electrostatic energy required to polarise a thinner stack of alternating charges, or the result of particularly polarisable lateral surface states (they can't be metallic, given their LDOS), then the polarisability density should be expected to remain constant with changes in nanorod length. The only one of the possibilities suggested at the end of section 4.5.3 that can explain the increase in polarisability density with nanorod length is the Fermi-level stabilisation argument: if the position of the Fermi-energy at the polar surfaces is very stable, due to the high density of surface states at E_f , the potential difference across the nanorod is pinned. When an electric field is applied, the electrons in a nanorod must respond to the potential difference that field produces across the

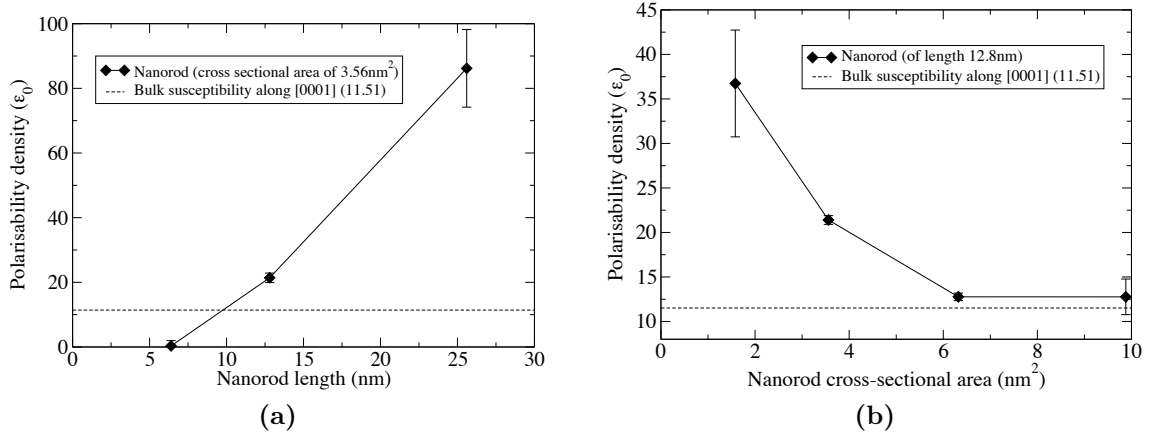


Figure 4.22 – The change in polarisability density nanorod H/H-r with length and cross-sectional area.

nanorod length in order to mitigate it. A nanorod of twice the length must respond to twice the applied potential difference, resulting in an accumulation of twice as much compensating charge on the surfaces. Hence the linear relationship of length with polarisability density.

From figure 4.22b we find that the polarisability density decreases rapidly with nanorod cross-sectional area towards a value close to (but probably slightly above) the bulk susceptibility. We have already seen in section 4.3 that as the nanorod becomes thicker, the relative stability of the position of the Fermi-level decreases, for a combination of reasons: the increase in energy required to polarise the bonds in the nanorod's bulk, the reduction of the decay rate of the nanorod's internal electric field due the reduction of 'fringe' loss, and the contraction of the energy window containing the nanorod's electronic states due to the reduced role of quantum confinement. The first two of these reasons are also important considerations for the polarisability, since the second fact will increase the amount of charge redistribution needed to fully compensate for the applied field, and the first means that a much higher energy barrier must be overcome to do so, which may mean that the 'occupancy-shifting' mechanism may play an increased role.

In figure 4.23 we compare the slab-by-slab LDOS of a thicker rod (of cross-sectional area 6.32 nm^2) in zero-field and in a large field of $E=0.05 \text{ V/nm}$, in order

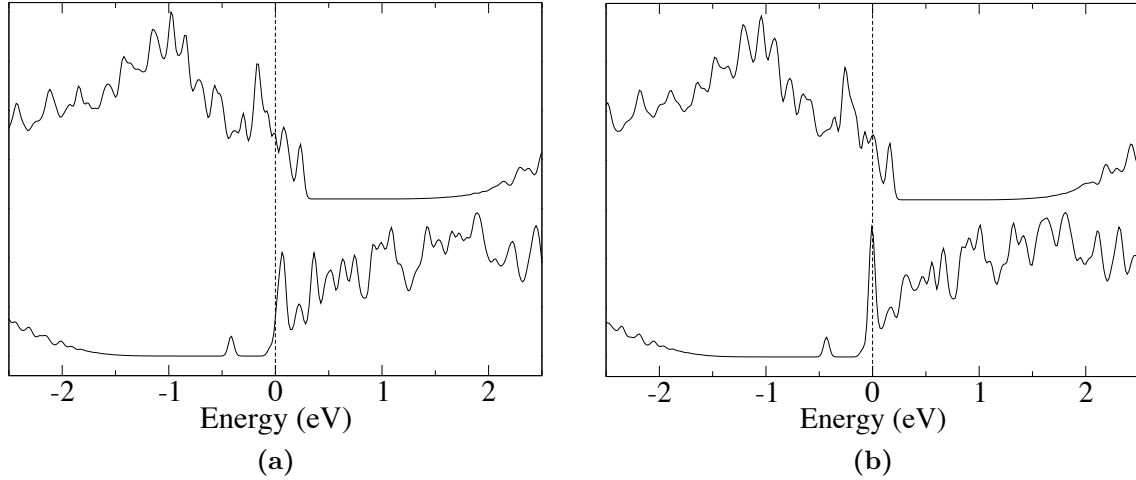


Figure 4.23 – LDOS (a) without and (b) with Efield.

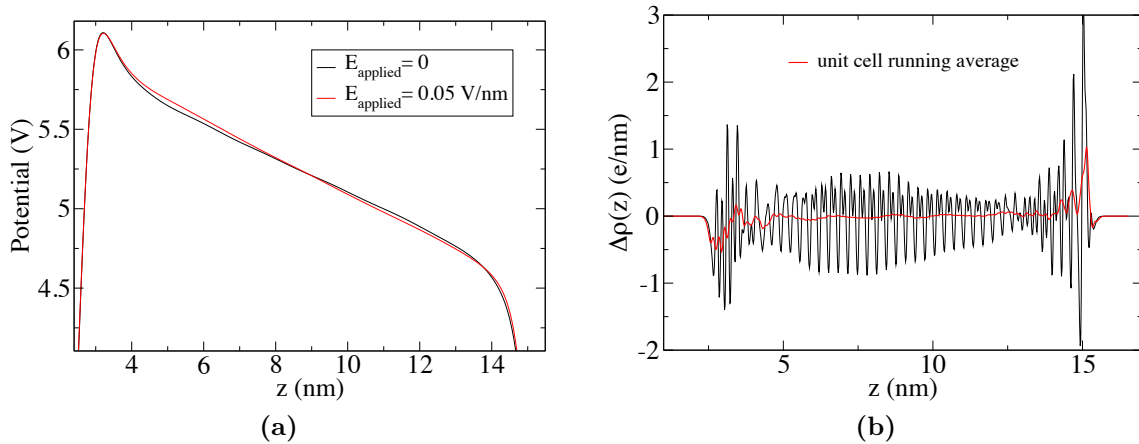


Figure 4.24 – (a) the potential across the rod. The potential difference remains almost identical, despite the applied field. The nanorod's internal field has increased slightly - 0.089 V/nm to -0.10 V/nm. (b) Difference between charge densities, at $E=0.05$ V/nm and $E=0$. A net charge of $+0.41e$ has been transferred from one end to the other as a result of the field, resulting in a change in dipole moment from -818 D to -610 D.

to see whether we can account for the smaller polarisability density of thicker rods. Again, just as for the thinner rod in figure 4.21, we find that the position of the Fermi-level at the surfaces is very stable. Figure 4.24a shows that the potential difference across the rod is almost identical.

Given, then, that occupancy shifting cannot explain the fast decline in polarisability density with thickness, since the Fermi-level position on the polar surfaces seems just as stable to an applied field in thick rods as in thin rods, we propose that it is likely to be caused by the faster decay of the internal polarisation field in thinner rods compared with thicker rods, due to the finite-size effect of ‘fringe loss’: in order to maintain the stabilised position of Fermi-level on the polar surfaces when a field is applied, the potential difference between these surfaces must remain the same. A much greater density of polarisation charge is required on the ends of the thinner rod, in response to the applied field, in order to maintain this potential difference, due to the faster decay of the induced potential in the thinner geometry.

In figure 4.24b we plot the difference in density generated by the electric field. We observe a transfer of $+0.41e$ from the As surface to the Ga surface, and a change in dipole moment from -818 D to -610 D. Though this is a larger transfer of charge than the $+0.3e$ found in the thinner nanorod of cross-sectional area 3.56 nm^2 , it is much smaller as a proportion of the cross-sectional area.

4.6 The effect of pairwise interactions between nanorods

4.6.1 Introduction

Polar nanorods in colloidal solution are known to self-assemble into a variety of different 1-, 2-, and 3-dimensional structures (see discussion in chapter 1). Within these structures, or while these structures are forming, nanorods with large permanent dipole moments and potentially large polarisabilities (as found in section 4.5) exist in close proximity to each other. Under such conditions, charge polarisation is likely to be significant both for the stability of the possible superstructures that can be

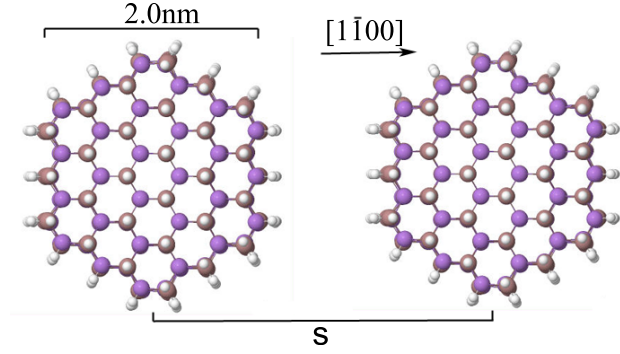


Figure 4.25 – Two aligned nanorods separated by a distance s . We are looking down on the (0001) surface of the nanorods.

formed, and for the kinetics of the self-assembly process, as well as other properties. Models of self-assembly of such nanorods and nanoparticles usually treat the charge distribution on a nanoparticle as fixed, which may not be a good approximation, particularly at a high density of nanoparticles. In this section, we consider the effect of close proximity between two nanorods on their charge distributions.

4.6.2 Results and discussion

We model two nanorods in a supercell, isolated from their periodic images using the cutoff-Coulomb potential. The nanorods we study are 12.8 nm in length, and 3.56 nm^2 in cross-sectional area i.e. the same ‘default’ size as in previous sections. Pseudo-ions are fixed at the equilibrium positions obtained by a geometry relaxation of an isolated nanorod in zero external field (nanorod H/H-r). We consider two different relative orientations for the two nanorods: (1) both nanorods lined up next to each other along $[1\bar{1}00]$, at a separation s (illustrated in figure 4.25) and oriented parallel to each other in all three spatial directions; (2) the same relative positioning in space, but with the second nanorod flipped with respect to the first in $[0001]$ (i.e. anti-parallel relative orientation in along $[0001]$), while remaining in parallel orientation along $[1\bar{1}00]$.

In Figure 4.26 we show how the component of the dipole moment along $[0001]$ for one of the rods in the pair varies with different values of the separation s . We

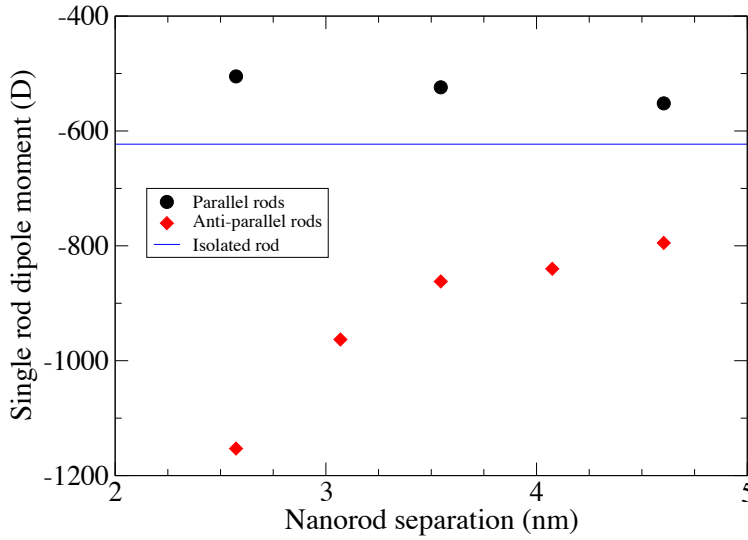


Figure 4.26 – The variation of the dipole moment of one of the pair of nanorods, with separation, in the aligned and anti-aligned configurations.

find that parallel rods have their dipole moments reduced as they get closer together, and anti-parallel rods see a much more prominent increase in their dipole moment as they get closer, with nearly a doubling of the dipole moment from its isolated value of -614 D, to a value of -1153 D at a separation of 2.57 nm between the nanorod's central axes i.e. a distance of 0.57 nm between the facing lateral surfaces of the nanorods.

When nanorods are parallel, the electric field produced by each one serves to depolarise its neighbour, whereas in the anti-parallel configuration, there is a mutual reinforcement of each others' dipole moment. We can understand something about the character of this interaction by considering the following simple model: rod 1 produces a field which is felt by rod 2, and vice-versa. If we assume very little variation of the electric field of rod 1 over the volume of space occupied by rod 2, then we may express the electric field due to rod 1 at the location of rod 2 in terms of the dipole moment d of rod 1 as

$$E = -\lambda(s)d$$

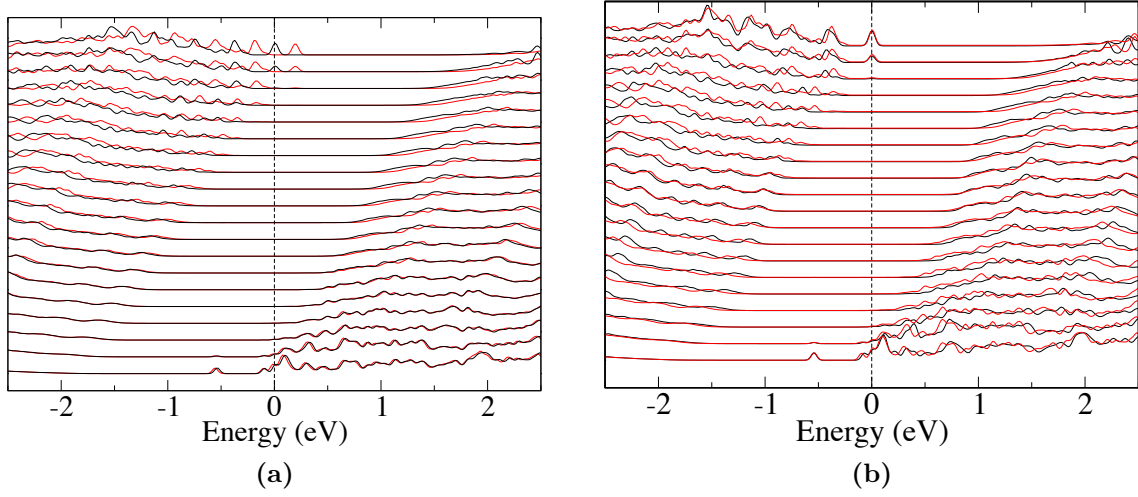


Figure 4.27 – The LDOS of an isolated rod (black) and of one of the pair of nanorods separated by a distance 2.57 nm between the rod’s central axes (red) in the cases where the rods are (a) aligned; and (b) anti-aligned

where $\lambda(s)$ is a positive constant which decreases with nanorod separation s . The dipole moment of rod 2 changes when an external electric field is applied like

$$d = d_0 + \alpha E$$

assuming we are in the linear regime, where α is the electric linear-polarisability in the z -direction (a positive constant) and d_0 is the dipole moment in zero field. Substituting, we obtain the self-consistent dipole moment. When the rods are aligned $|d|$ reduces as the rods are brought closer like,

$$|d| = \frac{d_0}{1 + \alpha\lambda(s)}. \quad (4.1)$$

When the rods are anti-aligned, $|d|$ increases towards a singularity (at $\alpha\lambda(r) = 1$) as the rods are brought closer together, like

$$|d| = \frac{d_0}{1 - \alpha\lambda(s)}. \quad (4.2)$$

Though this model appears to account qualitatively for the observe polarity

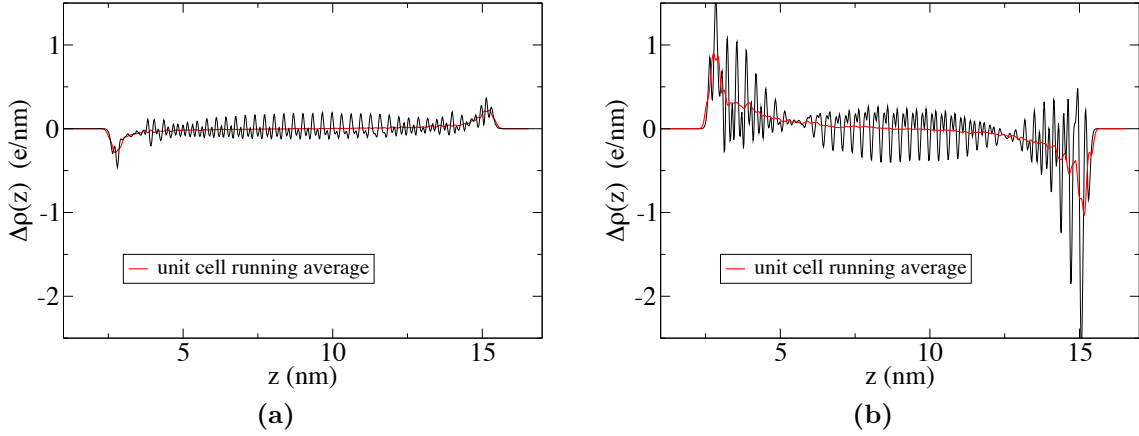


Figure 4.28 – Difference in electron density along the length of the rod between one of the pair of nanorods separated by a distance 2.57 nm between the rod’s central axes, and the isolated rod, in the cases where the rods are (a) aligned; and (b) anti-aligned. The charge transfer from the the left hand end to the right of the rod is 0.1e in (a), and -1.15 in (b).

behaviour, the LDOS plots of one of the rods in the pair, in both the aligned and anti-aligned cases in Figures 4.27 (a) and (b) reveals that there is more to consider. In the case of aligned rods, there has been a shift in the position of the Fermi-level relative to the surface states, resulting in a shift in occupancies compared with that of the isolated rod. The occupied mid-gap state on the Ga-rich end of the rod has become unoccupied, and electrons transferred to a state on the As end. This has allowed an increase in the potential difference across the nanorod, represented by the shift in the LDOS of the top slab of 0.18eV.

In contrast, in the anti-aligned case, (Figure 4.27(b)), there is no shift in the potential difference across the nanorod, or the position of the nanorod, or the internal field, despite the large field’s from the neighbouring rod. The redistribution of charge has completely screened out the field from the neighbouring rod, and this redistribution is wholly the result of the mechanism of polarisation of already occupied orbitals, and not from the mechanism of occupancy shifting.

In Figure 4.28 we show the difference in charge density between one of the rods in the pair, and the isolated rod. We find a transfer of charge from the As-rich

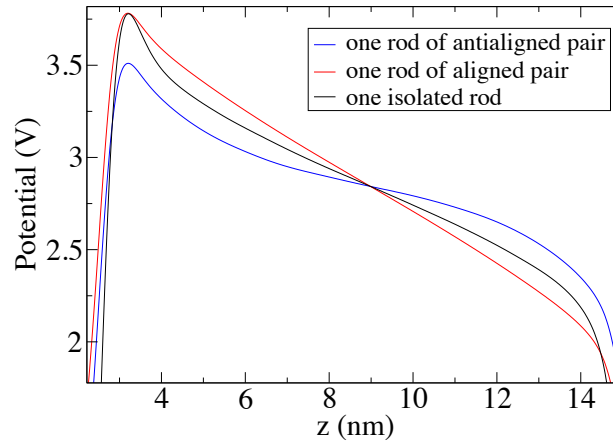


Figure 4.29 – The potential across the nanorod when aligned with (at separation 2.57 nm), anti-aligned with, and isolated from another nanorod. The decay of the internal field is most prominent when rods are anti-aligned, and least prominent when aligned.

end to the Ga-rich end of $+0.1e$, in the case of aligned rods, and $-1.15e$ in the case of anti-aligned rods.

We have seen in section 4.5.5 how the polarisability density of a nanorod can be dependent on the decay of the polarisation induced potential within the rod: thin rod geometries which result in faster decay of the induced depolarisation field require more charge transfer per unit surface area in order to maintain the potential difference between the polar ends of the nanorod – required by the stability of the Fermi-level at the surfaces. In Figure 4.29 we plot the potential across one of the nanorods in the pair, for each case, and compare with that of an isolated rod. We find that a nanorod which is anti-aligned with its neighbour has an internal electric field which decays faster than in an isolated nanorod, which, in turn, decays faster than that of a rod aligned with its neighbour. This is due to the mutual cancellation of the polarisation field produced when anti-aligned rods polarise. The result is that anti-aligned rods must exhibit a higher polarisability, and they are able to endure far high electric fields from their neighbours before the occupancy shifting mechanism of charge polarisation is preferred over the bond polarisation mechanism, resulting in an extremely stable Fermi-level position.

This picture, which predicts that the polarisability of the nanorod is itself

dependent on its interaction with nearby polarisable rods, suggests that the polarisability α that appears in Equations 4.1 and 4.2 must be higher in the anti-aligned case than in the aligned case. We test this using our data from Figure 4.26. Using more specific labels for the polarisability, $\alpha_{\uparrow\uparrow}$ and $\alpha_{\uparrow\downarrow}$, to denote aligned and anti-aligned rods respectively, we eliminate $\lambda(s)$ in the pair of simultaneous equations 4.1 and 4.2, to obtain

$$\frac{\alpha_{\uparrow\uparrow}}{\alpha_{\uparrow\downarrow}} = \frac{\frac{d_0}{d_{\uparrow\uparrow}} - 1}{1 - \frac{d_0}{d_{\uparrow\downarrow}}}, \quad (4.3)$$

which holds at fixed s . Substituting the values of $d_{\uparrow\uparrow}$ and $d_{\uparrow\downarrow}$, at $s=2.57$ nm, 3.55 nm and 4.60 nm, we obtain three independent estimates for the ratio $\alpha_{\uparrow\uparrow}/\alpha_{\uparrow\downarrow}$ of 0.51, 0.68, and 0.59 respectively – confirming the prediction that $\alpha_{\uparrow\uparrow} < \alpha_{\uparrow\downarrow}$.

4.7 Summary and conclusion

In this chapter we have sought to use developments in linear-scaling DFT, implemented in the ONETEP code, to simulate polar semiconductor nanorods of realistic size, with a particular focus on investigating the factors affecting a nanorod’s charge distribution. We have investigated the effect of surface terminating species (section 4.1), geometry relaxation (section 4.2), nanorod size (section 4.3), semiconductor type (section 4.4), external electric fields (section 4.5), and interactions with other nanorods (section 4.6).

Over the course of this chapter, a particular explanatory picture has emerged which has allowed us to rationalise the observed patterns in the data from our simulations. In this section we summarise the most significant findings from each of the sections in this chapter, we discuss their significance for nanorod synthesis and technological application, and we describe this explanatory picture which ties many of the observations together.

In sections 4.1 and 4.2 we found that the nanorod’s dipole moment depends strongly on its surface termination – particularly on its polar surfaces. Full hy-

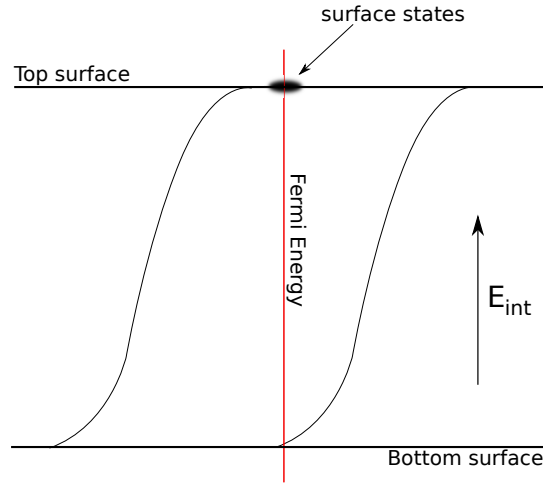


Figure 4.30 – Typical band-edge diagram of the nanorods in this work. The top and bottom surfaces here represent the end polar surfaces of the nanorod.

drogen, or pseudo-hydrogen termination on polar surfaces produces negative dipole moments, bare polar surfaces produce positive dipole moments, and a fractional coverage of the surface with hydrogen atoms produce dipole moment's somewhere in between. The spontaneous polarisation of the underlying wurtzite crystal lattice appears to play a much less important role than has been assumed in much of the literature (reviewed in chapter 1). These observations suggest that the dipole moments of nanorods may be tuneable, if sufficiently different levels of surface coverage are attainable by growth under different environmental conditions.

The potential difference across a nanorod due to its large dipole moment, shows up in the local density-of-states as a shifting of the energy of the states as one moves along the length of the rod. In all of the nanorods, we found that at the polar (end) surfaces, the Fermi-level is stabilised at an energy coinciding with a high density of mid-gap states, or if there are no mid-gap states, it coincides with the high density of states found at or near the band-edges. We give a typical example of this using a band-edge diagram in Figure 4.30.

These are very stable positions for the Fermi-level because small deviations from these positions would result in changes in occupancy, resulting in a redistribution of potentially a large amount of charge, generating a potential opposing the

initial small change.

The band-gap, then, effectively imposes an upper limit on the potential difference across the nanorods, beyond which it is very difficult to exceed because of the large density of states within the bands. This is why we see reduced dipole moment magnitudes in nanorod which have, for whatever reason, a narrower band-gap; in sections 4.1 we found that rods with large numbers of mid-gap polar surface states had lower dipole moment magnitudes; in section 4.2, we saw that the presence of lateral-surface states in the gap reduces the dipole moment magnitude, and relaxation of these lateral surfaces opens up the band-gap, therefore increasing the dipole moment magnitude; in section 4.4 we found that semiconductors, such as AlN, with large band-gaps support much larger polarisation magnitudes than smaller band-gap semiconductors, such as GaAs. Additionally, we find some evidence in section 4.3 for the possibility that the effect of quantum confinement in very thin nanorods – which increases the size of the band-gap – may be at least partially responsible for the increased polarisation seen in thinner nanorods.

Simple bond-electron counting models cannot account for the magnitudes of the excesses of charge on the ends of the nanorod. These models greatly overestimate these excesses because they do not account for the restrictions imposed by this Fermi-level stabilisation effect, or for the polarisability of the electrons in nanocrystal.

In section 4.5 we found nanorods to be much more polarisable in a uniform electric field than their bulk counterparts along the long axis of the nanorod, but much less polarisable along their shorter axes. The enhanced polarisability in small fields is likely to be due to the much faster decay of the induced depolarisation field in thin rods, due to finite size effects (c.f. “fringe loss” in a capacitor with small plate area), which necessitates a much larger increase in charge density on the ends in order to maintain the potential difference between the ends of the rod, which is fixed by the stabilised position of the Fermi-level on the polar surfaces, resulting from the significant density of states at the Fermi-level on those surfaces.

For thin rods in small fields, bonds within the nanocrystal polarise so as to

entirely screen the applied field from the nanorod's bulk, which maintains the position of the Fermi-level and prevents occupancy shifting. Only at strong fields does the energy cost of polarising the bonds become sufficiently large so that a shift in the position of the Fermi-level with respect to the stabilising polar surface states (occupancy shifting) becomes the preferred mechanism for stabilising the potential difference between the nanorod ends. In thicker rods, the bond polarisability is lower (i.e. there is a much faster decay in the induced depolarisation field) due to the reduction of finite size effects, and therefore; (1) the potential-difference stabilising mechanism of occupancy shifting becomes more significant relative to bond-polarisation; (2) relatively much less extra polar surface charge density is required to maintain the potential difference across the rod, than for thinner rods.

In very much stronger fields, or in very much bigger nanocrystals, more states cross the Fermi-level, and the polar surfaces become metallic, and are likely to be very unstable. Such surfaces are likely to reconstruct so as to remove the large potential difference generating the instability – consistent with the known phenomenon in semi-infinite polar surfaces of the ‘polar instability’, which says that charged semi-infinite surfaces are unstable to reconstruction.

In section 4.6 we investigated the effect of interaction with a neighbouring nanorod on polarity. We found that the dipole moment is slightly reduced when it is located beside a nanorod aligned parallel with it, and it is greatly enhanced when located beside a nanorod which is anti-aligned with it, as their electric fields mutually reinforce each other's dipole moments. Also, the polarisability itself of each rod is enhanced in the anti-aligned configuration, and reduced in the aligned configuration, due to the reinforcement (cancellation) of the internal induced depolarisation field in the aligned (anti-aligned) case, which also destabilised (stabilised) the position of the Fermi-level on the polar surfaces. These are important effects to consider for both nanorod self-assembly and in nano particle superlattices – particularly those exhibiting anti-ferroelectric ordering (reviewed in chapter 1).

A significant amount of the analysis in this thesis relies on results obtained from nanorods in which the Fermi-energy coincides with the energies of a significant

density of states on the polar (end) surfaces of the nanorod, which we have suggested play a significant role in determining many of their polar properties. This scenario may or may not represent the majority of polar nanocrystals found in nature, under real experimental conditions. It may be the case that the majority of nanorods in nature exhibit a kind of surface reconstruction that leaves the internal potential difference across the nanorod significantly smaller than its band-gap, and leaves the Fermi-level on the polar surfaces far away from any polar surface states. Under these conditions, the polar properties may be significantly different in some respects to the nanorods we have been studying here. It would therefore be an interesting topic for future investigation to see whether there are significant differences in the phenomena we have been concerned with, in such nanorods.

The most important question in this area, in my opinion, is the question of the stability of polar nanostructures: investigating what the constraints of thermodynamic stability mean for the polar properties of nanocrystals, and how they are affected by environmental conditions. This is a question we have not significantly touched on in this thesis.

A

Simulation details and preliminary calculations

It is important to ensure that our simulations capture the physics of bonding and respond correctly to polarisation. Previous DFT studies of compounds containing Gallium have found that accurate treatment of the response of its $3d$ electrons to valence charge is essential for obtaining accurate structures [108, 109]. In particular, pseudisation of $3d$ electrons in the the core produce inaccurate structures [110]. The solutions to this are either to include the $3d$ electrons explicitly (which is computationally expensive), or to use non-linear core corrections. We generate Ga and As pseudopotentials with $3d$ electrons frozen in to the core using the ‘on-the-fly’ generator in the CASTEP code [111], and test them with NLCC in the CASTEP code. Total energies and relaxed lattice parameters are presented at a variety of cutoff energies in Table A.1. We find that at a cutoff energy of 400 eV, the total energy is converged to 0.020% of the 800 eV value. This represents a convergence of the total energy at 400 eV to within 0.004 eV per atom. The c lattice parameter is converged to 0.020% of 800 eV value, c/a converged to 0.0018% and u converged to 0.0003%.

We find $\sim 1.3\%$ overbinding on a and c compared with the experimental values, which is typical of LDA, and we find an extremely accurate c/a ratio. We also

A. Simulation details and preliminary calculations

Ecutoff (eV)	Etot (eV)	a (Å)	c (Å)	c/a rat.	u (frac)
300	-923.684017	3.938078	6.491377	1.64836	0.373539
350	-923.747476	3.936210	6.488742	1.64847	0.373529
400	-923.755040	3.935243	6.487250	1.64850	0.373526
450	-923.757946	3.934981	6.486765	1.64849	0.373526
500	-923.762422	3.934840	6.486460	1.64847	0.373527
550	-923.765895	3.934709	6.486297	1.64848	0.373525
600	-923.767826	3.934695	6.486305	1.64849	0.373527
650	-923.768873	3.934684	6.486056	1.64843	0.373527
700	-923.769440	3.934546	6.486029	1.64848	0.373528
750	-923.770119	3.934480	6.485973	1.64850	0.373527
800	-923.770972	3.934464	6.485926	1.64849	0.373527
300K expt		3.989	6.564	1.64550	0.374

Table A.1 – Convergence data for the total energy and lattice parameters

tested the pseudopotentials on Ga_2 and As_2 dimers. At 400eV, the As_2 bond-length was calculated as 2.082 Å (1.8% smaller than the experimental value in Ref. [112]), and the Ga_2 was calculated as 2.373 Å (we weren't able to find an experimental bond length for Ga_2 for comparison). The pseudopotentials appear to display good transferability.

Reference [113] compares the convergence properties of the total energy and the forces in linear-scaling ONETEP and plane-wave method CASTEP, and finds them to be broadly similar, though the cutoff energies need to be slightly higher (by $\sim 15\%$) in ONETEP for the same level of convergence.

For our calculations of GaAs nanorods using ONETEP, we use cut-off energies of between 450 eV and 520 eV, throughout this thesis. The total energy of our nanorod system in ONETEP is found to have converged to within a tolerance of 0.03 eV per atom at a cutoff energy of 450 eV. This is more than adequate for our purposes, since we are using GaAs nanorods as a vehicle for investigating general principles governing nanorod polarity, rather than in order to obtain extremely accurate values for, say, the energy, or other properties. We use a very generous radius for the localisation region of the orbitals [107], of 0.53 nm (10 a_0) for each atomic species, and the density kernel is set at 1000 a_0 (i.e. effectively infinite) so

A. Simulation details and preliminary calculations

that metallic and insulating structures are treated on an equal footing.

B

Considerations of nanorod stability and environment

In the sections above, we have studied nanorods which support large dipole moments and large internal electric fields, without reference to the thermodynamic stability of these structures. Some of the most important outstanding questions in this area involve how polarization is stabilised in such systems given the high energetic cost of maintaining fields through the system and the multitude of available mechanisms for compensating fields under standard conditions in colloidal solutions. Or, more specifically, given a surface termination (say, Ga-rich, or H-covered) and an environment (i.e., some solvent or gas), it would be very useful to be able to determine what the minimum free-energy surface structures would be, and how that affects the polar behavior. For example, we have seen in section 4.1 how different surface termination (in particular, different levels of hydrogen coverage on the polar surfaces) can greatly change the nanorod's dipole moment, but if only a very small subset of possible surface configurations are thermodynamically stable under attainable environmental conditions, then the range of possible nanorod polarities which can exist may be equally limited.

In this section, we do not attempt to pursue the broad problem of determining nanorod structures which minimise the free energy in a given environment, since as

we will see, it is a problem with far too many variables for the amount of time and computational resources available. Instead, we use the techniques of ab-initio thermodynamics and some very restrictive assumptions to address the much less ambitious aims of trying to determine: (1) whether there exists a stable level of hydrogen coverage on the polar surfaces; and (2), whether this level of coverage has significant variability over a realistic range of environmental hydrogen chemical-potentials, so that a significant range of nanorod dipole moments may be accessible.

Equilibrium, under constant temperature and pressure, is defined as the configuration which minimises the Gibbs free energy

$$G = H - TS$$

where H is the enthalpy, T the temperature, and S the entropy.

We assume that a realistically attainable environment in which a nanorod may be found, is one in which water molecules are thermodynamically stable.

From the fact that oxygen and hydrogen in their elemental forms combust to produce water at standard temperature and pressure, we can be sure that the chemical potential per atom of both species is lower in water than it would be in their elemental form. Written explicitly, we have

$$\mu_O(T) \leq \mu_O^0(T) \tag{B.1}$$

$$\mu_H(T) \leq \mu_H^0(T) \tag{B.2}$$

where $\mu_O^0(T)$ and $\mu_H^0(T)$ are the chemical potentials per atom of oxygen and hydrogen in their elemental form (essentially, the Gibbs free energy per particle in their elemental form), and $\mu_O(T)$ and $\mu_H(T)$ are the chemical potentials per atom in water.

The free energy of formation of water is defined

$$\Delta G_f^{H_2O} = 2\mu_H + \mu_O - 2\mu_H^0 - \mu_O^0 \tag{B.3}$$

B. Considerations of nanorod stability and environment

Molecule	E_{DFT}	E_{f+v}
H ₂ O	-17.1159	0.17768
O ₂	-31.7193	0.00278
H ₂	-1.1364	0.00038

Table B.1 – Total energy of molecule, calculated in DFT, and the enthalpy of vaporisation plus fusion per molecule, obtained from Ref. [6]

Substituting the inequality in equation B.1 in to the expression in equation B.3 gives a lower bound to the chemical potential of hydrogen in water, and equation B.2 gives an upper bound so that the μ_H is confined in the range

$$\frac{1}{2}\Delta G_f^{H_2O}(T) + \mu_H^0(T) \leq \mu_H(T) \leq \mu_H^0(T)$$

At $T = 0$ K, $p_{O_2} = p_{H_2} = 0$, therefore $\mu_H^0 = \frac{1}{2}E_{DFT}[H_2]$, $\mu_O^0 = \frac{1}{2}E_{DFT}[O_2]$ and $\Delta G_f^{H_2O} = E_{DFT}[H_2O] - E_{DFT}[H_2] - \frac{1}{2}E_{DFT}[O_2]$.

These DFT calculations are carried out on isolated H₂, O₂ and H₂O molecules, and not a solid phase, as they are found at $T = 0$ K. In order to estimate the magnitude of the error from this approximation, we will consider the correction due to the enthalpy of fusion plus vaporisation, summarised along with the results of the calculations of the enthalpy of the molecules in table B.1.

Calculating these leaves

$$-0.6281 \leq \mu_H \leq -0.5683 \text{ Ha}$$

at $T = 0$ K, and we approximate that we have overestimated the lower bound, as a result of not accounting for the enthalpy of vaporisation and fusion of water by ~ 0.0178 Ha

Clearly, by setting $T = 0$ K, we have moved away from our desired objective of calculating a plausible range for the chemical potential of hydrogen in a *realistic* environment. The chemical potentials at finite T and p will be different, and will include contributions from the vibrational and configurational entropy. It is possible to account for some of the effects of temperature and pressure using an ideal gas

B. Considerations of nanorod stability and environment

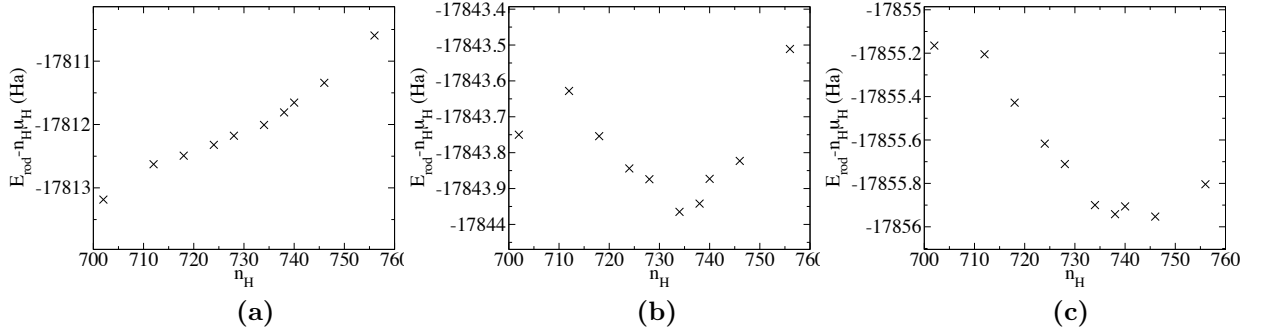


Figure B.1 – (a) lower bound, $\mu_H = -0.6281$ (b) a mid-range value, $\mu_H = -0.5845$ (c) upper bound, $\mu_H = -0.5683$

approximation, however we do not do so here. We continue, despite the inadequacy of setting $T = 0$.

In Figure 4.2 in chapter 4, we saw how the dipole moment varied with different surface polar surface coverage on the nanorod. Here, we use that same data set and we attempt to calculate which of the surface coverages is most stable over the range of ‘realistic’ chemical potentials calculated above. In Figure B.1 we plot the enthalpy ($E_{DFT}[\text{nanorod}] - n_H \mu_H$) of the nanorods as a function of n_H , over the range $n_H = [702 : 757]$ which corresponds to fractional polar surface coverages from 0 to 1. We do this at three different chemical potentials μ_H , over the ‘realistic’ range: the lower bound, the upper bound and a mid-range point of $\mu_H = -0.5845$ Ha.

This enthalpy differs with the enthalpy of formation of the nanorod by the same constant in each nanorod (namely the constant $-n_{Ga}\mu_{Ga} - n_{As}\mu_{As}$), therefore, the rod which minimises this enthalpy is the most stable.

Figure B.1 shows that, of the structures studied, the most stable value for the hydrogen coverage varies over a wide range, within the range of ‘realistic’ chemical potentials. At a mid-range value $\mu_H = -0.5845$ (Figure B.1(b)), the stable configuration has 11 hydrogen atoms missing from the 27 on each of the polar surfaces, which corresponds to a dipole moment of $\sim 20D$. At the upper bound of $\mu_H = -0.5683$, the stable configuration has 7 hydrogen atoms missing from each of the polar surfaces, corresponding to a dipole moment of $\sim -200D$.

Of course, there is no guarantee (in fact, it is quite unlikely) that the nanorod

with the minimum enthalpy of formation of the structures considered here, is stable with respect to changes other than the highly restricted variational freedom permitted in this analysis. The restrictions we have imposed are (1) only changes in the hydrogen composition are permitted (not Ga or As); (2) only on the polar surfaces (not the lateral surfaces); (3) the hydrogen composition of the two opposing polar surfaces must remain identical; (4) atoms must remain at their ‘ideal’ bulk-like positions (i.e. no geometry relaxation is allowed).

Nevertheless, despite the crudeness of this analysis, it may lend some advocacy to the suggestion in chapter 4 that it may be possible to tune the nanorod’s dipole moment over a fairly wide range, by varying its surface termination by tuning the environmental conditions of synthesis.

C

Description of relaxed geometries

The nature of the relaxation on the Ga polar surface on nanorod b/b is as follows: the Ga layer has sunk into the plane of the As layer below, resulting from a rehybridisation of the 4 sp^3 orbitals on the Ga atom to 3 sp^2 orbitals and a p orbital, so that the Ga atoms bond to their neighbouring As atoms with stronger $sp^3 - sp^2$ -like bonds, which are planar.

The Ga surface on rod h/b exhibits 4 Ga atoms that protrude a large distance outward from their original positions in the surface, making an angle of $39.8 \pm 1^\circ$ with their neighbouring As atoms and the plane perpendicular to the c-axis (the y-z plane in this work), compared with the ideal bulk value of 19.5° . The other Ga atoms on this surface have sunk into the As layer below, so that the angle between these, their neighbouring As atoms, and the y-z plane is only $1.70 \pm 0.4^\circ$. The As-terminated polar surface on the opposite end of the nanorod shows signs of a similar tendency to eject As atoms, but much less dramatically, with the angles between As surface atoms, Ga neighbouring atoms and the y-z plane ranging from 4.0° to 27.9° . In the former case (Figure 4.12(b/b)), we don't see ejection of any atoms on either of the polar surfaces. Instead, we see on the Ga terminated end, the sinking of the Ga layer into the As layer, so that for 9 out of the 12 surface Ga atoms, the angle between them, their neighbouring As atoms and the y-z plane is $6.4 \pm 1^\circ$. The three anomalous Ga atoms lie on the corners of the nanorod, and have not

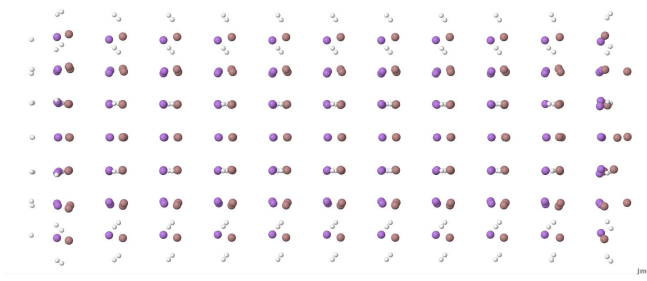


Figure C.1 – Geometry relaxation of rod with H-terminated As-rich polar surface and bare Ga-rich surface

changed position significantly from the bulk value, with an angle of $20.5 \pm 1^\circ$. On the As terminated end, there has been no significant reconstruction, with angles in the range $19.9 \pm 1^\circ$.

We consider the possibility that the reason for the qualitative difference between polar surface reconstructions in b/b and h/b may have to do with the fact that the surfaces are much more strongly charged in h/b compared with b/b, since it has a much larger dipole moment: it may be that the surface energy would be reduced in h/b by trying to neutralise the charge on the surfaces by ejecting cations (anions) from positively (negatively) charged surfaces - a typical reconstruction of polar surfaces. In b/b, surface energy is reduced by this re-hybridisation instead, since the charge accumulated on the surfaces isn't large enough to generate the behaviour seen in h/b.

This analysis is shown to be wrong (or at least, unlikely) by the rod in Figure C.1, which is a rod which is hydrogen terminated on all surfaces but the Ga-rich polar surface. This rod has a dipole moment opposite to that of h/b of -45 D (compared with h/h's +67 D) but it undergoes the same sort of relaxation i.e. it is exhibiting the same tendency to eject 4 of the Ga atoms - contrary to the arguments related to the polar instability. This relaxation instead may be driven by the H terminated bonds on corners of this nanorod's polar surfaces, which is also a feature of nanorod h/b, but does not feature in nanorod b/b.

In contrast to these bare surfaces, H terminated surfaces show no sign of instability - ionic positions remain remarkably close to their pristine values on both

the lateral and the polar surfaces, despite the fact that the H terminated polar surfaces on h/h and b/h (relaxed) support the largest magnitude of surface charge of the four nanorods studied. H passivation appears to have a strong stabilising effect, especially on the polar surfaces.

References

- [1] J. J. Shiang, A. V. Kadavanich, R. K. Grubbs, and A. P. Alivisatos. Symmetry of annealed wurtzite cdse nanocrystals: Assignment to the $c3v$ point group. *The Journal of Physical Chemistry*, 99(48):17417–17422, 1995.
- [2] Jason Taylor, Tadd Kippeny, and Sandra J. Rosenthal. Surface stoichiometry of cdse nanocrystals determined by rutherford backscattering spectroscopy. *Journal of Cluster Science*, 12:571–582, 2001. 10.1023/A:1014246315331.
- [3] Walter A. Harrison. Theory of polar semiconductor surfaces. *Journal of Vacuum Science and Technology*, 16(5):1492–1496, September 1979.
- [4] Jacek Goniakowski, Fabio Finocchi, and Claudine Noguera. Polarity of oxide surfaces and nanostructures. *Reports on Progress in Physics*, 71(1):016501, 2008.
- [5] Arash A. Mostofi. On linear scaling methods for quantum mechanical first-principles calculations. *Thesis*, 2003.
- [6] National institute of standards and technology (nist) online database.
- [7] Mostafa A. El-Sayed. Small is different: shape-, size-, and composition-dependent properties of some colloidal semiconductor nanocrystals. *Accounts of Chemical Research*, 37(5):326–333, 2004. PMID: 15147173.
- [8] T. Vossmeier, L. Katsikas, M. Giersig, I. G. Popovic, K. Diesner, A. Chemseddine, A. Eychmueller, and H. Weller. Cds nanoclusters: Synthesis, characterization, size dependent oscillator strength, temperature shift of the excitonic transition energy, and reversible absorbance shift. *The Journal of Physical Chemistry*, 98(31):7665–7673, 1994.
- [9] A. N. Goldstein, C. M. Echer, and A. P. Alivisatos. Melting in semiconductor nanocrystals. *Science*, 256(5062):1425–1427, 1992.
- [10] Dmitri V. Talapin, Elena V. Shevchenko, Christopher B. Murray, Alexey V. Titov, and Petr Král. Dipole-dipole interactions in nanoparticle superlattices. *Nano Letters*, 7(5):1213–1219, 2007. PMID: 17397231.

-
- [11] Zhiyong Tang, Nicholas A. Kotov, and Michael Giersig. Spontaneous organization of single cdte nanoparticles into luminescent nanowires. *Science*, 297(5579):237–240, 2002.
- [12] S. C. Glotzer, M. J. Solomon, and N. A. Kotov. Self-assembly: From nanoscale to microscale colloids. *AIChE Journal*, 50(12):2978–2985, 2004.
- [13] Zhihong Nie, Alla Petukhova, and Eugenia Kumacheva. Properties and emerging applications of self-assembled structures made from inorganic nanoparticles. *Nat Nano*, 5(1):15–25, 01 2010.
- [14] X. Michalet, F. F. Pinaud, L. A. Bentolila, J. M. Tsay, S. Doose, J. J. Li, G. Sundaresan, A. M. Wu, S. S. Gambhir, and S. Weiss. Quantum dots for live cells, in vivo imaging, and diagnostics. *Science*, 307(5709):538–544, 2005.
- [15] Nir Tessler, Vlad Medvedev, Miri Kazes, ShiHai Kan, and Uri Banin. Efficient near-infrared polymer nanocrystal light-emitting diodes. *Science*, 295(5559):1506–1508, 2002.
- [16] M. Kazes, D.Y. Lewis, Y. Ebenstein, T. Mokari, and U. Banin. Lasing from semiconductor quantum rods in a cylindrical microcavity. *Advanced Materials*, 14(4):317–321, 2002.
- [17] Wendy U. Huynh, Janke J. Dittmer, and A. Paul Alivisatos. Hybrid nanorod-polymer solar cells. *Science*, 295(5564):2425–2427, 2002.
- [18] Chuan-Jian Zhong, Jin Luo, Bin Fang, Bridgid N Wanjala, Peter N Njoki, Rameshwori Loukrakpam, and Jun Yin. Nanostructured catalysts in fuel cells. *Nanotechnology*, 21(6):062001, 2010.
- [19] SurendraK. Martha, Judith Grinblat, Ortal Haik, Ella Zinigrad, Thierry Drenzen, JamesH. Miners, Ivan Exnar, Andreas Kay, Boris Markovsky, and Doron Aurbach. Limn0.8fe0.2po4: An advanced cathode material for rechargeable lithium batteries. *Angewandte Chemie International Edition*, 48(45):8559–8563, 2009.
- [20] Andreas Moser, Kentaro Takano, David T Margulies, Manfred Albrecht, Yoshiaki Sonobe, Yoshihiro Ikeda, Shouheng Sun, and Eric E Fullerton. Magnetic recording: advancing into the future. *Journal of Physics D: Applied Physics*, 35(19):R157, 2002.
- [21] Dorota Bartczak, Otto L. Muskens, Timothy M. Millar, Tilman Sanchez-Elsner, and Antonios G. Kanaras. Laser-induced damage and recovery of plasmonically targeted human endothelial cells. *Nano Letters*, 0(0), 0.
- [22] *Nanoparticles: Synthesis, Stabilization, Passivation, and Functionalization*, Copyright, Foreword, chapter 0, pages i–v.

-
- [23] E. Ciliberto C. Altavilla. *Inorganic Nanoparticles: Synthesis, Applications, and Perspectives*. Taylor and Francis Group, 2011.
- [24] Conyers Herring. Some theorems on the free energies of crystal surfaces. *Phys. Rev.*, 82(1):87–93, Apr 1951.
- [25] P.D. Cozzoli, T. Pellegrino, and L. Manna. Synthesis, properties and perspectives of hybrid nanocrystal structures. *Chem. Soc. Rev.*, 35:1195–1208, 2006.
- [26] Iwan Moreels, Bernd Fritzing, José C. Martins, and Zeger Hens. Surface chemistry of colloidal pbse nanocrystals. *Journal of the American Chemical Society*, 130(45):15081–15086, 2008.
- [27] Deborah M. Aruguete, Matthew A. Marcus, Liang-shi Li, Andrew Williamson, Sirine Fakra, Francois Gygi, Giulia A. Galli, and A. Paul Alivisatos. Surface structure of cdse nanorods revealed by combined x-ray absorption fine structure measurements and ab initio calculations. *The Journal of Physical Chemistry C*, 111(1):75–79, 2007.
- [28] Roby Cherian and Priya Mahadevan. Size dependence of lattice constants of semiconductor nanocrystals. *Applied Physics Letters*, 92(4):043130, 2008.
- [29] Aaron Puzder, A. J. Williamson, François Gygi, and Giulia Galli. Self-healing of cdse nanocrystals: First-principles calculations. *Phys. Rev. Lett.*, 92(21):217401, May 2004.
- [30] P Atkins. *Physical Chemistry*. Oxford University Press, 7th edition edition, 2002.
- [31] J. P. Foster and F. Weinhold. Natural hybrid orbitals. *Journal of the American Chemical Society*, 102(24):7211–7218, 1980.
- [32] Yosuke Kayanuma. Quantum-size effects of interacting electrons and holes in semiconductor microcrystals with spherical shape. *Phys. Rev. B*, 38(14):9797–9805, Nov 1988.
- [33] Sean A. Blanton, Robert L. Leheny, Margaret A. Hines, and Philippe Guyot-Sionnest. Dielectric dispersion measurements of cdse nanocrystal colloids: Observation of a permanent dipole moment. *Phys. Rev. Lett.*, 79(5):865–868, Aug 1997.
- [34] Kyung-Sang Cho, Dmitri V. Talapin, Wolfgang Gaschler, and Christopher B. Murray. Designing pbse nanowires and nanorings through oriented attachment of nanoparticles. *Journal of the American Chemical Society*, 127(19):7140–7147, 2005.

-
- [35] M Shim and P Guyot-Sionnest. Permanent dipole moment and charges in colloidal semiconductor quantum dots. *JOURNAL OF CHEMICAL PHYSICS*, 111(15):6955–6964, OCT 15 1999.
- [36] Thomas Nann and Jrgen Schneider. Origin of permanent electric dipole moments in wurtzite nanocrystals. *Chemical Physics Letters*, 384(1-3):150 – 152, 2004.
- [37] Liang-shi Li and A. Paul Alivisatos. Origin and scaling of the permanent dipole moment in cdse nanorods. *Phys. Rev. Lett.*, 90(9):097402, Mar 2003.
- [38] Rishikesh Krishnan, Megan A. Hahn, Zhiheng Yu, John Silcox, Philippe M. Fauchet, and Todd D. Krauss. Polarization surface-charge density of single semiconductor quantum rods. *Phys. Rev. Lett.*, 92(21):216803, May 2004.
- [39] M.-H. Tsai, Z.-F. Jhang, J.-Y. Jiang, Y.-H. Tang, and L. W. Tu. Electrostatic and structural properties of gan nanorods/nanowires from first principles. *Applied Physics Letters*, 89(20):203101, 2006.
- [40] Sachin Shanbhag and Nicholas A. Kotov. On the origin of a permanent dipole moment in nanocrystals with a cubic crystal lattice: effects of truncation, stabilizers, and medium for cds tetrahedral homologues. *The Journal of Physical Chemistry B*, 110(25):12211–12217, 2006. PMID: 16800539.
- [41] Eran Rabani. Structure and electrostatic properties of passivated cdse nanocrystals. *The Journal of Chemical Physics*, 115(3):1493–1497, 2001.
- [42] Giancarlo Cicero, Andrea Ferretti, and Alessandra Catellani. Surface-induced polarity inversion in zno nanowires. *Phys. Rev. B*, 80(20):201304, Nov 2009.
- [43] Nguyen Que Huong and Joseph L. Birman. Origin of polarization in polar nanocrystals. *The Journal of Chemical Physics*, 108(5):1769–1772, 1998.
- [44] M. W. Finnis. *Surf. Sci. Newslett*, 24(2), 1999.
- [45] P W Tasker. The stability of ionic crystal surfaces. *Journal of Physics C: Solid State Physics*, 12(22):4977, 1979.
- [46] R. Nosker. Polar surfaces of wurtzite and zincblende lattices. *Surface Science*, 19:291–317, February 1970.
- [47] M. D. Pashley. Electron counting model and its application to island structures on molecular-beam epitaxy grown gaas(001) and znse(001). *Phys. Rev. B*, 40(15):10481–10487, Nov 1989.
- [48] Claudine Noguera. Polar oxide surfaces. *Journal of Physics: Condensed Matter*, 12(31):R367, 2000.

-
- [49] Claudine Noguera, Ariana Pojani, Petr Casek, and Fabio Finocchi. Electron redistribution in low-dimensional oxide structures. *Surface Science*, 507-510:245 – 255, 2002.
- [50] E. Artacho N. C. Bristowe, P. B. Littlewood. The net charge at interfaces between insulators. *Journal of Physics: Condensed Matter*, 23:081001, 2011.
- [51] Andreas Eichler and Georg Kresse. First-principles calculations for the surface termination of pure and yttria-doped zirconia surfaces. *Phys. Rev. B*, 69(4):045402, Jan 2004.
- [52] Jacek Goniakowski, Claudine Noguera, and Livia Giordano. Prediction of uncompensated polarity in ultrathin films. *Phys. Rev. Lett.*, 98(20):205701, May 2007.
- [53] Younjin Min, Mustafa Akbulut, Kai Kristiansen, Yuval Golan, and Jacob Israelachvili. The role of interparticle and external forces in nanoparticle assembly. *Nat Mater*, 7(7):527–538, 07 2008.
- [54] Eran Rabani, Balazs Hetenyi, B. J. Berne, and L. E. Brus. Electronic properties of cdse nanocrystals in the absence and presence of a dielectric medium. *The Journal of Chemical Physics*, 110(11):5355–5369, 1999.
- [55] Xi Zhang, Zhenli Zhang, and Sharon C. Glotzer. Simulation study of dipole-induced self-assembly of nanocubes. *The Journal of Physical Chemistry C*, 111(11):4132–4137, 2007.
- [56] Alex Y. Sinyagin, Artem Belov, Zhioyng Tang, and Nicholas A. Kotov. Monte carlo computer simulation of chain formation from nanoparticles. *The Journal of Physical Chemistry B*, 110(14):7500–7507, 2006.
- [57] M. Born and J. P. Oppenheimer. Zur quantentheorie der molekein. *Ann. Phys.*, 84:457–484, 1927.
- [58] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136(3B):B864–B871, Nov 1964.
- [59] Mel Levy. Electron densities in search of hamiltonians. *Phys. Rev. A*, 26(3):1200–1208, Sep 1982.
- [60] Mel Levy. Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v-representability problem. *Proceedings of the National Academy of Sciences*, 76(12):6062–6065, 1979.
- [61] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140(4A):A1133–A1138, Nov 1965.

-
- [62] Murray Gell-Mann and Keith A. Brueckner. Correlation energy of an electron gas at high density. *Phys. Rev.*, 106(2):364–368, Apr 1957.
- [63] J. P. Perdew and Alex Zunger. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B*, 23(10):5048–5079, May 1981.
- [64] M. Nusair S. H. Vosko, L. Wilk. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian Journal of Physics*, 58(8):1200–1211, 1980.
- [65] Maziar Nekovee, W. M. C. Foulkes, and R. J. Needs. Quantum monte carlo analysis of exchange and correlation in the strongly inhomogeneous electron gas. *Phys. Rev. Lett.*, 87(3):036401, Jun 2001.
- [66] David Mermin Neil Ashcroft. *Solid State Physics*. Brooks Cole, 1976.
- [67] L. P. Bouckaert, R. Smoluchowski, and E. Wigner. Theory of brillouin zones and symmetry properties of wave functions in crystals. *Phys. Rev.*, 50(1):58–67, Jul 1936.
- [68] Hendrik J. Monkhorst and James D. Pack. Special points for brillouin-zone integrations. *Phys. Rev. B*, 13(12):5188–5192, Jun 1976.
- [69] B. J. Austin, V. Heine, and L. J. Sham. General theory of pseudopotentials. *Phys. Rev.*, 127(1):276–282, Jul 1962.
- [70] D. R. Hamann, M. Schlüter, and C. Chiang. Norm-conserving pseudopotentials. *Phys. Rev. Lett.*, 43:1494–1497, Nov 1979.
- [71] N. Troullier and José Luriaas Martins. Efficient pseudopotentials for plane-wave calculations. *Phys. Rev. B*, 43(3):1993–2006, Jan 1991.
- [72] Leonard Kleinman and D. M. Bylander. Efficacious form for model pseudopotentials. *Phys. Rev. Lett.*, 48(20):1425–1428, May 1982.
- [73] Steven G. Louie, Sverre Froyen, and Marvin L. Cohen. Nonlinear ionic pseudopotentials in spin-density-functional calculations. *Phys. Rev. B*, 26(4):1738–1742, Aug 1982.
- [74] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos. Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients. *Rev. Mod. Phys.*, 64(4):1045–1097, Oct 1992.
- [75] W. Kohn. Density functional and density matrix method scaling linearly with the number of atoms. *Phys. Rev. Lett.*, 76(17):3168–3171, Apr 1996.

-
- [76] E. Prodan and W. Kohn. Nearsightedness of electronic matter. *Proceedings of the National Academy of Sciences of the United States of America*, 102(33):11635–11638, 2005.
- [77] Chris-Kriton Skylaris, Peter D. Haynes, Arash A. Mostofi, and Mike C. Payne. Implementation of linear-scaling plane wave density functional theory on parallel computers. *physica status solidi (b)*, 243(5):973–988, 2006.
- [78] Chris-Kriton Skylaris, Peter D. Haynes, Arash A. Mostofi, and Mike C. Payne. Introducing onetep: Linear-scaling density functional simulations on parallel computers. 122(8):084119, 2005.
- [79] R. McWeeny. Some recent advances in density matrix theory. *Rev. Mod. Phys.*, 32(2):335–369, Apr 1960.
- [80] Gregory H. Wannier. The structure of electronic excitation levels in insulating crystals. *Phys. Rev.*, 52(3):191–197, Aug 1937.
- [81] Walter Kohn. Theory of the insulating state. *Phys. Rev.*, 133(1A):A171–A181, Jan 1964.
- [82] Lixin He and David Vanderbilt. Exponential decay properties of wannier functions and related quantities. *Phys. Rev. Lett.*, 86(23):5341–5344, Jun 2001.
- [83] Christian Brouder, Gianluca Panati, Matteo Calandra, Christophe Mourougane, and Nicola Marzari. Exponential localization of wannier functions in insulators. *Phys. Rev. Lett.*, 98(4):046402, Jan 2007.
- [84] Otto F. Sankey and David J. Niklewski. Ab initio multicenter tight-binding model for molecular-dynamics simulations and other applications in covalent systems. *Phys. Rev. B*, 40(6):3979–3995, Aug 1989.
- [85] B. Delley. From molecules to solids with the dmol3 approach. 113(18):7756–7764, 2000.
- [86] S. D. Kenny, A. P. Horsfield, and Hideaki Fujitani. Transferable atomic-type orbital basis sets for solids. *Phys. Rev. B*, 62(8):4899–4905, Aug 2000.
- [87] Javier Junquera, Óscar Paz, Daniel Sánchez-Portal, and Emilio Artacho. Numerical atomic orbitals for linear-scaling calculations. *Phys. Rev. B*, 64(23):235111, Nov 2001.
- [88] José M Soler, Emilio Artacho, Julian D Gale, Alberto García, Javier Junquera, Pablo Ordejón, and Daniel Sánchez-Portal. The siesta method for ab initio order- n materials simulation. *Journal of Physics: Condensed Matter*, 14(11):2745, 2002.

-
- [89] Eduardo Anglada, José M. Soler, Javier Junquera, and Emilio Artacho. Systematic generation of finite-range atomic basis sets for linear-scaling calculations. *Phys. Rev. B*, 66(20):205101, Nov 2002.
- [90] Christopher A. White, Benny G. Johnson, Peter M. W. Gill, and Martin Head-Gordon. Linear scaling density functional calculations via the continuous fast multipole method. *Chemical Physics Letters*, 253(3-4):268 – 278, 1996.
- [91] Matthew C. Strain, Gustavo E. Scuseria, and Michael J. Frisch. Achieving linear scaling for the electronic quantum coulomb problem. *Science*, 271(5245):51–53, 1996.
- [92] Chris-Kriton Skylaris, Arash A. Mostofi, Peter D. Haynes, Oswaldo Diéguez, and Mike C. Payne. Nonorthogonal generalized wannier function pseudopotential plane-wave method. *Phys. Rev. B*, 66(3):035119, Jul 2002.
- [93] Emilio Artacho and Lorenzo Miláns del Bosch. Nonorthogonal basis sets in quantum mechanics: Representations and second quantization. *Phys. Rev. A*, 43(11):5770–5777, Jun 1991.
- [94] J. F. Janak. Proof that $\frac{\partial e}{\partial n_i} = \epsilon$ in density-functional theory. *Phys. Rev. B*, 18(12):7165–7168, Dec 1978.
- [95] P. D. Haynes and M. C. Payne. Corrected penalty-functional method for linear-scaling calculations within density-functional theory. *Phys. Rev. B*, 59(19):12173–12176, May 1999.
- [96] X.-P. Li, R. W. Nunes, and David Vanderbilt. Density-matrix electronic-structure method with linear system-size scaling. *Phys. Rev. B*, 47(16):10891–10894, Apr 1993.
- [97] Arash A. Mostofi, Peter D. Haynes, Chris-Kriton Skylaris, and Mike C. Payne. Preconditioned iterative minimization for linear-scaling electronic structure calculations. 119(17):8842–8848, 2003.
- [98] Chris-Kriton Skylaris, Arash A. Mostofi, Peter D. Haynes, Chris J. Pickard, and Mike C. Payne. Accurate kinetic energy evaluation in electronic structure calculations with localized functions on real space grids. *Computer Physics Communications*, 140(3):315 – 322, 2001.
- [99] Arash A. Mostofi, Chris-Kriton Skylaris, Peter D. Haynes, and Mike C. Payne. Total-energy calculations on a real space grid with localized functions and a plane-wave basis. *Computer Physics Communications*, 147(3):788 – 802, 2002.
- [100] R. W. Nunes and David Vanderbilt. Generalization of the density-matrix method to a nonorthogonal basis. *Phys. Rev. B*, 50(23):17611–17614, Dec 1994.

-
- [101] John M. Millam and Gustavo E. Scuseria. Linear scaling conjugate gradient density matrix search as an alternative to diagonalization for first principles electronic structure calculations. *The Journal of Chemical Physics*, 106(13):5569–5577, 1997.
- [102] Elias Rudberg and Emanuel H Rubensson. Assessment of density matrix methods for linear scaling electronic structure calculations. *Journal of Physics: Condensed Matter*, 23(7):075502, 2011.
- [103] M. R. Jarvis, I. D. White, R. W. Godby, and M. C. Payne. Supercell technique for total-energy calculations of finite charged and polar systems. *Phys. Rev. B*, 56(23):14972–14978, Dec 1997.
- [104] Carlo A. Rozzi, Daniele Varsano, Andrea Marini, Eberhard K. U. Gross, and Angel Rubio. Exact coulomb cutoff technique for supercell calculations. *Phys. Rev. B*, 73(20):205119, May 2006.
- [105] N. D. M Hine, J. Dziedzic, P. D. Haynes, and Chris-Kriton Skylaris. Electrostatic interactions in finite systems treated with periodic boundary conditions: Application to linear-scaling density functional theory. *pending publication*, 2011.
- [106] Xiangyang Huang, Eric Lindgren, and James R. Chelikowsky. Surface passivation method for semiconductor nanostructures. *Phys. Rev. B*, 71(16):165328, Apr 2005.
- [107] Chris-Kriton Skylaris, Peter D Haynes, Arash A Mostofi, and Mike C Payne. Using ONETEP for accurate and efficient $O(N)$ density functional calculations. *Journal of Physics: Condensed Matter*, 17(37):5757, 2005.
- [108] Alberto García and Marvin L. Cohen. First-principles ionicity scales. i. charge asymmetry in the solid state. *Phys. Rev. B*, 47(8):4215–4220, Feb 1993.
- [109] A. Qteish and R. J. Needs. Pseudopotential calculations of the valence-band offsets at the znse/ge, znse/gaas, and gaas/ge (110) interfaces: Effects of the ga and zn 3d electrons. *Phys. Rev. B*, 43(5):4229–4235, Feb 1991.
- [110] Roby Cherian and Priya Mahadevan. Bulk and nanoscale gan: Role of *gad* states. *Phys. Rev. B*, 76(7):075205, Aug 2007.
- [111] Stewart J. Clark, Matthew D. Segall, Chris J. Pickard, Phil J. Hasnip, Matt I. J. Probert, Keith Refson, and Mike C. Payne. First principles methods using castep. *Zeitschrift für Kristallographie*, 220(5-6-2005):567–570, 2011/07/27 2005.

- [112] D. A. Murdick, H. N. G. Wadley, and X. W. Zhou. Condensation mechanisms of an arsenic-rich vapor on gaas (001) surfaces. *Phys. Rev. B*, 75(12):125318, Mar 2007.
- [113] Nicholas D. M. Hine, Mark Robinson, Peter D. Haynes, Chris-Kriton Skylaris, Mike C. Payne, and Arash A. Mostofi. Accurate ionic forces and geometry optimization in linear-scaling density-functional theory with local orbitals. *Phys. Rev. B*, 83(19):195102, May 2011.